

ANÁLISE DE DADOS CATEGORIZADOS
INCOMPLETOS:
FUNDAMENTOS, MÉTODOS E APLICAÇÕES

Carlos Daniel Mimoso Paulino

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM
ESTATÍSTICA

ORIENTADOR: PROF. DR. CARLOS ALBERTO DE BRAGANÇA PEREIRA

- São Paulo, outubro de 1988 -

<u>Pág.</u>	<u>Linha</u>	<u>Onde está</u>	<u>Deve estar</u>
19	13	não em	não tem
21	18	$(1 \ 0'_{(n+1)})$	$(1 \ 0'_{(n-1)})$
21	20	$(\underset{\sim}{I} \ 0) \underset{\sim}{\theta}$	$(\underset{\sim}{I} \ 0) \underset{\sim}{\theta}$
22	24	$\underset{\sim}{Y}$	$\underset{\sim}{y}$
30	expr. de $\underset{\sim}{B}$	4	5
32	enunciado do teor. 2.5	≡	$\langle \underset{\sim}{\cdot} \rangle$
33	01		
33	08		
35	02	$(\Psi_{01} \dots \Psi_{0m})$	$(\Psi_{01} \dots \Psi_{0m})$
36	01	$\underset{\sim}{C}'_k$	$\underset{\sim}{c}'_k$
36	22	$I(\underset{\sim}{\theta}_0, \underset{\sim}{\theta})$	$I(\underset{\sim}{\theta}_0, \underset{\sim}{\theta})$
37	08	$\frac{\partial \underset{\sim}{h}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}$	$\frac{\partial \underset{\sim}{h}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}$
37	19	$\underset{\sim}{\theta}$	$\underset{\sim}{\theta}^*$
37	24	$(\underset{\sim}{\theta}^* \ 1-1'_{q-1} \ \underset{\sim}{\theta})'$	$(\underset{\sim}{\theta}^* \ 1-1'_{q-1} \ \underset{\sim}{\theta}^*)'$
45	última	$\underset{\sim}{S}_0$	$\underset{\sim}{X}_0$
49	02	2.1.3	2.1.4
51	01	$\underset{\sim}{X}' \underset{\sim}{X}$	de $\underset{\sim}{X}' \underset{\sim}{X}$
57	03	mis	mais
94	05	$\underset{\sim}{\lambda}^t, \underset{\sim}{\beta}_i$	$\underset{\sim}{\lambda}^t, \underset{\sim}{\beta}'_i$
97	10	$\sum_{K=1}$	$\sum_{k=1}$
99	antepenúltima	$\underset{\sim}{\phi}(2j) (\underset{\sim}{\theta})$	$\underset{\sim}{\rho}(2j) (\underset{\sim}{\theta})$
100	14	j_h	j_h
100	25	$\underset{\sim}{Z}'(2h) \underset{\sim}{\theta}(3h)$	$\underset{\sim}{Z}'(2h) \underset{\sim}{\theta}(3h)$
110	penúltima	$j \ 1, \dots$	$j=1, \dots$
111	07	$\underset{\sim}{Z}'_{\underset{\sim}{t}} \bar{\underset{\sim}{\theta}}$	$\underset{\sim}{Z}'_{\underset{\sim}{t}} \bar{\underset{\sim}{\theta}}$
111	09	$[-\underset{\sim}{H}_{\underset{\sim}{t}}(\bar{\underset{\sim}{\theta}})]_{N_t}$	$[-\underset{\sim}{H}_{\underset{\sim}{t}}(\bar{\underset{\sim}{\theta}})]_{N_t}$
111	12	$\bar{\underset{\sim}{P}}_T$	$\bar{\underset{\sim}{P}}'_T$

<u>Pág.</u>	<u>Linha</u>	<u>Onde está</u>	<u>Deve estar</u>
122	13	o_p	o_p
122	23	da ordem $o_p(N_t^{-1/2})$	de ordem $o_p(\max_t N_t^{-1/2})$
123	21	$\underline{\underline{C\bar{\theta}}}$	$\underline{\underline{C\tilde{\theta}}}$
123	26	H	H_o
126	5	$\underline{\underline{E\alpha}}$	$\underline{\underline{E\tilde{\alpha}}}$
127	penúltima	(5.50)	(4.60)
148	12-13	denotaremos	denotámos
148	penúltima	$\underline{\underline{ y...}}$	$\underline{\underline{ ...}}$
149	expr.(5.19)	μ_i	$\mu_{i\delta}$
152	20	(3.11)	(3.15)
154	antepenúltima	(65c)	(65b)
154	expr. (5.39)	$[Q'S]$	$[Q'S]$
157	18	$E(\gamma_o)E(\alpha_{\delta_o}^i)$	$E(\gamma_o \underline{n})E(\alpha_{\delta_o}^i \underline{n})$
170	16	$m_i.$	$M_i.$
178	6	$x_1 \underline{\underline{ }} x_2 x_3$	$x_3 \underline{\underline{ }} x_2 x_1$
179	5		
193	última	completicidade de	completicidade de qualquer subálgebra F_1
195	22	(Y,B)	(Y,A)
198	expr.(16)	$g(\theta)$	$q(\theta)$
202	15	$f(\theta,y)$	$g(\theta,y)$
206	10	$F_3 =$	$F_3 \subset$
214	11	ρ_j^i	$\rho^i(j)$
216	expr.(18)	K	k
217	expr.(22)	$i \in \Delta$	$j \in \Delta$
226	3	v_k	b_k
227	16	$\underline{\underline{z'b}}$	$\underline{\underline{z'b}},$
227	17	$\underline{\underline{zb}}$	$\underline{\underline{z'b}}$
236	17	omisso	(43)
237	21	$(\underline{\underline{b}}, \underline{\underline{z}}, c)$	$(\underline{\underline{b}}, \underline{\underline{z}}, -c)$



Aos meus pais

I N D I C E

1.	INTRODUÇÃO	1
2.	A IDENTIFICABILIDADE DE MODELOS ESTATÍSTICOS PARAMÉTRICOS	9
2.1.	A identificabilidade no contexto clássico	10
2.1.1	Conceitos e resultados básicos da teoria da identificabilidade	10
2.1.2	Exemplos ilustrativos	20
2.1.3	Identificabilidade versus informação amostral ..	31
2.1.4	Identificabilidade versus estimabilidade	38
2.1.5	Formas de identificação do modelo	45
2.2.	A identificabilidade no contexto Bayesiano	51
2.2.1	Considerações e instrumentos preliminares	51
2.2.2	A dualidade entre a teoria da identificabilidade paramétrica e a teoria da suficiência amostral	55
2.2.2.1	A suficiência paramétrica Bayesiana e os conceitos Bayesiano e clássico de suficiência amostral	55
2.2.2.2	A suficiência paramétrica Bayesiana minimal	62
2.2.2.3	A identificabilidade Bayesiana e o conceito amostral de "necessidade".	66
2.2.3	A identificabilidade e o conceito de invariância paramétrica	67
2.2.4	A não identificabilidade e o tratamento Bayesiano..	70
3.	FORMULAÇÃO E DISCUSSÃO DO PROBLEMA DE DADOS CATEGORIZADOS INCOMPLETOS	77
3.1.	Introdução	77
3.2.	Caso de um padrão genérico de censura	80
3.2.1.	Processo de registo Bayesianamente ignorável...	84
3.2.2.	Processo de registo "classicamente" ignorável..	90

3.3.	Caso de um padrão de censura em partições genéricas	93
3.4.	Caso de um padrão de censura em partições encaixadas	98
4.	ANÁLISE CLÁSSICA CONDICIONAL DE DADOS COM CATEGORIZAÇÃO COMPLETA E PARCIAL	105
4.1.	Introdução	105
4.2.	Breve revisão da literatura	105
4.2.1.	Método de máxima verosimilhança	106
4.2.2	Outros métodos	108
4.3.	Métodos de máxima verosimilhança	109
4.3.1.	Solução das equações de verosimilhança para o modelo "saturado"	109
4.3.2.	Máxima verosimilhança versus outros métodos de estimação	114
4.3.3.	Testes de ajustamento de vários modelos ...	115
4.3.4.	Alguns comentários adicionais	119
4.4.	Métodos de mínimos quadrados ponderados	120
4.4.1.	Metodologia do χ^2 modificado em uma fase ..	120
4.4.2.	Metodologias dos mínimos quadrados ponderados em duas fases e híbridas	124
4.4.3.	Relação com outros métodos inferenciais ...	126
4.5.	Aplicações	130
5.	ANÁLISE BAYESIANA DE DADOS COM CATEGORIZAÇÃO COMPLETA E PARCIAL . .	143
5.1.	Introdução com breve revisão da literatura	143
5.2.	Solução Bayesiana para um mecanismo de censura não ignorável	145

5.2.1. Distribuições à priori	145
5.2.2. Distribuição preditiva à priori	147
5.2.3. Distribuições à posteriori	149
5.2.4. Momentos à posteriori	153
5.2.5. Inferências sob dependência à priori dos pa râmetros perturbadores	159
5.2.6. Inferências condicionadas nos parâmetros perturbadores	162
5.3. Solução Bayesiana para um mecanismo de censura ignorável	164
5.3.1. Caso de um padrão genérico de censura	164
5.3.2. Caso de um padrão de censura em partições..	166
5.4. Aplicações	171
APÊNDICE A - Alguns resultados sobre espaços vectoriais	185
APÊNDICE B	188
B.1. Funções, partições, estatísticas e σ -álgebras	188
B.2. Algumas notas sobre a estrutura do modelo Bayesiano	194
B.3. Alguns resultados sobre independência condicional	203
B.4. O dual do teorema de factorização clássico: demonstrações	208
APÊNDICE C	211
C.1. Algumas notas sobre as famílias de distribuições Dirichlet e Dirichlet-Multinomial	211
C.2. Funções hipergeométricas	222
C.2.1. Séries hipergeométricas	222
C.2.2. Funções hipergeométricas R de Carlson	226
C.2.3. Funções hipergeométricas R de Carlson	232
C.3. Distribuições de Dirichlet generalizadas	236
REFERÊNCIAS BIBLIOGRÁFICAS	245

NOTA DE AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Carlos Alberto de Bragança Pereira, pela clareza com que delineou o meu projecto de pesquisa, pelo apoio concedido e pela confiança demonstrada ao longo de todo o trabalho.

Ao Prof. Dr. Júlio da Motta Singer, pela sua constante predisposição na concessão de estímulo e de apoio concretizado, particularmente, nos seus úteis comentários e sugestões e na obtenção, por seu intermédio, dos dados que são objecto de análise - cujo uso foi permitido pela Profa. Dra. Tatiana Rozov, a quem ficamos igualmente gratos.

Ao Giovani, pela assistência computacional na efectivação de alguns cálculos do capítulo 5, e à Marita, pela sua colaboração no trabalho de dactilografia e de revisão.

À Antônia, pela paciência e competência demonstradas na dactilografia desta extensa dissertação, e ao Sargento Gomes, pela sua breve mas eficaz participação nesse trabalho.

À Fundação Calouste Gulbenkian, à CAPES e ao CNPQ, pela sucessiva contribuição, com o seu apoio financeiro, para a concretização desta longa tarefa.

SUMMARY

Analyses of incomplete categorical data described in the literature present several deficiencies. An important one refers to the frequent omission of aspects on which these analyses must be founded.

The unidentifiability problem is one of these aspects which is also underestimated in the general statistical literature. In this monography attention is given to this problem through a systematic, unified, illustrative and self-contained description of the parametric identifiability theory. Emphasis is given to questions that are relevant to the statistician and to their illustration with incomplete categorical data situations.

In order to analyse fully and partially categorized data, a report process is introduced to explain the observed pattern of incompleteness. This new element generates an unidentifiable statistical model. The understanding of this fact and the consideration of several reduced models for the structure of such process allow us to justify properly why some of the usual probabilistic models may be used in specific situations. It is possible now to realize in what conditions the report process can be ignored for several methods of inference about the sampling probabilities.

"Classical" and Bayesian methodologies for this problem of incomplete data are described separately.

From the classical inference viewpoint, attention is directed to the maximum likelihood (ML) and weighted least squares (WLS) methodologies. Some interrelationships among these and other methodologies are also referred to. Within the ML procedure, it must be enhanced that the tools necessary to fit log-linear models via general optimization algorithms are derived. The WLS procedure is well-founded by a complete description of the underlying statistical theory.

Distinct unidentifiability implications under the Bayesian approach are confirmed when considering, for the same data set, non-ignorable report mechanisms. In fact, Bayesian solutions are obtained under such assumptions. Many of these solutions are quite tractable in various senses and all of them complement the Bayesian analyses developed in the literature. These latter ones are of a different nature inasmuch as they are based on either the report mechanism ignorability or the conditioning on the nuisance parameters.

Finally, the main methodologies developed are illustrated by several examples with original and real data summarized into contingency tables.

RESUMO

O tratamento de dados categorizados incompletos dominante na literatura apresenta várias deficiências. Uma delas, de enorme importância, diz respeito aos aspectos que fundamentam as análises desenvolvidas e que são geralmente negligenciados.

Um desses aspectos, com uma gama de aplicações mais lata, reporta-se ao problema de identificabilidade - aliás, claramente subestimado na literatura estatística. Daí a atenção que lhe é aqui dedicada através de uma descrição sistemática, unificada, ilustrativa e auto-suficiente das questões da teoria da identificabilidade paramétrica de maior relevância para o estatístico. Neste tratamento confere-se um destaque especial às situações de dados categorizados incompletos.

No problema de dados completa e parcialmente categorizados, a introdução de um processo de registo, a quem se atribui a responsabilidade pela natureza incompleta dos dados, gera um modelo estatístico não identificável. A compreensão deste facto e a consideração de vários modelos reduzidos para a estrutura daquele processo permitem clarificar os modelos probabilísticos para as frequências observadas, geralmente adoptados sem uma fundamentação clara. Deste modo, ficam evidenciadas as diferentes condições que tornam ignorável o mecanismo de registo para os vários tipos de inferências sobre as probabilidades amostrais.

No que concerne à análise estatística do problema de dados com categorização completa e parcial, descrevem-se, em separado, algumas metodologias "clássicas" e Bayesianas.

Na óptica inferencial clássica, a relevância é atribuída às metodologias de máxima verosimilhança (MV) e de mínimos quadrados ponderados (MQP), sob um mecanismo de registo ignorável. No entanto, referem-se também algumas inter-relações entre essas e outras metodologias. No quadro do procedimento de MV destaca-se a derivação dos instrumentos necessários para o ajuste de modelos log-lineares via algoritmos gerais de optimização. O procedimento de MQP é fundamentado através da descrição completa da teoria estatística subjacente.

A diferenciação das implicações da falta de identificabilidade na abordagem Bayesiana é comprovada através da derivação de soluções Bayesianas para o referido problema, sob um mecanismo de registo não ignorável. Estas soluções, muitas das quais bem tratáveis, complementam assim as análises Bayesianas já desenvolvidas, e que repousam ou na ignorabilidade do mecanismo de registo ou no condicionamento nos seus parâmetros.

Finalmente, as principais metodologias desenvolvidas neste trabalho são ilustradas através de vários exemplos com dados reais e originais estruturados em tabelas de contingência.

1. INTRODUÇÃO

A falta de identificabilidade estatística é um dos problemas que não conhece fronteiras. Ele ocorre em áreas científicas extremamente diversas bem como em diferentes campos da mesma área. A plausibilidade da sua ocorrência aumenta consideravelmente com a complexidade do fenómeno e do modelo postulado para o descrever. Daí a importância que lhe é dedicada, por exemplo, pela literatura econométrica.

Esta atitude já não encontra correspondência na literatura estatística, que lhe dispensa, quando é o caso, uma atenção bem modesta apesar do seu surgimento em vários campos da modelagem estatística. A dispersão do material mais relevante para o estatístico é um indício claro dessa situação. Mais, são raros os livros de texto de inferência estatística que mencionam, mesmo implicitamente, o conceito de identificabilidade. As escassas tentativas de elucidação desse conceito pecam por um tratamento incompleto e desequilibrado. Em particular, as rotineiras ilustrações com modelos econométricos não são, certamente as mais apropriadas para o esclarecimento e motivação para a importância desse problema daquele que trabalha em áreas mais "standard" da Estatística. A referência ao conceito quando existe, não é, em geral nada incisiva. Várias vezes, ele surge diluído ou obscurecido por alusão a conceitos inferenciais que não lhe são equivalentes em geral. É o que sucede na grande maioria da literatura de modelos lineares que descreve o conceito sob a terminologia, já familiar, de estimabilidade (linear).

Todos estes aspectos têm, na nossa opinião, contribuído para limitar, em largos sectores da comunidade estatística, a compreensão das noções que gravitam em torno da questão identificabilidade versus não-identificabilidade e das suas reais implicações inferenciais. Sentimos, da nossa experiência quotidiana como docente e pesquisador de Estatística, que o conhecimento do problema é de facto superficial no seio de larga maioria dos estatísticos.

Efectivamente, não é invulgar depararmo-nos com respostas insatisfatórias a questões centrais como: O que é realmente um modelo estatístico identificável? Identificabilidade e estimabilidade são apenas termos sinónimos? Quais as limitações que a presença de um modelo não identificável induz nos procedimentos inferenciais? Como desbloquear as inferências sobre parâmetros não identificáveis? Como actua a informação à priori injectada no modelo na sua identificação? As implicações inferenciais da falta de identificabilidade são indiferentes ao tipo de abordagem (ortodoxa ou Bayesiana) que se pretende adoptar?

A fortiori, pouca ou nenhuma ideia se faz de questões mais específicas como: Que ligação se pode estabelecer entre identificabilidade e informação amostral? É possível visualizar algum paralelismo entre a teoria da suficiência amostral e o corpo de resultados que configura o que pode ser apelidado de teoria da identifica

bilidade? Designadamente, como se pode constatar a autêntica dualidade entre as suficiências amostral e paramétrica no âmbito do modelo estatístico Bayesiano? Que utilidade podem ter os conceitos da teoria da invariância na interpretação do conceito de identificabilidade?

Face ao exposto, no capítulo 2 tentamos dar o nosso contributo para a clarificação das respostas às questões supra-mencionadas através de uma descrição sistemática, unificada e auto-suficiente dos aspectos da teoria da identificabilidade paramétrica de maior relevância para o estatístico. Os resultados de maior interesse são sistematizados, comentados, clarificados e, alguns deles, desenvolvidos.

De entre os aspectos mais inovadores da secção 2.1 destacamos:

- A visualização dos conceitos de funções identificáveis e suficientes em termos da relação entre as partições por elas induzidas e a partição induzida pela relação de equivalência observacional, o que nos parece facilitar a derivação e compreensão de alguns resultados analíticos;
- O enquadramento de resultados de Rothenberg (1971) e de Reiersol (1963) como consequências da caracterização de funções identificáveis de Kadane (1972);
- A descrição detalhada do modo de actuação de restrições exactas, necessariamente não lineares, na identificação do espaço paramétrico, o que permite, na sua aplicação a contextos lineares, descrita numa linguagem de espaços vectoriais, justificar os modelos restringidos identificáveis vulgarmente usados.

Nesta secção 2.1 devemos realçar ainda a orientação que é dada à parte ilustrativa. A ilustração dos conceitos e resultados é feita através de exemplos minimamente familiares ao estatístico que retratam situações reais sem caírem em aplicações desnecessárias. Confere-se um destaque especial a exemplos com dados categorizados pelo facto de constituírem uma fonte potencial de problemas de identificabilidade, derivados da própria génese do modelo multinomial a partir de processos de Poisson, do uso do modelo estrutural log-linear em frequentes análises de dados e num plano menos específico, da possível natureza incompleta de certas observações. Analisam-se em pormenor vários problemas com dados, contínuos ou categorizados, diversamente incompletos. O relacionamento entre os modelos estruturais para os dados quando completos e os decorrentes modelos para os dados incompletos clarificam a relação de dependência da identificabilidade destes últimos para com aqueles e com a própria configuração dos dados observados.

Mesmo quando um subconjunto próprio da amostra é completamente observado - situação em que se baseiam os restantes capítulos - a incorporação de um processo de registo na formulação do problema suscita questões de não identificabilidade no espaço paramétrico ampliado. O enquadramento harmonioso deste capítulo no conjunto

da dissertação alicerça-se pois na compreensão de que a questão de identificabilidade constitui parte integrante dos fundamentos das análises de dados incompletos.

No tocante ao conteúdo da secção 2.2, exposto geralmente na linguagem abstracta da teoria da medida e integração, à semelhança do tratamento da suficiência amostral de Halmos-Savage - Bahadur, o nosso maior contributo concentra-se nos seguintes aspectos:

- Alguns trabalhos (Barankin [1960], Picci [1977], Florens et al. [1983]) evidenciaram a dualidade que existe entre os conceitos Bayesianos de suficiência amostral e paramétrica e dos correspondentes elementos minimais. Neste quadro, a definição do análogo dual do conceito de estatística completa limitada permite-nos estabelecer um critério de suficiência minimal Bayesiana que é a versão Bayesiana do dual do correspondente resultado de Lehmann-Scheffé. Uma vez seleccionada uma σ -álgebra paramétrica suficiente minimal, visualiza-se linearmente o conceito Bayesiano de σ -álgebra (estatística ou partição) identificável como dual do conceito de σ -álgebra necessária de Bahadur;
- Com base no conhecimento prévio das condições que garantem a equivalência da suficiência amostral Bayesiana e Fisheriana, define-se um análogo dual do conceito Fisheriano de suficiência para o espaço paramétrico e demonstra-se, por dualidade com aquele resultado, a sua equivalência com o correspondente conceito Bayesiano adoptado. Posto isto, obtém-se um critério de suficiência paramétrica que é o dual do teorema de factorização clássico de Bahadur;
- Recorrendo à visualização de funções identificáveis à luz da invariância face a um determinado grupo de transformações (resultado de Kadane [1972]), obtém-se uma nova caracterização de identificabilidade baseada no conceito de invariância maximal paramétrica;
- O tratamento abstracto desta secção é "diluído" pela ilustração dos conceitos e resultados definidos através de alguns exemplos originais ligados a dados incompletos e pelos comentários sobre a nossa interpretação dos problemas suscitados pela falta de identificabilidade nas inferências clássica e Bayesiana.

Os capítulos seguintes da dissertação tratam do problema de dados categorizados incompletos e das metodologias inferenciais mais vulgarmente utilizadas na análise de dados envolvendo categorização completa e parcial.

Face à falta de unidade, generalidade e de clareza no tratamento de vários problemas com dados perdidos, dominante na literatura, o capítulo 3 tenta remar contra tais deficiências objectivando "pôr a casa em ordem". A concretização deste objectivo passa pela interpretação da natureza incompleta dos dados e pelo esclarecimento

das dificuldades e das vicissitudes analíticas, tornando patentes os pressupostos que viabilizam e validam muitas das análises secamente descritas.

Este capítulo começa por uma descrição unificada do problema genérico de dados incompletos de modo a integrar os vários tipos de classificação incompleta passíveis de ocorrência por perda de certa informação no processo classificatório das unidades amostrais. Esta ocorrência é interpretada à luz de uma intervenção de um mecanismo de registo, que vai responsabilizar-se pelo tipo de classificação sofrido por cada unidade, cujo resultado do processo de amostragem é suposto não observado.

A consideração de processos de amostragem e de registo multinomiais clarifica a forma correcta da verosimilhança para todos os dados e torna patente a falta de identificabilidade paramétrica. Dada a frequência com que se obtém a categorização completa de um subconjunto próprio da amostra, a atenção concentra-se no problema que admite tal padrão de registo. A falta de identificabilidade do modelo probabilístico gerador dos dados, geralmente omitida na literatura, permite compreender, do ponto de vista das inferências sobre as probabilidades amostrais, o habitual refúgio no condicionamento às probabilidades condicionais de registo, ou na admissão de modelos para estes últimos parâmetros que tornam ignorável o mecanismo de registo. Alguns destes modelos são discutidos com a dupla finalidade de clarificar as suas implicações distribucionais e as desiguais condições para a ignorabilidade do processo de registo exigidas pelas inferências baseadas na verosimilhança e na teoria da amostragem.

Dada a sua ocorrência frequente em tabelas de contingência, atribui-se um destaque especial ao caso de um padrão de censura em partições, no qual se evidencia a simplificação interpretativa de vários resultados distribucionais, sob a ignorabilidade do mecanismo de registo. Esta simplificação estende-se ao plano computacional, no acto de traçar inferências baseadas na verosimilhança, quando as partições formam uma estrutura encaixada. Exemplifica-se, em particular, para o encaixe em dois níveis, o método de factorização da verosimilhança, conducente a resultados explícitos para as estimativas de máxima verosimilhança.

O capítulo 4 trata das inferências clássicas condicionadas a um mecanismo de registo ignorável, baseadas na aplicação, ao problema de dados com categorização completa e parcial estruturados em partições, das metodologias de máxima verosimilhança e de mínimos quadrados generalizados (ou ponderados, como é usual).

Começa-se por uma breve digressão pela literatura "clássica" de dados incompletos, cuja análise dá uma ideia do porquê da abordagem de vários aspectos que são, até ao momento, insuficientemente descritos e/ou esclarecidos, ou mesmo, omitidos.

No tocante à metodologia de máxima verosimilhança, a estimação no modelo "saturado" é examinada em termos dos algoritmos de Newton-Raphson, EM e de Hocking and Oxspring (1971), procedendo-se à sua comparação em padrões especiais de censu-

ra. Descreve-se o ajustamento de modelos reduzidos salientando-se a derivação dos instrumentos necessários para a implementação do método de Newton-Raphson no ajuste de modelos log-lineares. Deste modo se contribui para tornar aplicável um procedimento largamente vantajoso, pela sua generalidade e eficiência, quando comparado com a técnica mais comum que combina os algoritmos EM e de ajustamento proporcional iterativo (usada, e.g., por Fuchs [1982]).

A estimação de máxima verosimilhança, sob o modelo multinomial subjacente, é ainda relacionada com os métodos do χ^2 de Pearson mínimo e dos mínimos quadrados ponderados. Alguns comentários são feitos sobre a comparação das soluções de máxima verosimilhança com as que se obteriam se o modelo probabilístico subjacente fosse multinomial-produto.

Com base em Koch, Imrey and Reinfurt (1972), desenvolve-se a teoria inferencial subjacente à aplicação ao referido problema da metodologia de mínimos quadrados ponderados de Grizzle, Starmer and Koch (1969). Nela se destaca a diferenciação que existe no ajuste de modelos estritamente lineares, por um lado, e de outros modelos funcionais lineares (e.g., modelos log-lineares), por outro. Descrevem-se ainda algumas relações existentes com outras metodologias "clássicas", particularmente com aquela baseada na teoria da informação.

Finalmente, as metodologias de máxima verosimilhança e de mínimos quadrados ponderados são ilustradas cabalmente através de dois exemplos originais com dados reais provenientes de um estudo feito no Hospital das Clínicas da Universidade de São Paulo.

O capítulo 5 respeita à análise Bayesiana do problema geral de dados completa e parcialmente categorizados e inicia-se com uma breve revisão da correspondente literatura.

A análise detalhada do conteúdo destes trabalhos revela vários aspectos que ainda não foram resolvidos. No contexto de um esquema de amostragem multinomial com um padrão geral de censura, as inferências sobre os parâmetros de interesse, sob um mecanismo de registo informativo, têm revestido uma natureza condicional às probabilidades condicionais de registo. Esta característica é partilhada compreensivelmente pelas correspondentes soluções clássicas devido ao problema de falta de identificabilidade. Contudo, como se argumenta no capítulo 2, este último problema não suscita a mesma gravidade no quadro Bayesiano.

Neste capítulo desenvolve-se precisamente a teoria para a realização de inferências não condicionais, baseadas no uso de distribuições à priori Dirichlet generalizadas para o vector das probabilidades dos resultados conjuntos dos processos de amostragem e de registo.

Nas sub-seções 5.2.1 - 5.2.4 derivam-se os principais resultados com o uso de distribuições Dirichlet independentes para os vários tipos de parâmetros, as

quais implicam uma distribuição à priori Dirichlet para o referido vector das probabilidades conjuntas das categorias e dos diferentes tipos de classificação.

Em certos casos pode acontecer que as crenças à priori apontem no sentido de haver dependência à priori entre os vários parâmetros, o que leva à inadequação do tipo de distribuições à priori acima citadas. Por isso, na sub-secção 5.2.5 generaliza-se o tratamento anterior, adoptando para aquelas probabilidades conjuntas uma distribuição Dirichlet generalizada que traduza a dependência à priori entre os vectores das probabilidades amostrais e das várias probabilidades condicionais de registo. Os resultados obtidos revestem compreensivelmente uma forma muito menos tratável que no primeiro caso.

Na sub-secção 5.2.6 descrevem-se os resultados necessários para traçar inferências condicionadas nos parâmetros perturbadores, baseadas numa distribuição à priori Dirichlet generalizada deles dependente, estendendo assim os resultados de Dickey, Jiang and Kadane (1987).

A secção 5.3 trata da teoria inferencial sob um mecanismo de registo ignorável, desenvolvida por esses mesmos autores, com uma dupla finalidade:

- registar os resultados à posteriori em forma fechada por aplicação do método das expansões multinomiais;
- derivar as expressões explícitas de parâmetros à posteriori no caso de um padrão de censura encaixada em dois níveis.

Finalmente, as soluções correspondentes a um mecanismo ignorável e não ignorável, descritas respectivamente em 5.3 e 5.2.1 - 5.2.4, são ilustradas e comparadas (juntamente com a solução usual relativa apenas aos dados completos) através de dois problemas com dados reais e originais, suscitados por estudos realizados no Hospital das Clínicas da Universidade de São Paulo.

Com o propósito de evitar uma extrema densidade do corpo do texto, optou-se por colocar em apêndice algumas notas sobre certos conceitos e resultados mais ou menos desconhecidos, e a derivação de outros resultados com um grau variável de originalidade, ambos frequentemente exigidos pelo desenvolvimento dos assuntos focados.

A estrutura desta extensa dissertação permite assim diferentes opções de leitura parcial. O leitor mais interessado pelo problema de identificabilidade pode concentrar-se no capítulo 2. A secção 2.1 poderá, por si só, satisfazer as suas pretensões, dado que nela encontrará, segundo cremos, resposta às questões mais relevantes que se prendem com o conceito de identificabilidade. O leitor mais atraído por uma abordagem abstracta de conceitos da estatística matemática, num quadro Bayesiano, poderá manter um contacto adicional com a teoria da identificabilidade

através da secção 2.2. O leitor que pretende compreender a formulação e o tratamento básico do problema de dados categorizados incompletos encontrará no capítulo 3 material que poderá contribuir para a consecução dos seus objectivos. Consoante as tendências e/ou interesses em matéria inferencial, o capítulo 4, para uma óptica clássica, e o capítulo 5, para uma óptica Bayesiana, constituem instrumentos que dotam o estatístico com várias soluções para o problema de dados completa e parcialmente categorizados.

Por fim, algumas palavras sobre aspectos notacionais adoptados. Em geral, não se faz qualquer distinção notacional entre vectores aleatórios e seus valores observados. A distinção de significado tornar-se-á clara a partir do contexto. Os vectores aleatórios observáveis são designados por letras latinas e os parâmetros (desconhecidos) das suas distribuições, entendidos ou não como vectores aleatórios, por letras gregas. Porém, os parâmetros (conhecidos) das distribuições paramétricas são denotados por letras latinas.

As funções densidade (probabilidade) de vectores aleatórios distintos são frequentemente designadas, apesar da imperfeição técnica, pelo mesmo símbolo, sendo a distinção operada pelo argumento dessas funções. O valor esperado de alguma função aleatória menciona expressamente a distribuição envolvida. Por exemplo, $E_{\underline{X}|\underline{Y}}[h(\underline{X})]$, ou em certos casos, $E[h(\underline{X})|\underline{Y}]$, denota o valor esperado com respeito à distribuição do vector aleatório \underline{X} condicional ao vector \underline{Y} - as quantidades matriciais são sempre acompanhadas por um til - da função vectorial $h(\underline{X})$. As matrizes são sempre designadas por uma letra latina maiúscula.

As distribuições multivariadas terão sempre indicada a respectiva dimensão. Quando a distribuição se reporta a um vector de k componentes sujeitas a uma restrição de igualdade, a letra k no símbolo indicador da distribuição indica que esta é de dimensão $k-1$. É o que sucede, por exemplo, com os símbolos M_k , D_k e DM_k para representar as distribuições Multinomial, Dirichlet e Dirichlet-Multinomial. Achou-se conveniente adoptar uma notação sobreparametrizada para estas distribuições. Por exemplo, sendo \underline{n} um vector $k \times 1$ de inteiros não negativos tal que $\sum_{k=1}^k n_k = N$, escreveremos $\underline{n} | N, \underline{\theta} \rightsquigarrow M_k(N, \underline{\theta})$ para significar que qualquer sub-vector de $k-1$ elementos de \underline{n} segue uma distribuição multinomial $(k-1)$ -dimensional com parâmetros N e $\underline{\theta}$, onde $\underline{\theta}$ é tal que $\sum_{k=1}^k \theta_k = 1$.

Finalmente, será usada a notação usual $\underline{X} \perp\!\!\!\perp \underline{Y} | \underline{Z}$ para indicar que os vectores aleatórios \underline{X} e \underline{Y} são condicionalmente independentes dado \underline{Z} .

2. A IDENTIFICABILIDADE DE MODELOS ESTATÍSTICOS PARAMÉTRICOS

O denominado problema de identificação tem a particularidade de estar presente em campos extremamente variados. Nalguns deles, como na ciência de sistemas e controle determinístico, a estrutura física subjacente não envolve estocasticidade. Esta diversidade de situações torna compreensível as várias formulações e sentidos em que têm sido usados os termos identificabilidade/identificação. Preocupar-nos-emos aqui com o conceito mais relevante para a Estatística, onde a estrutura do modelo é definida por variáveis aleatórias, e que é, às vezes, designado por identificabilidade estatística ou identificabilidade por distribuição. Usaremos apenas a designação usual de identificabilidade para nos referirmos a esse conceito, o qual está relacionado com a possibilidade de determinação unívoca dos parâmetros a partir da distribuição das variáveis aleatórias observadas.

Historicamente, os primeiros problemas de identificabilidade surgiram nas áreas de econometria e de psicometria (análise factorial). A literatura econométrica, em particular, está recheada de estudos de problemas de identificabilidade, onde se destacam os exemplos clássicos dos chamados modelos estruturais e dos modelos com erros nas variáveis. A área estatística de modelos lineares e planeamento de experiências também constitui, desde os primeiros tempos, um campo padrão para a ocorrência de situações de não identificabilidade paramétrica. Os problemas com dados incompletos são, independentemente da área em que surgem, exemplos privilegiados para o confronto do pesquisador com modelos paramétricos não identificáveis.

Apesar do conceito de identificabilidade poder estar englobado em contextos não paramétricos (vide, e.g., Koopmans and Reiersol [1950], Basu [1983]), circunscrever-nos-emos à sua aplicação em modelos estatísticos paramétricos. A definição adoptada do conceito é independente do procedimento inferencial. Contudo, a estrutura mais complexa do modelo estatístico Bayesiano relativamente ao modelo "clássico" leva a que certos resultados possam ter interpretações adicionais no primeiro caso. Este aspecto, juntamente com o facto da não identificabilidade ter implicações inferenciais distintas nas abordagens clássica e Bayesiana, justifica a separação da análise do problema nos dois contextos, feita através da divisão do capítulo em duas secções.

2.1. - A identificabilidade no contexto clássico

2.1.1. - Conceitos e resultados básicos da teoria da identificabilidade

Consideremos o espaço estatístico $(Y, \mathcal{A}, \mathcal{P})$, onde Y é o espaço amostral, \mathcal{A} uma σ -álgebra nele definida e \mathcal{P} uma família de medidas de probabilidade em (Y, \mathcal{A}) especificada por $\mathcal{P} = \{P_{\theta} : \theta \in E\}$, onde E designa o espaço paramétrico.

Def. 2.1: Dois pontos de E , θ_0 e θ_1 , dizem-se observacionalmente equivalentes (e escreve-se $\theta_1 \sim \theta_0$) se $P_{\theta_1}(A) = P_{\theta_0}(A)$, $\forall A \in \mathcal{A}$.

Denotando $F(y|\theta)$ a função de distribuição do vector aleatório observado $y \in Y$, então $\theta_1 \sim \theta_0$ sse $F(y|\theta_1) = F(y|\theta_0)$, $\forall y \in Y$, ou equivalentemente, sse $dF(y|\theta_1) = dF(y|\theta_0)$ a menos de conjuntos de medida $dF(y|\theta_0)$ nula. Deste modo, no caso dominado, a função de verosimilhança é, essencialmente para todos os dados, constante em todos os pontos observacionalmente equivalentes. E para cada par de pontos não equivalentes nesse sentido, ela assume valores diferentes para pelo menos um $y \in Y$.

É fácil constatar que \sim é uma relação de equivalência, induzindo assim em E uma partição nas classes de equivalência $[\theta_0] = \{\theta_1 \in E : P_{\theta_1} = P_{\theta_0}\}$. Denotaremos essa partição, habitualmente designada por conjunto quociente de E segundo \sim , por E/\sim .

Def. 2.2: i) O ponto particular $\theta_0 \in E$ diz-se (globalmente) identificável se $[\theta_0] = \{\theta_0\}$.

ii) E (ou o modelo \mathcal{P}) diz-se identificável quando $[\theta] = \{\theta\}$, $\forall \theta \in E$, isto é, quando E/\sim consiste na partição mais fina possível, $E/\sim = \{\{\theta\}, \theta \in E\}$.

A definição 2.2.ii) ilustra que E não é identificável se existe pelo menos um $\theta_1 \in E$ tal que $[\theta_1]$ não é singular. Assim, é possível teoricamente que E não seja identificável mas haja pontos que satisfaçam a definição 2.2.i). Por exemplo, se para um dado $y \in Y$, $F(y|\theta)$ tem um máximo absoluto único, atingido no ponto $\theta_0 \in E$, então θ_0 é claramente identificável, independentemente de haver ou não equivalência observacional nos restantes pontos de E . No entanto, nas situações usuais de não identificabilidade, nenhuma classe de equivalência é singular. Em contextos paramétricos, E é habitualmente um subconjunto de um espaço métrico, o que permite definir o conceito mais fraco de identificabilidade local:

Def. 2.2: iii) O ponto $\theta_0 \in E$ diz-se localmente identificável se $[\theta_0] \cap V(\theta_0) = \{\theta_0\}$, onde $V(\theta_0)$ denota uma vizinhança aberta de θ_0 .

Este conceito tem a sua utilidade porque em certos modelos não identificáveis, mesmo sem nenhum ponto globalmente identificável, uma dada classe de equivalência $[\underline{\theta}_0]$ pode ser singular numa determinada vizinhança de $\underline{\theta}_0$, o que permite, como veremos em 2.1.3, estabelecer uma relação com medidas de informação amostral.

A identificabilidade de Ξ é assim definida em termos da injectividade da aplicação de Ξ em $[0,1], (\underline{\theta}) \rightarrow F(y|\underline{\theta})$, para pelo menos um $y \in \mathcal{Y}$. É interessante notar já que a injectividade desta aplicação é necessariamente garantida quando $\underline{\theta}$ é directamente interpretável como uma característica de $F(y|\underline{\theta})$. Mais concretamente, se existe uma função $T(y)$ tal que $\underline{\theta} = E[T(y)|\underline{\theta}]$, a relação $F(y|\underline{\theta}_1) = F(y|\underline{\theta}_0)$, $\forall y \in \mathcal{Y}$ para qualquer par de pontos $\underline{\theta}_0$ e $\underline{\theta}_1$ de Ξ implica $\underline{\theta}_1 = \underline{\theta}_0$, ou seja, a identificabilidade de Ξ .

A definição de identificabilidade ilustra claramente que a sua presença ou ausência é uma característica da especificação adoptada para o modelo que gera as observações na experiência estatística em causa, sendo pois independente do procedimento inferencial que se pretenda seguir.

Contudo, o problema da identificabilidade está intimamente associado com o problema da inferência estatística propriamente dita no sentido em que as limitações deste estão fortemente dependentes daquele. Com efeito, uma vez observados os dados $y \in \mathcal{Y}$, a efectivação de um procedimento inferencial sobre $\underline{\theta}$ (ou outras características distribucionais desconhecidas) exige não só, e obviamente, o relacionamento de tais características com a distribuição geradora das observações, mas também a injectividade da aplicação $\underline{\theta} \rightarrow F(\cdot|\underline{\theta})$ (ou da aplicação correspondente à característica em estudo). Como num modelo não identificável existem vários pontos de Ξ que produzem as observações com a mesma plausibilidade, uma realização amostral dá apenas informação sobre uma dada classe de equivalência, não conseguindo operar uma selecção adicional dentro desta. Particularmente, se $\hat{\underline{\theta}} = \hat{\underline{\theta}}(y)$ é um ponto que maximiza a função de verosimilhança de um parametro não identificável, todo o elemento $\tilde{\underline{\theta}} = \tilde{\underline{\theta}}(y) \in [\hat{\underline{\theta}}]$, distinto de $\hat{\underline{\theta}}$, pode ser adoptado como um outro estimador de máxima verosimilhança.

É conveniente realçar que a não unicidade do estimador de máxima verosimilhança (bem como de qualquer outro tipo de estimador, como veremos à frente) de um parâmetro não identificável é essencialmente independente dos dados. Este problema tem assim uma natureza distinta daquele que ocorre em vários modelos identificáveis, onde a não unicidade referida decorre da não concavidade global da função de verosimilhança para alguns (eventualmente todos) conjuntos particulares de observações. Situações com dados categorizados incompletos, do tipo daquelas citadas em 2.1.2., constituem muitas vezes exemplos onde o método da máxima verosimilhança não produz estimativas únicas de parâmetros identificáveis para certos conjuntos de dados. Contudo, para os dados mais frequen

tes, já se obtêm estimativas únicas. Outras situações existem onde a concavidade da função de verosimilhança não depende do conjunto particular de dados. Consideremos, por exemplo, que $\underline{y} = (y_1, \dots, y_n)'$ é uma amostra aleatória de dimensão n de uma distribuição uniforme em $[\theta - 1/2, \theta + 1/2]$, $\theta \in \mathbb{R}$. É bem sabido que $f(\underline{y} | \theta) = I_{(-\infty, \theta + 1/2]}(y_{(n)}) I_{[\theta - 1/2, \infty)}(y_{(1)})$, onde $y_{(n)} = \max_{1 \leq i \leq n} y_i$,

$y_{(1)} = \min_{1 \leq i \leq n} y_i$ e $I_A(\cdot)$ é a função indicadora do conjunto A , o que reflecte claramente a identificabilidade de θ . Como a função de verosimilhança é definida por $l(\theta | \underline{y}) = I_{[\underline{y}_{(n)} - 1/2, \underline{y}_{(1)} + 1/2]}(\theta)$, fica claro que independentemente de \underline{y} , qualquer valor de $\hat{\theta}_t = t \underline{y}_{(n)} - 1/2 + (1-t)(\underline{y}_{(1)} + 1/2)$, $t \in [0, 1]$ é um estimador de máxima verosimilhança de θ . A não unicidade deste tipo de estimação, inerente à forma específica da função de verosimilhança, não é assim sinónimo de falta de identificabilidade. Em situações deste género, a não unicidade da estimação não é independente do método de estimação. O recurso a outros métodos possibilita em geral a obtenção de estimadores únicos, como acontece neste caso com a aplicação do método dos momentos, por exemplo.

Para além da arbitrariedade da estimação de um parâmetro não identificável, outro aspecto indesejável, com sérias implicações na teoria assintótica de testes de hipóteses, é a ausência de estimadores consistentes. De facto, sendo θ_0 o verdadeiro valor do parâmetro, a não identificabilidade de θ_0 implica que a distribuição de qualquer estimador, $\tilde{\theta}_n$, seu, qualquer que seja o número n de observações, é a mesma para todos os pontos $\theta_1 \in [\theta_0] \neq \{\theta_0\}$. Se $\tilde{\theta}_n$ fosse um estimador consistente (em probabilidade) para θ_0 , $P_{\theta_0} [\|\tilde{\theta}_n - \theta_0\| < \varepsilon] \rightarrow 1, \forall \varepsilon > 0$ e, $n \rightarrow \infty$

como consequência da não identificabilidade, dever-se-ia verificar igualmente $\tilde{\theta}_n \xrightarrow{P} \theta_1$, para todo o $\theta_1 \in [\theta_0]$, o que é um absurdo por contrariar a definição de consistência. Assim, a existência de estimadores consistentes para um parâmetro é condição suficiente para a sua identificabilidade. Deve, contudo, realçar-se que a proposição recíproca não é verdadeira. Sob a verificação de certas condições de regularidade, a identificabilidade implica a existência de estimadores consistentes (não necessariamente o estimador de máxima verosimilhança) quando a amostra é aleatória (LeCam [1956]). Esta situação pode mudar de figura se a amostra não é constituída por observações i.i.d. - veja-se o exemplo do modelo de regressão de Gabrielsen (1978). Zacks (1971, sec. 5.3) fornece ainda novos exemplos da inconsistência do estimador de máxima verosimilhança de um parâmetro identificável.

Os problemas inferenciais citados acima não implicam necessariamente que na da haja a fazer com um modelo originalmente não identificável. Se o interesse está em testar a significância de hipóteses paramétricas sob as quais o correspondente espaço paramétrico restringido é identificável, pode acontecer que a aplicação de um dado critério de teste não seja inviabilizada (vide 2.2.4). Por

outro lado, muitos aspectos desconhecidos do modelo podem ser traduzidos por funções paramétricas $\underline{\phi} = \underline{\phi}(\underline{\theta}): \Xi \rightarrow \Phi$ que são constantes em cada classe de equivalência $[\underline{\theta}]$. Isto significa que se pode estabelecer uma correspondência, não necessariamente biunívoca entre Ξ/\sim e $\underline{\phi}(\Xi)$ de modo que valores distintos de $\underline{\phi}$ correspondem a valores diferentes de $F(\underline{y}|\underline{\theta})$, $\forall \underline{y} \in \mathcal{Y}$. A função $\underline{\phi}(\underline{\theta})$ é assim identificada pelos dados, o que possibilita a realização de inferências inequívocas sobre ela.

Esta última situação evidencia que o conceito de partição é mesmo mais fundamental que o de função na averiguação da identificabilidade. Por um lado, cada função $\underline{\phi}(\underline{\theta})$ induz em Ξ uma partição $\Pi = \{\pi_{\underline{\phi}_0} : \underline{\phi}_0 \in \underline{\phi}(\Xi)\}$, onde o elemento da partição indexado pelo valor $\underline{\phi}_0$ é definido por $\pi_{\underline{\phi}_0} = \{\underline{\theta} \in \Xi: \underline{\phi}(\underline{\theta}) = \underline{\phi}_0\}$.

Mas, por outro, uma dada partição pode ser induzida por várias funções, cada uma delas assumindo um único valor em cada elemento da partição e valores distintos em elementos distintos da partição. Além disso, o raciocínio em termos de partições vai permitir pôr logo em destaque, de uma forma simples, um certo paralelismo existente entre os conceitos que definiremos aqui e certos conceitos que gravitam em torno da suficiência de partições do espaço amostral. Esta analogia vai poder ser desenvolvida e tornada rigorosa em 2.2, uma vez que se torna indispensável a criação de uma estrutura de medida em Ξ para se poder vislumbrar a autêntica dualidade entre os conceitos básicos da teoria da identificabilidade paramétrica e da teoria da suficiência amostral.

Consideremos então o conjunto das partições de Ξ e sejam Π_1 e Π_2 dois membros desse conjunto.

Def. 2.3: Diz-se que Π_1 é uma redução de Π_2 (ou que Π_1 é mais "grossa" do que Π_2), e escreve-se $\Pi_1 < \Pi_2$, se toda a parte de Π_1 é uma união de partes de Π_2 .

É fácil constatar que a relação binária $<$ goza das propriedades reflexiva, anti-simétrica lata e transitiva. Ela é, de facto, a relação de ordem lata associada à relação de ordem parcial (que verifica a anti-reflexividade, a anti-simetria, a transitividade mas não a tricotomia) " Π_1 é uma redução estrita de Π_2 " (ou " Π_1 é estritamente mais grossa que Π_2 ") traduzindo a ideia de que, no contexto anterior, pelo menos uma parte de Π_1 é uma união de mais do que uma parte de Π_2 .

Num espaço paramétrico não identificável, a partição mais fina (com mais partes) de Ξ , $\Pi = \{\{\underline{\theta}\}, \underline{\theta} \in \Xi\}$, não consegue ser discriminada pelos dados. Por construção, a partição mais fina que é detectável pelos dados é aquela que denotamos por Ξ/\sim . Se uma dada partição Π é uma redução de Ξ/\sim , ela pode ser identificada pelos dados, uma vez que partes distintas de Π incluem classes de e

quivalência observacional distintas, ou equivalentemente, que cada $[\theta]$ está contido numa parte de Π . Quando Π é definida por uma função $\phi(\theta)$, isto significa que ϕ é constante em cada classe de equivalência. Em suma:

Def. 2.4: i) A partição Π de E é identificável se Π é uma redução de E/\sim .

ii) A função $\phi: E \rightarrow \Phi$ é identificável se a partição por ela induzida é identificável, isto é, $\forall \theta_1, \theta_0 \in E, \theta_1 \in [\theta_0] \Rightarrow \phi(\theta_1) = \phi(\theta_0)$.

Esta definição incorpora obviamente como caso particular a definição 2.2.ii) e mostra que qualquer partição com um único elemento (ou qualquer função constante) é trivialmente identificável.

A descoberta, em modelos não identificados, de uma partição não singular (respectivamente, função não constante) identificável permite definir novas partições (funções) identificáveis na base do seguinte teorema:

Teor. 2.1: i) Se Π_2 é identificável e $\Pi_1 < \Pi_2$, então Π_1 é também identificável.

ii) Se $\phi: E \rightarrow \Phi$ é identificável e $\rho: \Phi \rightarrow \Psi$, então a função $\psi = \rho \circ \phi: E \rightarrow \Psi$ é também identificável.

Dem: A aplicação da def. 2.4 i) e a transitividade da relação $<$ asseguram que $\Pi_1 < E/\sim$, o que demonstra i). A parte ii) é uma reformulação do enunciado i). Basta considerar que ϕ e ψ são funções definidoras de Π_2 e Π_1 , respectivamente, e atender ao facto de que $\Pi_1 < \Pi_2$ implica que ψ é uma função (necessariamente não injectiva, a não ser que Π_1 coincida com Π_2) de ϕ . Δ

Este teorema mostra assim que uma partição Π_1 não identificável não pode ser redução de uma partição Π_2 identificável, o que não significa, sublinhe-se, que Π_1 seja necessariamente estritamente mais fina do que Π_2 , ou equivalentemente, que Π_2 seja redução de Π_1 . De facto, as partições de E são apenas parcialmente ordenadas pela relação de redução já que esta não verifica a propriedade dicotómica. O seguinte exemplo artificial ilustra claramente essa afirmação:

Exemplo: Suponhamos $E = \{\theta_i, i=0,1,\dots,9\}$ e as seguintes partições

$$E/\sim = \{\{\theta_0, \theta_1, \theta_2\}, \{\theta_3, \theta_4\}, \{\theta_5, \theta_6\}, \{\theta_7, \theta_8, \theta_9\}\}$$

$$\Pi_0 = \{\{\theta_0, \theta_1\}, \{\theta_2\}, \{\theta_3, \theta_4\}, \{\theta_5, \theta_6\}, \{\theta_7\}, \{\theta_8, \theta_9\}\}$$

$$\Pi_1 = \{\{\theta_0, \theta_1\}, \{\theta_2\}, \{\theta_3, \theta_4\}, \{\theta_5, \theta_6, \theta_7, \theta_8, \theta_9\}\}$$

$$\Pi_2 = \{\{\theta_0, \theta_1, \theta_2, \theta_3, \theta_4\}, \{\theta_5, \theta_6\}, \{\theta_7, \theta_8, \theta_9\}\}$$

A partição Π_2 é claramente identificável enquanto que Π_0 e Π_1 são não identificáveis, porque não são redução de \mathbb{E}/\sim . Consequentemente, Π_0 e Π_1 também não são redução de Π_2 . Contudo, Π_2 é uma redução de Π_0 mas não é uma redução de Π_1 . Δ

Em termos de funções, o teorema 2.1 indica que uma função não identificável não pode ser função de uma função identificável. Em particular, dada uma função identificável $\phi(\theta)$, a não identificabilidade de θ implica a não injectividade da função $\phi(\cdot)$.

A definição de uma função identificável não implica que ela assuma valores diferentes em classes de equivalência distintas. Barankin (1960) designa toda a função que satisfaz esta última propriedade como suficiente no sentido em que ela é "suficiente" para identificar medidas idênticas em \mathcal{P} . No mesmo sentido, diremos que uma partição é suficiente quando cada classe de equivalência observacional é uma união de partes dessa partição. Em síntese:

Def. 2.5: i) A partição Π é suficiente se \mathbb{E}/\sim é uma redução de Π .

ii) A função $\tau: \mathbb{E} \rightarrow \mathcal{T}$ é suficiente se a partição por ela induzida é suficiente, isto é, se $\forall \theta_1, \theta_0 \in \mathbb{E}, \tau(\theta_1) = \tau(\theta_0) \Rightarrow \theta_1 \in [\theta_0]$.

Esta definição significa, por outras palavras, que todos os pontos situados no mesmo conjunto da partição, qualquer que ele seja, geram distribuições iguais (no caso dominado, funções densidade ou de probabilidade, $f(y|\theta)$, essencialmente iguais).

Em ordem a compreender, no contexto clássico, a razão de ser do epíteto suficiente na definição 2.5, façamos uma digressão pelo conceito de suficiência amostral em situações específicas do caso dominado. Mais concretamente, seja \mathcal{P} definida nos conjuntos mensuráveis de um subconjunto \mathcal{Y} de um espaço euclídeo, com \mathbb{E} igualmente euclídeo, dominada por uma medida σ -finita m e denotemos por $f(y|\theta)$ uma função densidade (ou de probabilidade) de \mathcal{P}_θ relativamente a m . Suponhamos ainda que para todo o $\theta \in \mathbb{E}, f(y|\theta) > 0, \forall y \in \mathcal{Y}$ e, se \mathcal{Y} é não numerável, que $f(y|\theta)$ é uma função contínua em $\mathcal{Y}, \forall \theta \in \mathbb{E}$.

Definamos agora em \mathcal{Y} a relação binária

$$y_1 * y_0 \iff f(y_1|\theta) = c(y_1, y_0) f(y_0|\theta), \forall \theta \in \mathbb{E} \quad (2.1)$$

onde $c(y_1, y_0) > 0$ é independente de θ . É fácil verificar que $*$ é uma relação de equivalência induzindo em \mathcal{Y} a partição $\mathcal{Y}/* = \{\pi_{y_0}\}$, onde π_{y_0} denota a classe de pontos amostrais y equivalentes a y_0 no sentido em que as respectivas funções de verosimilhança - denotadas por $l(\theta|y)$ e $l(\theta|y_0)$ - são proporcionais. Fixemos $\theta_0 \in \mathbb{E}$ e consideremos o logaritmo da razão das verosimilhanças em θ e θ_0

como uma função de \underline{y} indexada por $\underline{\theta}$. Denotando

$$T_{\underline{\theta}}(\underline{y}) = \log \frac{\lambda(\underline{\theta}|\underline{y})}{\lambda(\underline{\theta}_0|\underline{y})}, \quad \underline{\theta} \in \Xi \quad (2.2)$$

constata-se de imediato que a família de funções $\{T_{\underline{\theta}}: \underline{\theta} \in \Xi\}$ induz a partição $\mathcal{Y}/*$, ou seja, cada uma das partes desta pode ser equivalêntemente definida por

$$\pi_{\underline{y}_0} = \{\underline{y} \in \mathcal{Y} : T_{\underline{\theta}}(\underline{y}) = T_{\underline{\theta}}(\underline{y}_0), \forall \underline{\theta} \in \Xi\} \quad (2.3)$$

É sabido que $\mathcal{Y}/*$ é a partição suficiente minimal construída pelo conhecimento do método de Lehmann-Scheffé e, por conseguinte, toda a partição amostral Π de que $\mathcal{Y}/*$ seja uma redução é suficiente. Do mesmo modo, toda a partição Π que seja uma redução de $\mathcal{Y}/*$ é necessária, para usar a terminologia usual da teoria da suficiência clássica.

Voltando ao espaço paramétrico, a partição Ξ/\sim pode definir-se, nas condições assumidas (vide observação 2.1 a seguir), por

$$\pi_{\underline{\theta}_0} \equiv [\underline{\theta}_0] = \{\underline{\theta} \in \Xi : f(\underline{y}|\underline{\theta}) = f(\underline{y}|\underline{\theta}_0), \forall \underline{y} \in \mathcal{Y}\} \quad (2.4)$$

Fixemos $\underline{y}_0 \in \mathcal{Y}$ e definamos a família de funções paramétricas $\{\psi_{\underline{y}}: \underline{y} \in \mathcal{Y}\}$, onde

$$\psi_{\underline{y}}(\underline{\theta}) = \log \frac{f(\underline{y}|\underline{\theta})}{f(\underline{y}_0|\underline{\theta})} \quad (2.5)$$

Verifica-se facilmente que

$$\pi_{\underline{\theta}_0} = \{\underline{\theta} \in \Xi : \psi_{\underline{y}}(\underline{\theta}) = \psi_{\underline{y}}(\underline{\theta}_0), \forall \underline{y} \in \mathcal{Y}\} \quad (2.6)$$

A comparação dos conceitos de partição amostral suficiente (e necessária) e das expressões (2.2) e (2.3), por um lado, com a definição 2.5.i) (e 2.4.i)), (2.5) e (2.6) patenteia um estreito paralelismo que, nomeadamente, ajuda a perceber o porquê da designação de suficiência paramétrica adoptada por Barankin.

Este paralelismo vai poder ser aprofundado e generalizado no contexto de um modelo Bayesiano, como veremos na secção 2.2. Adiante-se, desde já, que na situação específica analisada e sob o uso de uma função densidade a priori, $q(\underline{\theta})$, estritamente positiva em Ξ , a definição de $\pi_{\underline{y}_0}$ com base em (2.1) é equivalente a afirmar, representando $q(\underline{\theta}|\underline{y})$ a função densidade a posteriori, que

$$\pi_{\underline{y}_0} = \{\underline{y} \in \mathcal{Y} : q(\underline{\theta}|\underline{y}) = q(\underline{\theta}|\underline{y}_0), \forall \underline{\theta} \in \Xi\} \quad (2.7)$$

A comparação de (2.7) com (2.4) revela a autêntica dualidade existente entre Ξ/\sim e $\mathcal{Y}/*$.

Estabelece-se, em seguida, uma caracterização das funções suficientes em

termos de uma factorização das medidas P_{θ} , que é, como veremos em 2.2.2, o análogo dual da caracterização Bayesiana das estatísticas suficientes.

Teor.2.2: A função $\tau : \Xi \rightarrow T$ é suficiente sse existe uma função $P_{\tau}^*(.)$ definida em $\tau(\Xi) \times \mathcal{A}$ tal que $P_{\theta}(A) = P_{\tau(\theta)}^*(A)$, $\forall A \in \mathcal{A}$, $\forall \theta \in \Xi$.

Dem. (Parte suficiente): Verificando-se a factorização referida, considere mos θ_1 e θ_0 tal que $\tau(\theta_1) = \tau(\theta_0)$. Segue-se então que $P_{\theta_1}(A) = P_{\theta_0}(A)$, $\forall A \in \mathcal{A}$, o que traduz pela definição 2.5 a suficiência de τ .

(Parte necessária): Supondo τ suficiente, então $P_{\theta_1}(A) = P_{\theta_0}(A)$, $\forall A \in \mathcal{A}$ para todo o par de pontos θ_1 e θ_0 de Ξ tal que $\tau(\theta_1) = \tau(\theta_0)$. Mas isto significa que a medida P_{θ} depende de θ apenas através de $\tau(\theta)$. Definindo para cada $A \in \mathcal{A}$ a função $P_{\tau}^*(A)$ como $P_{\tau(\theta)}^*(A) = P_{\theta}(A)$ para θ tal que $\tau(\theta) = \tau_0$, fica demonstrada a factorização referida. Δ

Observações:

2.1. - Esta caracterização, usada por Picci (1977) para definir uma função paramétrica suficiente, é uma generalização dos resultados para o caso dominado

registados, nomeadamente, em Barankin (1960). Fazendo $f(\underline{y}|\theta) = \frac{dP_{\theta}(\underline{y})}{dm(\underline{y})}[m]$ com m a medida σ -finita em (Y, \mathcal{A}) dominadora de \mathcal{P} , obtém-se a correspondente factorização de $f(\underline{y}|\theta)$ para todo o $\theta \in \Xi$ e quase todo o $\underline{y} \in Y$. No caso mais particular tratado por Barankin (1960), no qual Y é um espaço euclídeano, \mathcal{A} é a classe dos conjuntos Lebesgue-mensuráveis, m é a correspondente medida de Lebesgue e $f(\underline{y}|\theta)$ é contínua em Y para todo θ , a referida factorização aplica-se para todo o $(\underline{y}, \theta) \in Y \times \Xi$ devido a que $P_{\theta_1} = P_{\theta_0} \implies f(\underline{y}|\theta_1) = f(\underline{y}|\theta_0)$, $\forall \underline{y} \in Y$.

2.2. - Este teorema exprime que a função de distribuição, independentemente do conjunto de observações, depende de θ apenas através dos conjuntos de uma partição suficiente. Consequentemente, todos os valores de θ situados no elemento da partição a que pertence θ_0 são necessariamente equivalentes observacionalmente a θ_0 . Saliente-se que isto não significa que esse elemento da partição suficiente englobe todos os pontos de $[\theta_0]$. Repare-se na partição Π_0 do exemplo acima que representa, juntamente com Ξ/\sim , as únicas partições suficientes do conjunto de partições considerado. Aliás, basta pensar na partição $\{\{\theta\}, \theta \in \Xi\}$ que é sempre, e trivialmente, suficiente. Δ

Teor. 2.3: i) Seja Π_2 uma partição suficiente qualquer. Se a partição Π_1 é identificável, então Π_1 é uma redução de Π_2 .

ii) Seja $\tau: \Xi \rightarrow T$ uma função suficiente qualquer. Se a função $\phi: \Xi \rightarrow \Phi$ é identificável, então existe uma função $\rho: T \rightarrow \Phi$ tal que $\phi = \rho \tau$.

Dem.: O recurso às definições 2.4. i) e 2.5. i) e a transitividade da relação $<$ provam imediatamente a afirmação i). O enunciado ii) pode ser provado de i) considerando τ e ϕ funções indutoras respectivamente das partições Π_2 e Π_1 . Δ

Em virtude da propriedade anti-simétrica lata da relação de redução, Ξ/\sim constitui a única partição que é simultaneamente identificável e suficiente. Designá-la-emos, optando pela terminologia de Kadane (1974), por partição identificante. Ela é assim a partição suficiente que não admite nenhuma redução suficiente, como acontece com a partição amostral $Y/*$ e que, por isso, é denominada de partição suficiente minimal (necessária e suficiente).

Def. 2.6: A função $\psi: \Xi \rightarrow \Psi$ diz-se identificante se induz a partição identificante.

Toda função identificante está assim em correspondência biunívoca com qualquer outra função identificante. Desta forma, os elementos de Ξ/\sim podem ser definidos com base em qualquer função identificante ψ por $[\underline{\theta}_0] = \{\theta \in \Xi: \psi(\theta) = \psi(\underline{\theta}_0)\}$, $\theta_0 \in \Xi$. Um exemplo, válido para todos os casos, é dado pela função complexa formada a partir da função característica de $\underline{y} = (y_1, \dots, y_n)'$, $h(\underline{t} | \theta) = E[e^{i\underline{t}'\underline{y}} | \theta]$, fixando $\underline{t} = (t_1, \dots, t_n)' \in \mathbb{R}^n$, isto é, por $\psi_{\underline{t}}(\theta) = h(\underline{t} | \theta)$, em virtude da correspondência biunívoca existente entre as funções de distribuição e as funções características. O mesmo argumento pode ser aplicado à função geradora de momentos sempre que ela existir.

Uma função identificante é assim

- uma função suficiente que é função de toda a função suficiente, compreendendo-se desta forma a designação suficiente minimal usada por Barankin (1960);
- uma função identificável que corresponde à máxima discriminação de Ξ permitida pelos dados - daí a designação identificável maximal de Picci (1977).

Atendendo à definição 2.6, a junção dos teoremas 2.1 e 2.3 proporciona uma caracterização das funções identificáveis que permite justificar a designação identificante:

Teor. 2.4: Seja $\underline{\psi}: \Xi \rightarrow \Psi$ uma função identificante. A função $\underline{\phi}: \Xi \rightarrow \Phi$ é identificável sse existe uma função $\underline{\rho}: \Psi \rightarrow \Phi$ tal que $\underline{\phi} = \underline{\rho} \circ \underline{\psi}$.

Esta caracterização, estabelecida por Kadane (1974) através da aplicação dos conceitos de identificabilidade e de suficiência de funções paramétricas, revela a vantagem do uso destes conceitos em partições, já que a sua dedução nos parece mais clara e directa.

O teorema 2.4, ou equivalentemente, a definição de Ξ/\sim através de uma função $\underline{\psi} = \underline{\psi}(\underline{\theta})$ identificante, permite compreender que a não identificabilidade do próprio parâmetro indexante implica (e decorre de) a não injectividade da função $\underline{\psi}(\underline{\theta})$. Como o parâmetro $\underline{\psi}$ pode ser inequivocamente estimado, digamos que pelo valor $\underline{\psi}^*$, obtido por aplicação de algum critério de estimação, a inexistência de solução única do sistema em $\underline{\theta}$, $\underline{\psi}(\underline{\theta}) = \underline{\psi}^*$, materializa assim a não unicidade da determinação de quaisquer estimadores para $\underline{\theta}$. Fica pois provado que não tem sentido proceder à estimação de um parâmetro não identificável.

Se Ξ está restringido por $\underline{\lambda}(\underline{\theta}) = (\lambda_1(\underline{\theta}), \dots, \lambda_m(\underline{\theta}))' = \underline{0}_{(m)}$, a identificabilidade de $\underline{\theta}$ verifica-se se e só se a função $\underline{\psi}(\underline{\theta})$ com o seu domínio restringido a $\bar{\Xi}$, o espaço paramétrico restrito, é injectiva. Por outras palavras, se e só se $\underline{\theta}_0$ é a única solução do sistema

$$\begin{cases} \underline{\psi}(\underline{\theta}) = \underline{\psi}(\underline{\theta}_0) \\ \underline{\lambda}(\underline{\theta}) = \underline{0}_{(m)} \end{cases}$$

para todo o $\underline{\theta}_0 \in \bar{\Xi}$. O recurso a funções identificantes permite assim visualizar a averiguação da identificabilidade como um problema de solução de sistemas de equações, possivelmente não-lineares.

Frequentemente, Ξ é um subconjunto aberto de \mathbb{R}^k e as funções reais $\lambda_i(\underline{\theta})$, $i=1, \dots, m$ são continuamente diferenciáveis. Admitamos ainda que a função identificante $\underline{\psi} = (\psi_1 \dots \psi_q)'$ é tal que as funções reais $\psi_i(\underline{\theta})$, $i=1, \dots, q$ são também continuamente diferenciáveis. Estas condições permitem linearizar localmente as funções $\underline{\psi}(\underline{\theta})$ e $\underline{\lambda}(\underline{\theta})$ e deste modo, obter uma condição necessária e suficiente de identificabilidade local. Concentrando a nossa atenção num dado $\underline{\theta}_0$ e desenvolvendo as referidas funções em fórmula de Taylor linear em torno de $\underline{\theta}_0$, o sistema anterior é equivalente a

$$\underline{W}(\underline{\theta}^*)(\underline{\theta} - \underline{\theta}_0) = \underline{0}_{(m+q)}$$

onde $\underline{\theta}^* = t\underline{\theta} + (1-t)\underline{\theta}_0$, $0 < t < 1$ e $\underline{W}(\underline{\theta}^*)$ é a matriz jacobiana $(m+q) \times k$

$(\underline{W}_1'(\underline{\theta}) \underline{W}_2'(\underline{\theta}))'$, com $\underline{W}_1(\underline{\theta}) = \partial \underline{\psi} / \partial \underline{\theta}'$ e $\underline{W}_2(\underline{\theta}) = \partial \underline{\lambda} / \partial \underline{\theta}'$, avaliada em $\underline{\theta}^*$. Se a matriz dos coeficientes deste sistema homogêneo tem posto constante na vizi-

nhança de $\underline{\theta}_0$, então $\underline{\theta}_0$ é localmente identificável sse o posto de $\underline{W}(\underline{\theta}_0)$ é igual a k . A averiguação da identificabilidade global deixa de ser problemática no caso em que $\underline{\psi}(\cdot)$ e $\underline{\lambda}(\cdot)$ são lineares. Neste caso, a matriz jacobiana \underline{W} é constante e a análise padrão dos sistemas de equações lineares permite-nos concluir que $\underline{\theta}$ é identificável se e só se \underline{W} tem posto k . Deste modo, estes resultados, várias vezes usados em problemas de econometria (vide e.g., Rothenberg [1971], teors. 5 e 6), podem ser vistos como consequências do teorema 2.4 supra-citado.

2.1.2. Exemplos ilustrativos

A literatura de várias áreas aplicadas, como a econometria, está recheada de exemplos de modelos não identificáveis, alguns dos quais têm sido objeto de uso frequente para ilustração dos conceitos e resultados básicos do que denominamos teoria da identificabilidade. O nosso objectivo aqui está em proceder a essa ilustração através de modelos, por um lado, estruturalmente menos complexos do que muitos modelos econométricos e, por outro, mais familiares ao estatístico. Dedicaremos uma atenção especial aos modelos para dados categorizados, interligando os vários tipos de dados incompletos nas suas implicações em identificabilidade/não identificabilidade. Começaremos por focar o exemplo clássico do modelo linear normal pela sua notoriedade e pelo facto de permitir simplificar a discussão de muitos aspectos aplicáveis quer aos modelos estritamente lineares gerais, quer aos modelos funcionais lineares, como é o caso do modelo log-linear para dados categorizados.

EXEMPLO I: Consideremos o modelo linear univariado $\underline{y} = \underline{X}\underline{\beta} + \underline{u}$, onde \underline{X} é a matriz $n \times p$ de delinearmento com posto $r < p < n$, $\underline{\beta} = (\beta_1, \dots, \beta_p)'$ é um vector de parâmetros desconhecidos e \underline{u} , o vector aleatório não observável dos erros, tem distribuição normal de parâmetros $E(\underline{u}) = \underline{0}_{(n)}$ e $E(\underline{u}\underline{u}') = \sigma^2 \underline{I}_{(n)}$. Seja $\underline{\theta} = (\underline{\beta}'\sigma^2)' \in \Xi = \{\underline{\theta} \in \mathbb{R}^{p+1} : \beta_i \in \mathbb{R}, i=1, \dots, p, \sigma^2 \in \mathbb{R}^+\}$. O facto da distribuição de \underline{y} depender de $\underline{\theta}$ através de $\underline{X}\underline{\beta}$ e de σ^2 , conjugado com a hipótese de posto incompleto para \underline{X} , leva a que existam vários valores distintos de $\underline{\beta}$ compatíveis com o mesmo vector média de \underline{y} . Segue-se daí a não identificabilidade de $\underline{\beta}$ e, consequentemente, de $\underline{\theta}$. Deste modo, a não identificabilidade deste modelo reside na relação funcional entre o vector observado \underline{y} e o vector não observável \underline{u} , sendo independente da forma matemática da distribuição de \underline{u} .

Em ordem a obter uma ilustração mais detalhada, consideremos que o modelo acima corresponde à especificação condensada do modelo ANOVA II balanceado sem interacção, $y_{ijk} = \mu_{ij} + u_{ijk}$, $i=1, \dots, a$; $j=1, \dots, b$; $k=1, \dots, l$ ($a, b, l > 1$) com

$\mu_{ij} = \eta + \alpha_i + \gamma_j$. Assim, $\underline{X}\underline{\beta} = \underline{D}\underline{X}_0\underline{\beta} \equiv \underline{D}\underline{\mu}$ onde \underline{D} é a matriz $n \times q$ ($n=q\ell$, $q=ab$) diagonal em blocos, $\text{diag}(1_{\ell}, \dots, 1_{\ell})$, \underline{X}_0 é a matriz $q \times p$ ($p=a+b+1$) de 0's e 1's obtida de \underline{X} por omissão das linhas repetidas e $\underline{\beta} = (\eta \ \alpha_1 \dots \alpha_a \ \gamma_1 \dots \gamma_b)'$. O vector da média das celas, $\underline{\mu} = \underline{X}_0\underline{\beta}$, pertence assim ao subespaço gerado pelas colunas de \underline{X}_0 denotado por $M(\underline{X}_0)$. Como o posto de \underline{X}_0 é $r(\underline{X}_0) = r(\underline{X}) \equiv r$, então $M(\underline{X}_0)$ é um subespaço de \mathbb{R}^q de dimensão r , onde $r = a+b-1$, o que implica que o sistema de equações lineares $\underline{X}_0\underline{\beta} = \underline{\mu}$ para cada $\underline{\mu} \in \mathbb{R}^q$ seja indeterminado. Sejam, por exemplo,

$\underline{\beta}_0 = (\eta^0 \ \alpha_1^0 \dots \alpha_a^0 \ \gamma_1^0 \dots \gamma_b^0)'$ e $\underline{\beta}_1 = (\eta^1 \ \alpha_1^1 \dots \alpha_a^1 \ \gamma_1^1 \dots \gamma_b^1)'$ dois pontos distintos de \mathbb{R}^p tais que $\eta^1 = \eta^0 + c + d$, $\alpha_i^1 = \alpha_i^0 - c$, $i=1, \dots, a$ e $\gamma_j^1 = \gamma_j^0 - d$, $j=1, \dots, b$ ($c, d > 0$) e façamos $\underline{\theta}_0 = (\underline{\beta}_0' \ \sigma_0^2)'$, $\underline{\theta}_1 = (\underline{\beta}_1' \ \sigma_0^2)'$, $\sigma_0^2 > 0$. Ora, $\underline{\theta}_1 \neq \underline{\theta}_0$ mas $F(\underline{y} | \underline{\theta}_1) = F(\underline{y} | \underline{\theta}_0)$, $\forall \underline{y} \in \mathbb{R}^n$, pelo facto de $\underline{X}_0\underline{\beta}_1 = \underline{X}_0\underline{\beta}_0$. Consequentemente, Ξ é não identificável e a classe de pontos observacionalmente equivalentes a $\underline{\theta}_0$ pode ser definida por

$$[\underline{\theta}_0] = \{(\underline{\beta}_1', \sigma_1^2)' \in \Xi : \underline{X}_0\underline{\beta}_1 = \underline{X}_0\underline{\beta}_0, \sigma_1^2 = \sigma_0^2\} = \{(\underline{\beta}_1', \sigma_1^2)' \in \Xi : \underline{X}\underline{\beta}_1 = \underline{X}\underline{\beta}_0, \sigma_1^2 = \sigma_0^2\}.$$

A aplicação da definição 2.4 (e/ou do teorema 2.1) permite mostrar que são identificáveis as funções $(\underline{0}'_{(p)} \ 1)_{\underline{\theta}} = \sigma^2$, $(\underline{X} \ \underline{0}'_{(n)})_{\underline{\theta}} = \underline{X}\underline{\beta}$, $((\underline{X}\underline{\beta})' \ \sigma^2)'$, $\underline{X}'\underline{X}\underline{\beta}$, $((\underline{X}'\underline{X}\underline{\beta})' \ \sigma^2)'$, $\underline{a}'_1 \underline{\theta} = \alpha_i - \alpha_i$, (\underline{a}_1 é o vector $(p+1) \times 1$ com 1 e -1 nas linhas $i+1$ e $i'+1$ respectivamente e 0's nas restantes), $\underline{a}'_2 \underline{\theta} = \gamma_j - \gamma_j$, (\underline{a}_2 é o vector $(p+1) \times 1$ com 1 e -1 nas linhas $a+1+j$ e $a+1+j'$ resp. e 0's nas restantes) e $[(1 \ \underline{0}'_{(n+1)}) \underline{X}\underline{\beta}]^2 = (\eta + \alpha_1 + \gamma_1)^2$. De modo idêntico, pode constatar-se facilmente que, além de $\underline{\theta}$ e de $(\underline{I}'_p \ 0)_{\underline{\theta}} = \underline{\beta}$, as funções $\alpha_i + \alpha_i$, ($i \neq i'$), $\gamma_j + \gamma_j$, ($j \neq j'$), $\sum_{i=1}^a \alpha_i$ e $\sum_{j=1}^b \gamma_j$ constituem outros exemplos de funções não identificáveis.

Tomando em consideração a definição de Ξ/\sim , conclui-se adicionalmente que $((\underline{X}\underline{\beta})' \ \sigma^2)'$ e $((\underline{X}'\underline{X}\underline{\beta})' \ \sigma^2)'$ são funções identificantes enquanto que $\underline{X}\underline{\beta}$ e $\underline{X}'\underline{X}\underline{\beta}$ não são suficientes (sê-lo-ão se σ^2 for conhecido). Dentro das funções identificáveis indicadas acima, as funções $\sigma^2, \alpha_i - \alpha_i$, ($i \neq i'$), $\gamma_j - \gamma_j$, ($j \neq j'$) e $(\eta + \alpha_1 + \gamma_1)^2$ constituem novos exemplos de funções não suficientes. Já o próprio $\underline{\theta}$ e qualquer função bijectiva sua são exemplos de funções suficientes triviais enquanto que $(\sum_{i=1}^a \alpha_i \ \sigma^2)$ constitui um exemplo de uma função não suficiente, dentro do conjunto de funções não identificáveis mencionado anteriormente.

Vimos que a não identificabilidade de $\underline{\beta}$ deriva da sobreparametrização que a formulação em "equações livres" do modelo propicia. Como a função $\underline{\mu} = \underline{X}_0\underline{\beta}$ é identificável, torna-se então possível obter uma nova formulação do modelo sem vestígios de não identificabilidade. Designando por \underline{W}' qualquer matriz $(q-r) \times q$ de posto completo tal que $\underline{W}'\underline{X}_0 = \underline{0}_{(q-r, p)}$, a identidade dos subespaços $M(\underline{X}_0)$ e $N(\underline{W}')$ de \mathbb{R}^q (proposição A.2 do apêndice) permite-nos definir alternativamente o modelo linear acima como $\underline{y} = \underline{D}\underline{\mu} + \underline{u}$ sujeito a $\underline{W}'\underline{\mu} = \underline{0}_{(q-r)}$. Esta formulação em termos de restrições é assim mais vantajosa quando o interesse se centra na média das celas e não nos parâmetros $\underline{\beta}$.

Contudo, a identificabilidade do modelo restringido da média das celas, inevitável em delineamentos com pelo menos uma observação em cada cela, pode desaparecer em delineamentos não balanceados com algumas celas não observadas. Com efeito, suponhamos uma experiência factorial d -dimensional, onde o factor j tem a_j níveis, $j=1, \dots, d$, delineada de modo a existirem n_i réplicas em cada cela i , $i=1, \dots, q$ ($q = \prod_{j=1}^d a_j$), sendo $n_i = 0$ para algumas delas. O valor médio do vector $n \times 1$ ($n = \sum_{i=1}^q n_i$) das observações, \underline{y} , pode ser ainda representado por $\underline{D}\underline{\mu}$, onde \underline{D} é agora a matriz $n \times q$ cuja coluna i é o vector $\underline{0}_{(n)}$ se a cela i não é observada e igual ao vector obtido deste por substituição dos elementos das linhas correspondentes a $y_{i1}, y_{i2}, \dots, y_{in_i}$ por 1, no caso contrário. Designando por \underline{D}^* a matriz $(q-m) \times q$, onde m é o número de celas não observadas, obtida de \underline{D} por remoção das linhas repetidas, o vector das médias das celas observadas é então $\underline{D}^* \underline{\mu}$. Torna-se então evidente que, contrariamente ao que se passa quando $n_i > 0$, $i=1, \dots, q$, o vector $\underline{\mu}$ não é identificável no modelo não restringido, já que o parâmetro indexante da distribuição é apenas o seu subvector componente $\underline{D}^* \underline{\mu}$, para além de σ^2 . Se postularmos as restrições $\underline{\mu} \in N(\underline{W}')$ onde \underline{W} é uma matriz $q \times (q-r)$ de posto completo, o respectivo modelo pode ser redefinido, por redução da dimensionalidade paramétrica, em termos de equações livres num subconjunto dos μ_i 's, conforme o procedimento usado por Murray and Smith (1985). Para o efeito, consideremos $\underline{W}' = (\underline{W}'_1 \ \underline{W}'_2)$, onde \underline{W}'_2 é uma matriz não singular de ordem $q-r$, e denotemos as partições induzidas respectivamente em $\underline{\mu}$ e \underline{D} por $(\underline{\mu}'_1 \ \underline{\mu}'_2)'$ e $(\underline{D}_1 \ \underline{D}_2)$. Incorporando as restrições $\underline{W}' \underline{\mu} = \underline{0}_{(q-r)}$, equivalentes a $\underline{\mu}'_2 = -\underline{W}'_2^{-1} \underline{W}'_1 \underline{\mu}'_1$ na relação entre $E(\underline{y})$ e $\underline{\mu}$ obtemos o modelo reduzido $\underline{Y} = \underline{V} \underline{\mu}'_1 + \underline{u}$, onde \underline{V} é a matriz $n \times r$ definida por $\underline{D}_1 - \underline{D}_2 \underline{W}'_2^{-1} \underline{W}'_1$. Esta conversão para a formulação em modelo linear não restringido não depende, assim, do número nem da posição das celas não observadas. Contudo, a identificabilidade do modelo já depende do delineamento. Basta notar que o modelo em $\underline{\mu}'_1$ está na forma do modelo de efeitos β com \underline{V} desempenhando o papel de \underline{X} . Como a condição de \underline{V} ser de posto completo é necessária e suficiente para a identificabilidade de $\underline{\mu}'_1$ conclui-se, pela relação entre $\underline{\mu}$ e o seu vector componente $\underline{\mu}'_1$, $\underline{\mu} = (\underline{I}_{(r)} \ -\underline{W}'_1 \underline{W}'_2^{-1})' \underline{\mu}'_1$, que $\underline{\mu}$ é identificável sse \underline{V} é de posto completo. Desta forma, em situações deste tipo em que há uma ausência completa de informação amostral para certos elementos de $\underline{\mu}$ (modelos ANOVA com celas "missing" ou modelos de delineamento em blocos incompletos, por exemplo) a identificabilidade de $\underline{\mu}$ depende não só do modelo postulado mas também do próprio delineamento.

Esta conclusão é obviamente independente da formulação adoptada para $\underline{\mu}$. Se \underline{X}_0 for uma matriz $q \times p$ de posto $r \leq p < q$ geradora do subespaço $N(\underline{W}')$

a que pertence μ , o modelo pode formular-se equivalentemente por $E(\underline{y}) = \underline{X}\beta$, onde $\underline{X} = \underline{D}\underline{X}_0$, sendo \underline{D} a matriz acima definida. Como $r(\underline{X}) = r(\underline{X}^*)$ onde $\underline{X}^* = \underline{D}^*\underline{X}_0$, segue-se que $r(\underline{X})$ pode ser diferente de $r(\underline{X}_0)$ contrariamente ao que se passa quando todas as celas são observadas. Se \underline{D}^* é tal que $r(\underline{X}^*) = r$ (o que exige $q-m \geq r$) podem ser identificados r μ_i 's e através destes e das $q-r$ restrições que o modelo define, os restantes elementos de μ .

Estas considerações sobre a formulação em termos de β do modelo linear têm igualmente as suas implicações na identificabilidade de β , devido à possibilidade do subespaço imagem de $M(\underline{X}_0)$ por \underline{D}^* ter dimensão inferior à de $M(\underline{X}_0)$. Por outras palavras, a presença de celas não observadas pode fazer com que a identificabilidade do modelo $M(\underline{X}_0)$ em \mathbb{R}^q , com $r=p$, deixe de existir no modelo correspondente em \mathbb{R}^{q-m} , $M(\underline{X}^*)$, a imagem de $M(\underline{X}_0)$ pela transformação linear \underline{D}^* . Esta situação surge sempre que $q-m < p$ mas pode igualmente ocorrer quando o número de parâmetros não excede o número de celas observadas. De facto, a remoção de linhas de \underline{X}_0 operada por \underline{D}^* pode alterar o seu posto e, sendo assim, cai-se numa situação análoga à considerada anteriormente para dados completos, quando se definiram modelos de efeitos com posto incompleto.

Finalmente, queremos deixar claro que as considerações anteriores não se confinam a modelos de análise de variância e são naturalmente generalizáveis a modelos lineares multivariados. O modelo de regressão linear univariada com multicolinearidade extrema é um exemplo típico de um modelo não identificável que se encaixa no esquema genérico definido no início do exemplo. Basta considerar que β representa um vector de coeficientes de regressão de \underline{y} em p variáveis explicativas e que a matriz \underline{X} dos valores destas variáveis tem $p-r$ dependências lineares exactas entre as suas colunas.

EXEMPLO II. A discussão anterior em torno dos conceitos de 2.1.1. pode ser aplicada, na sua essência, ao modelo log-linear para dados categorizados. Denotando agora \underline{y} o vector das frequências nas q celas duma tabela de contingência, de média $\underline{\mu} \in (\mathbb{R}^+)^q$, o modelo log-linear ordinário é definido por $\log \underline{y} = \underline{X}_0 \beta$, onde $\beta \in \mathbb{R}^p$ é um vector de parâmetros desconhecidos e a matriz \underline{X}_0 de especificação do modelo, com posto suposto igual a $r < p \leq q$, é tal que $M(\underline{X}_0)$ inclui os vectores indicadores da eventual partição do conjunto Q das q celas operada pelo delineamento experimental.

No caso de \underline{y} representar o registo simultâneo de q processos de Poisson independentes, por exemplo, durante um período de tempo fixo, o modelo probabilístico fica definido pela função de probabilidade

$$f(\underline{y} | \underline{\mu}) = \left(\prod_{i \in Q} (y_i!)^{-1} \right) \exp \{ \underline{y}' \cdot \log \underline{\mu} - \sum_{i \in Q} \mu_i \}, \quad \log \underline{\mu} \in M(\underline{X}_0) \quad (2.8)$$

Suponhamos agora que Q é particionado no conjunto $\{Q_k, k=1, \dots, s\}$ com os

Q_k 's disjuntos e exaustivos e onde Q_k é constituído por q_k celas. Seja $C = (c_1, \dots, c_s)$ a matriz $q \times s$ tal que c_k representa o vector indicador das celas de Q_k , $k=1, \dots, s$, de forma que a experiência é definida condicionalmente a $\{c_k' y = N_k; k=1, \dots, s\}$, onde y tem a distribuição supra-mencionada. Denotando por $\{y_k, k=1, \dots, s\}$ e $\{\mu_k, k=1, \dots, s\}$ as partições de y e μ , respectivamente, é sobejamente conhecido que a distribuição resultante é um produto de multinomiais independentes, podendo ser expressa por

$$f(y | \{N_k, v_k\}) = \left(\prod_{k=1}^s \frac{N_k!}{\prod_{i \in Q_k} y_i!} \right) \exp \left\{ \sum_{k=1}^s y_k' \log v_k - \sum_{k=1}^s N_k \log N_k \right\} \quad (2.9)$$

onde $v_k = (N_k / c_k' \mu) \mu_k$ é o vector média condicional de y_k . Como

$$\log v \equiv \log(v_1' \dots v_s')' = \log \mu + \sum_{k=1}^s \left(\log \left(\frac{N_k}{c_k' \mu} \right) \right) c_k$$

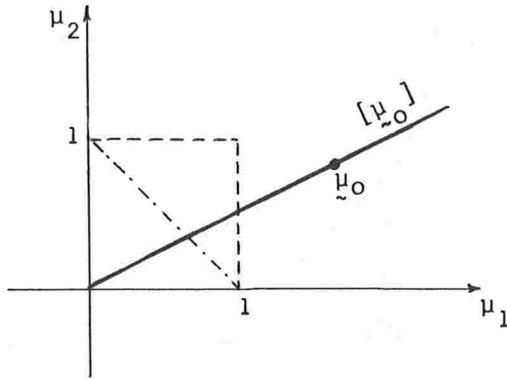
, fica claro que o mesmo modelo log-linear $M(X_0)$ pode ser aplicado a v , uma vez que $M(C) \subset M(X_0)$. A única diferença reside no facto de que a gama de valores possíveis para $\log v$ é o conjunto $M(X_0)$ restringido por $c_k' v = N_k$, $k=1, \dots, s$.

Em qualquer dos casos, a distribuição de y depende de β através da função $X_0 \beta$, o que atesta a não identificabilidade do modelo log-linear tal como foi definido, bem como a natureza identificante das funções $X_0 \beta$ e $\exp(X_0 \beta)$. As implicações inferenciais do problema da identificabilidade podem ser aqui menos relevantes do que no modelo linear normal. De facto, na abordagem por máxima verosimilhança dos modelos log-lineares hierárquicos, construídos similarmente aos modelos ANOVA, o interesse centra-se mais nas funções identificáveis μ ou v (ou nas probabilidades $v_k | N_k$, $k=1, \dots, s$) do que no próprio β . Desta forma, o modelo log-linear tradicional para dados categorizados completos é um exemplo característico de um problema de identificabilidade que não bloqueia certas inferências sobre os parâmetros de interesse. Não obstante, o uso da abordagem via mínimos quadrados ponderados já suscita os mesmos problemas que surgem na análise do modelo linear normal, quando não restringida a funções identificáveis.

EXEMPLO III: Mesmo sem a introdução de restrições log-lineares, o modelo (2.9) encobria uma situação de não identificabilidade que foi erradicada pela reparametrização da distribuição de μ para v . Para uma maior simplicidade na ilustração, vamos considerar o caso em que $q=2$ e $s=1$. Designemos agora por u o vector $(y_1 \ y_2)'$ de média $(\mu_1 \ \mu_2)'$ e por n_1 a v.a y_1 quando $1_2' u = N$. No modelo (2.8), definido por $y_i | \mu_i \rightsquigarrow \text{Poi}(\mu_i)$, $i=1,2$, $y_1 \perp\!\!\!\perp y_2 | \mu_1, \mu_2$, o espaço paramétrico $U = \{\mu = (\mu_1, \mu_2): \mu_i \in \mathbb{R}^+, i=1,2\}$ é claramente identificável. Tal situação já não ocorre no modelo correspondente a (2.9) definido por $n_1 | N_1, \mu \rightsquigarrow \text{Bi}(N, \mu_1 / (\mu_1 + \mu_2))$. De facto, é fácil constatar que, sendo $\mu_0 = (\mu_1^0 \ \mu_2^0)' \in U$,

$$[\underline{\mu}_0] = \left\{ \underline{\mu}^* \in U : \frac{\mu_1^*}{\mu_1^* + \mu_2^*} = \frac{\mu_1^0}{\mu_1^0 + \mu_2^0} \right\} = \{ \underline{\mu}^* \in U : \mu_1^* / \mu_2^* = \mu_1^0 / \mu_2^0 \} \neq \{ \underline{\mu}_0 \}.$$

Deste modo, U/\sim será o conjunto das semi-rectas do 1º quadrante emanando da origem e, portanto, cada classe de equivalência poderá ser indexada pelo declive da respectiva semi-recta (vide figura).



A definição da variável aleatória observada implica que o seu contradomínio seja definido no subconjunto de \mathbb{N}_0^2 (o contradomínio de \underline{u}) definido por $X = \{ \underline{u} \in \mathbb{N}_0^2 : \sum_{i=1}^2 u_i = N \}$, isto é, os valores de n_1 são os elementos da secção de X em y_1 , $X_1 = \{ y_1 : (y_1, y_2) \in X \} = \{ 0, 1, \dots, N \}$.

É esta relação entre n_1 e o vector gerador \underline{u} , juntamente com a distribuição assumida para este, que estão na base da não identificabilidade do parâmetro $\underline{\mu}$ no modelo probabilístico resultante. Esta situação constitui um exemplo onde a indistinguibilidade na experiência conjunta - no caso vertente, em

$(\mathbb{N}_0^2, P(\mathbb{N}_0^2), \{ P_{\underline{\mu}} : \underline{\mu} \in (\mathbb{R}^+)^2 \})$, onde $P(\mathbb{N}_0^2)$ designa o conjunto das partes de \mathbb{N}_0^2 e $P_{\underline{\mu}}$ é tal que a sua derivada de Radon-Nikodym é

$$\frac{dP_{\underline{\mu}}(\underline{u})}{dm(\underline{u})} = \prod_{i=1,2} (e^{-\mu_i} \mu_i^{u_i} / u_i!) [m] \text{ com } m \text{ a medida de contagem no espaço mensurável indicado - deixa de existir na experiência condicionada definida por}$$

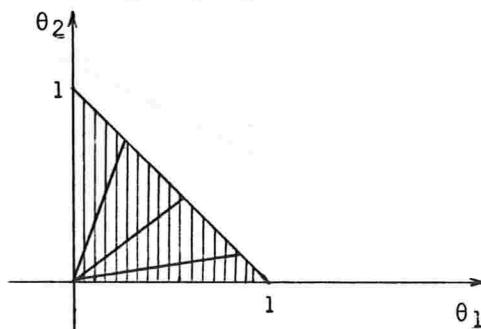
$(X_1, P(X_1), \{ P_{\underline{\mu}}^* : \underline{\mu} \in (\mathbb{R}^+)^2 \})$, onde $\frac{dP_{\underline{\mu}}^*(n_1)}{dm^*(n_1)} = \binom{N}{n_1} (\mu_1 / (\mu_1 + \mu_2))^{n_1} (\mu_2 / (\mu_1 + \mu_2))^{N-n_1} [m^*]$ com m^* a medida de contagem em $(X_1, P(X_1))$.

A definição de U/\sim permite concluir imediatamente que a função

$$\theta(\underline{\mu}) = \left(\frac{\mu_1}{\mu_1 + \mu_2} \frac{\mu_2}{\mu_1 + \mu_2} \right)' = \exp[\underline{A} \log(\underline{B}\underline{\mu})] \text{ onde } \underline{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \text{ e } \underline{B}' = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ é}$$

identificante. Em particular, a representação gráfica do seu contradomínio (o segmento de recta dentro do quadrado unitário da figura anterior) permite visualizar a singularização de cada classe de equivalência. É uma reparametrização deste tipo eliminadora da não identificabilidade que a expressão (2.9) materializa. Torna-se agora fácil arranjar novos exemplos de funções identificáveis, e.g., $E(n_1/N, \underline{\mu})$ e $\text{Var}(n_1/N, \underline{\mu})$, das quais só a primeira é suficiente (aplique-se a def. 2.5 ou o teor. 2.2).

Uma situação do mesmo gênero aparece em modelos multinomiais condicionados reflectindo a truncatura ou o não registo de certas frequências da tabela subjacente. Por exemplo, seja $\underline{u} = (y_1 \ y_2 \ y_3)'$ um vector com distribuição trinomial de parâmetros $N^* = \underline{1}'_3 \underline{u}$ e $\underline{\theta} = (\theta_1 \ \theta_2 \ \theta_3)'$ com $\theta_i > 0$ e $\underline{1}'_3 \underline{\theta} = 1$ e denotemos a variável aleatória y_1 quando $y_1 + y_2 = N$ por n_1 . É sobejamente conhecido que a distribuição de n_1 , que corresponde à distribuição de \underline{u} quando a frequência y_3 é "missing", é binomial de parâmetros N e $\theta_1 / (\theta_1 + \theta_2)$. O conjunto $E = \{(\theta_1, \theta_2) \in \mathbb{R}^2: 0 < \theta_1, \theta_2 < 1, \theta_1 + \theta_2 < 1\}$ é assim decomposto em infinitas classes de equivalência observacional representadas na figura ao lado pelos segmentos de recta emanando da origem contidos em E indicado a tracejado.



O problema da identificabilidade pode ser eliminado através da introdução de restrições no espaço paramétrico.

Consideremos, por exemplo, uma tabela 2×2 com vector de contagens $\underline{u} = (y_{ij}, i, j = 1, 2)$ e seja \underline{n} o vector representando as contagens observadas $\{y_{ij}, (i, j) \in S\}$ quando $\sum_{(i,j) \in S} y_{ij} = N$. Sob a hipótese de independência na tabela completa, é fácil

constatar a identificabilidade de $\underline{\lambda} = (\theta_{1.} \ \theta_{2.} \ \theta_{.1} \ \theta_{.2})'$ e, conseqüentemente, de $\underline{\theta} = (\theta_{ij}, i, j = 1, 2)'$ quando $S = \{(1,1), (1,2), (2,1)\}$. Contudo, a restrição do espaço paramétrico pode, por si só, não ser suficiente para a identificação. É o que acontece no caso anterior se, por exemplo $S = \{(1,1), (2,1)\}$ pois a função de probabilidade de \underline{n} não depende dos parâmetros $\theta_{.1}$ e $\theta_{.2}$. Esta constatação, agora num contexto de dados categorizados, é mais uma manifestação concreta de que a identificabilidade do parâmetro indexante da distribuição dos dados completos em problemas com observações "missing" depende do número e da localização destas observações, para além das próprias restrições impostas no espaço paramétrico.

Em alguns casos, o exame do problema leva a considerar certas celas como tendo probabilidades conhecidas a priori e que, por isso, são excluídas da análise. Esta situação encaixa-se no esquema anterior com a particularidade da soma das probabilidades relativas às celas observadas ser conhecida de antemão. Deste modo, fica assegurada a identificabilidade do parâmetro indexante do modelo probabilístico que descreve as chamadas tabelas incompletas com zeros estruturais. No entanto, podem surgir problemas de identificabilidade em modelos visando descrever a estrutura dessas tabelas, como nos chamados modelos quase-log-lineares. Retomando a notação do exemplo II, suponhamos que $\underline{y}^* = D^* \underline{y}$ representa o vector das frequências da tabela incompleta de dimensão

$q^* = q - m$, onde y tem a distribuição (2.9) com $s=1$, $N_1=N$, e seja $\underline{\theta} = \log y$. Denotando por S o conjunto das celas observadas e representando $\underline{\theta}^* = \{\theta_{ij}^* : i \in S\}$ o vector dos logaritmos dos valores esperados das celas da tabela incompleta, o modelo quási-log-linear correspondente ao modelo log-linear $M(\underline{X}_0)$ para $\underline{\theta}$ define-se fazendo $\underline{\theta}^* = \underline{D}^* \underline{\theta}$ para $\underline{\theta} = \underline{X}_0 \underline{\beta}$, ou seja, $\underline{\theta}^* = \underline{X}^* \underline{\beta}$, onde $\underline{X}^* = \underline{D}^* \underline{X}_0$ e $\underline{\beta} \in \mathbb{R}^p$, $p \leq q$.

À semelhança do que foi visto para os modelos lineares de análise de variância, o facto de $\underline{\theta}^*$ pertencer à imagem de $M(\underline{X}_0)$ pela transformação linear \underline{D}^* pode conduzir à não identificabilidade de $\underline{\beta}$ sob o modelo quási-log-linear, ainda que adoptemos uma formulação log-linear original com $\dim M(\underline{X}_0) = p$.

Por exemplo, consideremos a tabela $2^2 \times 3$ sob o modelo de quasi-independência,

$$\theta_{ijk}^* = \theta_{ijk} = \eta + \alpha_i + \gamma_j + \delta_k, \quad (i,j,k) \in S \text{ onde}$$

$$S = \{(1,1,1), (1,2,2), (2,1,2), (2,2,3)\} \text{ e } \sum_{i=1}^2 \alpha_i = \sum_{j=1}^2 \gamma_j = \sum_{k=1}^3 \delta_k = 0. \text{ O subespaço}$$

$M(\underline{X}_0)$ de \mathbb{R}^{12} tem dimensão $p=5$. Fazendo $\underline{\beta} = (\eta \ \alpha_1 \ \gamma_1 \ \delta_1 \ \delta_2)'$, então

$\underline{\theta}^* = \underline{X}^* \underline{\beta}$, onde \underline{X}^* de dimensão (4×5) tem posto 4. O facto da função de probabilidade de y^* depender de $\underline{\beta}$ através de $\underline{\theta}^*$ aliado a que o número de observações é inferior à dimensão de $\underline{\beta}$ evidencia a não identificabilidade do parâmetro do modelo quási-log-linear considerado. A indeterminação do sistema linear em $\underline{\beta}$ para um dado $\underline{\theta}^*$ pode ainda verificar-se quando $q^* \geq p$. Por exemplo, na tabela 2×3^2 sob o modelo de quasi-independência $\theta_{ijk}^* = \theta_{ijk} = \eta + \alpha_i + \gamma_j + \delta_k, (i,j,k) \in S$

com $S = \{(1,1,1), (1,2,2), (1,3,2), (2,1,2), (2,2,3), (2,3,3)\}$, $\sum_{i=1}^2 \alpha_i = \sum_{j=1}^3 \gamma_j = \sum_{k=1}^3 \delta_k = 0$, temos 6 observações e 6 parâmetros funcionalmente independentes. Definindo a matriz 18×6 de posto completo, \underline{X}_0 , em correspondência com

$\underline{\beta} = (\eta \ \alpha_1 \ \gamma_1 \ \gamma_2 \ \delta_1 \ \delta_2)'$, o modelo referido fica definido por $\underline{\theta}^* = \underline{X}^* \underline{\beta}$ onde

$$\underline{X}^* = (\underline{b}_1 \ \dots \ \underline{b}_6) \text{ com } \underline{b}_1 = \underline{1}_6, \ \underline{b}_2 = \begin{pmatrix} 1 & -1 \\ 3 & -3 \end{pmatrix}', \ \underline{b}_3 = (1 \ 0 \ -1 \ 1 \ 0 \ -1)',$$

$\underline{b}_4 = (0 \ 1 \ -1 \ 0 \ 1 \ -1)', \ \underline{b}_5 = (1 \ 0 \ 0 \ 0 \ -1 \ -1)' \text{ e } \underline{b}_6 = (0 \ 1 \ 1 \ 1 \ -1 \ -1)'$. É fácil constatar que $M(\underline{X}^*)$ é um subespaço de \mathbb{R}^6 de dimensão 5, pelo que caímos numa situação similar à dos modelos log-lineares para dados completos considerados no exemplo II.

Em suma, neste tipo de tabelas truncadas, a identificabilidade do modelo paramétrico postulado está dependente da configuração dos zeros estruturais e uma dada configuração pode implicar a identificabilidade de certos modelos e a não identificabilidade de outros.

EXEMPLO IV: Com as ilustrações anteriores ficou já patente a íntima ligação existente entre dados incompletos e problemas de identificabilidade nos parâmetros relativos aos dados completos. A natureza incompleta dos dados materializou-se na ausência total de informação amostral em certas celas determi-

nada por delineamento experimental (exemplo I) ou por truncatura (exemplo III). No contexto de dados categorizados é ainda frequente surgirem situações com ocorrência de uma perda parcial de informação amostral em certas celas e que não é necessariamente determinada pelo delineamento da experiência. A genética, por exemplo, é um campo fértil para a ocorrência de dados parcialmente categorizados em consequência da natureza do próprio fenômeno observável.

Um exemplo largamente citado na literatura é o do grupo sanguíneo ABO, no qual se pretende indagar a distribuição conjunta dos genes A, B e O dos progenitores dos indivíduos de uma dada população. Denotando por \underline{u} o vector das frequências genotípicas da tabela 3² correspondente a uma amostra de N indivíduos, o modelo probabilístico geralmente considerado estabelece uma distribuição multinomial para \underline{u} (u_{ij} , $i, j = A, B, O$) de parâmetros N e $\underline{\theta} = (\theta_{ij})$ com $\sum_{ij} \theta_{ij} = 1$, onde θ_{ij} , suposto positivo para todo o (i,j), representa a probabilidade de qualquer indivíduo herdar do pai e da mãe genes com alelos i e j respectivamente. Contudo, o processo observacional permite apenas determinar a classe fenotípica

		gene da mãe		
		A	B	O
gene do pai	A	u_{11}	u_{12}	u_{13}
	B	u_{21}	u_{22}	u_{23}
	O	u_{31}	u_{32}	u_{33}

(A, B, AB e O) a que cada indivíduo pertence. Ou seja, o vector observado é $\underline{y} = (y_A \ y_B \ y_{AB} \ y_O)'$ onde $y_A = u_{11} + u_{13} + u_{31}$, $y_{AB} = u_{12} + u_{21}$, $y_B = u_{22} + u_{23} + u_{32}$ e $y_O = u_{33} = N - (y_A + y_B + y_{AB})$, o que pode ser representado compactamente por $\underline{y} = D\underline{u}$, com

$$D = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Representando $\underline{\psi} = \underline{\psi}(\underline{\theta}) = D\underline{\theta} = (\psi_A \ \psi_B \ \psi_{AB} \ \psi_O)'$ o vector das probabilidades fenotípicas, a distribuição observada é caracterizada por

$$f(\underline{y} | N, \underline{\theta}) = N! \frac{\psi_A^{y_A}}{y_A!} \frac{\psi_B^{y_B}}{y_B!} \frac{\psi_{AB}^{y_{AB}}}{y_{AB}!} \frac{\psi_O^{y_O}}{y_O!}, \quad \begin{matrix} \sum_{i=1}^4 y_i = N \\ \sum_{i=1}^4 \psi_i = 1 \end{matrix}$$

Tal como acontece para o domínio de $\underline{\beta}$ no modelo linear normal, o espaço paramétrico $E = \{\underline{\theta} \in \mathbb{R}^9: 0 < \theta_{ij} < 1, \sum_{ij} \theta_{ij} = 1\}$ é particionado em classes de equivalência não singulares, sendo a classe observacionalmente equivalente a um dado $\underline{\theta}_0 \in E$ definida por

$$[\underline{\theta}_0] = \{\underline{\theta}_* \in E: D(\underline{\theta}_* - \underline{\theta}_0) = \underline{0}_{(4)}\} = \{\underline{\theta}_* \in E: \underline{\theta}_* = \underline{\theta}_0 + \underline{\alpha}, \underline{\alpha} \in N(D)\}$$

onde o espaço nulo de D é um subespaço pentadimensional de \mathbb{R}^9 . Como implicação temos a não identificabilidade de $\underline{\theta}$ e a natureza identificante de $\underline{\psi}$.

Neste problema genético, é usual assumir-se à partida a simetria marginal, isto é, que a distribuição dos alelos em ambos os progenitores é a mesma. Uma das questões de maior interesse é averiguar então se a população se encontra em equilíbrio de Hardy-Weinberg, o que em termos estatísticos se traduz pela hipótese de independência. Em conjunto, estas duas hipóteses exprimem-se em termos multiplicativos por $H: \theta_{ij} = \theta_i \cdot \theta_j$, $i, j = A, B, O$, onde θ_k representa a probabilidade do alelo $k = A, B, O$ em qualquer dos progenitores. Da relação entre ψ e $\lambda = (\theta_A \ \theta_B \ \theta_O)'$

$$\begin{aligned} \psi_A &= 2\theta_A - \theta_A^2 - 2\theta_A \theta_B ; \quad \psi_{AB} = 2\theta_A \theta_B \\ \psi_B &= 2\theta_B - \theta_B^2 - 2\theta_A \theta_B ; \quad \psi_O = \theta_O^2 = (1 - \theta_A - \theta_B)^2 \end{aligned}$$

é fácil verificar que

$$\begin{aligned} \theta_A &= 1 - (1 - c'_A \psi)^{1/2} ; \quad c'_A = (1 \ 0 \ 1 \ 0) \\ \theta_B &= 1 - (1 - c'_B \psi)^{1/2} ; \quad c'_B = (0 \ 1 \ 1 \ 0) \end{aligned}$$

o que evidencia, por aplicação repetida do teorema 2.1, que λ e consequentemente θ são parâmetros identificáveis sob o modelo restrito H.

Neste exemplo concreto, a informação "missing" resulta da "impossibilidade" de distinguir todas as celas da tabela de contingência original. Contudo, as frequências aglutinadas podem surgir noutras situações como resultado de um processo observacional e/ou de registo deficiente explicável pela intervenção de factores aleatórios (vide cap. 3). Por exemplo, suponhamos que ao pretender-se classificar N indivíduos pelas 3×4 categorias de duas variáveis foram registadas apenas as seguintes frequências:

$$\begin{aligned} y_a &= u_{11} + u_{12} + u_{21} + u_{22} ; \quad y_b = u_{13} + u_{14} + u_{23} + u_{24} ; \\ y_{3j} &= u_{3j} , \quad j=1,2,3,4 \end{aligned}$$

onde \underline{u} designa o vector não totalmente observado das 12 frequências originais. Designando \underline{y} o vector $(y_a \ y_b \ y_{31} \ y_{32} \ y_{33} \ y_{34})'$ podemos escrever $\underline{y} = \underline{D}\underline{u}$ onde \underline{D} é a matriz reflectindo a ligação entre as frequências observadas e não observadas. Supondo o modelo multinomial $M_{12}(N, \underline{\theta})$ com $\sum_{12} \theta = 1$ para \underline{u} , admitamos que o processo conducente à aglutinação de \underline{u} é tal que a distribuição observada pode ser caracterizada pelo modelo $M_6(N, \underline{D}\underline{\theta})$. Tal como no exemplo genético, os parâmetros observacionalmente equivalentes do simplex probabilístico de \mathbb{R}^{12} estão compreendidos no conjunto de hiperplanos paralelos ao subespaço hexadimensional de \mathbb{R}^{12} , $\underline{N}(\underline{D})$. Analogamente, a não identificabilidade de $\underline{\theta}$ torna impossível obter estimativas razoáveis dos θ_{ij} confundidos no contexto do modelo saturado. Ainda do mesmo modo, a estimação sob o modelo de independência H_I , correspondente a 6 graus de liberdade, torna-se em princípio possível, já que a ocorrência da tabela incompleta acarretou a perda de apenas 6 graus de liberdade.

Todavia, fazendo $\underline{\lambda} = (\theta_{1.} \theta_{2.} \theta_{3.} \theta_{.1} \theta_{.2} \theta_{.3} \theta_{.4})'$, a expressão da função identificante $\underline{\psi} = \underline{D}\underline{\theta}$ em termos de $\underline{\lambda}$ leva a que não possamos determiná-lo univocamente a partir de $\underline{\psi}$. Este facto transparece na função de probabilidade

$$f(\underline{y}|N, \underline{\lambda}) = \frac{N!}{y_a! y_b! \prod_{j=1}^4 y_{3j}!} (\theta_{1.} + \theta_{2.})^{y_a + y_b} \theta_{3.}^{y_{3.}} (\theta_{.1} + \theta_{.2})^{y_a} (\theta_{.3} + \theta_{.4})^{y_b} \prod_{j=1}^4 \theta_{.j}^{y_{3j}}$$

na qual os parâmetros $\theta_{1.}$ e $\theta_{2.}$ aparecem sempre juntos. Deste modo, a não identificabilidade de $\underline{\lambda}$ permanece sob o modelo H_I , enquanto que $\underline{B}\underline{\lambda}$, com

$$\underline{B} = \begin{pmatrix} \underline{I}_2 & \underline{O}'_4 \\ \underline{O}_{(4,2)} & \underline{I}_4 \end{pmatrix} \text{ é um exemplo de uma função identificante no espaço paramétrico restringido.}$$

Note-se ainda que se fosse obtido apenas o primeiro bloco confundido, isto é, se $\underline{y}^* = \underline{D}^* \underline{u}$ onde \underline{D}^* é a matriz (9x12) obtida da matriz \underline{D} anterior por substituição da 2ª linha pela submatriz

$$\begin{pmatrix} \underline{O} & \underline{I}_2 & \underline{O} & \underline{O} & \underline{O} & \underline{O} \\ \underline{O} & \underline{O} & \underline{O} & \underline{I}_2 & \underline{O} & \underline{O} \end{pmatrix} \text{ onde } \underline{O} \text{ é } 2 \times 2, \text{ a cor-}$$

respondente função de probabilidade sob as mesmas hipóteses seria

$$f(\underline{y}^*|N, \underline{\lambda}) = \frac{N!}{y_a! \prod_{i=1,2} \prod_{j=3,4} y_{ij}! \prod_{j=1}^4 y_{3j}!} \times (\theta_{1.} + \theta_{2.})^{y_a} \theta_{3.}^{y_{3.}} \theta_{1.}^{y_{13} + y_{14}} \theta_{2.}^{y_{23} + y_{24}} (\theta_{.1} + \theta_{.2})^{y_a} \prod_{j=1,2} \theta_{.j}^{y_{3j}} \prod_{j=3,4} \theta_{.j}^{y_{.j}}$$

Isto evidencia que $\underline{\lambda}$ passaria a ser identificável e portanto, o ajustamento do modelo de independência seria desbloqueado uma vez que, adicionalmente, o nº de observações independentes excederia em 3 o nº de parâmetros a estimar.

Estes exemplos mostram que o problema da ocorrência de dados parcialmente categorizados (com as categorias observadas disjuntas) pode também ser traduzido pela relação funcional $\underline{y} = \underline{D}\underline{u}$, entre o vector observado $m \times 1, \underline{y}$, e o vector $q \times 1$ ($q > m$) não observado (ou não observável) \underline{u} , onde \underline{D} é uma matriz de zeros e uns sem linhas nulas e com um único 1 em cada coluna. No esquema de amostragem multinomial, a correspondente relação entre os parâmetros $\underline{\psi} = \underline{D}\underline{\theta}(\underline{\lambda})$ (a mesma relação verifica-se para as médias do modelo Poisson), permite compreender que o problema da identificabilidade do parâmetro indexante está dependente não só do modelo funcional para $\underline{\theta}$ expresso genericamente pela relação $\underline{\theta}(\underline{\lambda})$, onde $\underline{\lambda}$ é $p \times 1$ ($p \leq q$), mas também do próprio processo observacional em termos da configuração de \underline{D} . Ainda que se considere um modelo identificável para $\underline{\theta}$ (caso dos modelos log-lineares do exemplo II com as usuais restrições de identificabilidade), a identificabilidade do modelo para $\underline{\psi}$ não fica assegurada, como o atesta o primeiro exemplo dado relativo

ã tabela 3x4. Obviamente que a configuração da matriz D é irrelevante sob um modelo saturado para $\theta(p=q)$, já que este tipo de modelo implica necessariamente a não identificabilidade do parâmetro indexante. As conclusões extraídas para os modelos linear ou log-linear, em classificações cruzadas com dados contínuos completamente "missing" ou com categorias truncadas, mantêm-se assim para os modelos usuais aplicados a dados categorizados de forma parcial.

Dado que ψ é um parâmetro identificante, a identificabilidade do parâmetro λ está assim intimamente associada com a existência de uma solução única para o sistema de m equações, frequentemente não lineares, a p incógnitas $D\theta(\lambda) = \psi$, no qual se considera que λ é seleccionado de modo que θ satisfaz a restrição imposta pelo modelo multinomial. Como frisámos em 2.1.1, a identificabilidade de λ não garante que, por exemplo, o método de máxima verosimilhança produza uma solução única, para um dado y . Em Haberman (1974), e.g., podemos encontrar exemplos deste tipo de problema com dados parcialmente categorizados, onde a função de verosimilhança para certos dados não é côncava.

Os juízos desenvolvidos para o modelo multinomial estendem-se naturalmente ao modelo produto de multinomiais. Um importante caso especial surge quando u se reporta a uma tabela de contingência multidimensional e a matriz D configura as observações nas margens de um subconjunto das variáveis categorizadas em estudo. Quando as variáveis restantes são usadas no sentido de tentar explicar as interações existentes entre as variáveis observadas, cai-se no contexto dos modelos de estrutura latente de Goodman (1974).

2.1.3. Identificabilidade versus informação amostral

A inspecção da identificabilidade através da definição, dada em 2.1.1, de muitos modelos parametricamente complexos não é tarefa fácil e, muito menos, imediata. É, várias vezes, o confronto com situações anómalas quando se tenta avançar na realização de inferências que leva o pesquisador a reanalisar, mais detalhadamente, a estrutura do modelo em estudo. Por outro lado, a maior parte dos primeiros resultados sobre o problema da identificabilidade dirigiram-se a modelos específicos - em particular, a modelos lineares da econometria - o que tendia a obscurecer a natureza efectivamente geral do problema devidamente salientada por Koopmans e Reiersol (1950). Foi o reconhecimento deste facto que levou Rothenberg (1971) a tentar derivar critérios gerais de identificabilidade. Compreendendo que a não identificabilidade é reflexo da falta de uma informação suficiente, na especificação do modelo, para a discriminação entre valores paramétricos alternativos, Rothenberg abordou o problema em termos da medida de informação de Fisher. Bowden (1973) generalizou esta abordagem através do uso da medida de informação de Kullback-Leibler. Nesta sub-secção, visa-se sistemati-

zar e discutir os principais resultados destes dois trabalhos, ilustrando-os através de exemplos descritos em 2.1.2.

Seja θ_0 um determinado valor do parâmetro $\theta \in \Xi$ indexante da família de distribuições $F(y|\theta)$ definida em (Y, \mathcal{A}) . A função de informação de Kullback-Leibler para discriminar $F(y|\theta_0)$ contra $F(y|\theta)$ por observação de $F(y|\theta_0)$ é a função real definida em $\Xi \times \Xi$ por

$$I(\theta_0, \theta) = H(\theta_0, \theta_0) - H(\theta_0, \theta) = E\left[\log \frac{dF(y|\theta_0)}{dF(y|\theta)} \Big| \theta_0\right] \quad (2.10)$$

onde $H(\theta_0, \theta) = E[\log dF(y|\theta) | \theta_0]$. O seu contradomínio é $[0, \infty]$ (Kullback [1959], pág. 11) e toma apenas valores finitos se $P_{\theta_0} \ll P_{\theta}$, onde o símbolo \ll entre duas medidas significa continuidade absoluta de qualquer uma em relação à outra (vide, e.g., Wilks [1962], pg. 347). Esta condição, doravante assumida, é verificada nos modelos usuais do caso dominado, com suporte independente de θ , onde $I(\theta_0, \theta)$ passa a ser definida pelo integral de Lebesgue estendido a Y da função $\log\left[\frac{f(y|\theta_0)}{f(y|\theta)}\right]f(y|\theta_0)$, com $f(y|\theta)$ designando a função densidade de probabilidade de ou a função de probabilidade de y .

Teor. 2.5: Sejam $\theta_1 \neq \theta_0$ dois pontos de Ξ para os quais $P_{\theta_1} \equiv P_{\theta_0}$. Então $\theta_1 \in [\theta_0]$ sse $I(\theta_0, \theta_1) = 0$.

Dem.: Fazendo $A(\theta_0, \theta_1) = \{y \in Y : dF(y|\theta_1) = dF(y|\theta_0)\}$ e denotando $\bar{A}(\theta_0, \theta_1)$ o seu complementar, a definição de equivalência observacional de θ_1 e θ_0 indica que $\int_{A(\theta_0, \theta_1)} dF(y|\theta_0) \equiv P_{\theta_0}[A(\theta_0, \theta_1)] = 1$. Suponhamos que $\theta_1 \in [\theta_0]$. Como $\log \frac{dF(y|\theta_0)}{dF(y|\theta_1)} = 0$ em $A(\theta_0, \theta_1)$ e $\int_{\bar{A}(\theta_0, \theta_1)} dF(y|\theta_0) = 0$, conclui-se por (2.10) que $I(\theta_0, \theta_1) = 0$. Admitamos agora que θ_1 não é observacionalmente equivalente a θ_0 . Então

$$I(\theta_0, \theta_1) = \int_{\bar{A}(\theta_0, \theta_1)} -\log\left[\frac{dF(y|\theta_1)}{dF(y|\theta_0)}\right]dF(y|\theta_0) > -\log \int_{\bar{A}(\theta_0, \theta_1)} dF(y|\theta_1)$$

por aplicação da desigualdade de Jensen, já que $-\log(\cdot)$ é uma função estritamente convexa e a função integranda não é constante em $\bar{A}(\theta_0, \theta_1)$. Por hipótese, $P_{\theta_0}[\bar{A}(\theta_0, \theta_1)] > 0$, o que implica $P_{\theta_1}[\bar{A}(\theta_0, \theta_1)] > 0$, pelo facto de $P_{\theta_0} \ll P_{\theta_1}$. Assim, $I(\theta_0, \theta_1) > 0$. Como o contradomínio de $I(\cdot, \cdot)$ é $[0, \infty)$, segue-se então que $I(\theta_0, \theta_1) = 0 \implies \theta_1 \in [\theta_0]$, como pretendíamos. Δ

De acordo com o resultado deste teorema e sob a hipótese de $P_{\underline{\theta}} \equiv P_{\underline{\theta}_0}$,

$\forall \underline{\theta} \in \Xi$, se $\underline{\theta}_0$ é globalmente identificável, então $I(\underline{\theta}_0, \underline{\theta}) > 0$, $\forall \underline{\theta} \neq \underline{\theta}_0$, i.e., $I(\underline{\theta}_0, \underline{\theta}) = 0$ só se verifica para $\underline{\theta} = \underline{\theta}_0$, o que significa que $\underline{\theta}_0$ é o único minimizante (maximizante) de $I(\underline{\theta}_0, \cdot)$ ($H(\underline{\theta}_0, \cdot)$). Se $\underline{\theta}_0$ não é globalmente identificável, existe então pelo menos um $\underline{\theta} \neq \underline{\theta}_0$ tal que $I(\underline{\theta}_0, \underline{\theta}) = 0$, ou seja, que é igualmente minimizante de $I(\underline{\theta}_0, \cdot)$. Desta forma, fica demonstrado, como corolário do teorema 2.5, o critério de identificabilidade de Bowden (1973):

Critério I: Supondo que $P_{\underline{\theta}} \equiv P_{\underline{\theta}_0}$, $\forall \underline{\theta} \in \Xi$, $\underline{\theta}_0$ é globalmente (localmente) identificável sse $\underline{\theta}_0$ é a única solução da equação $I(\underline{\theta}_0, \underline{\theta}) = 0$ em Ξ (numa vizinhança aberta de $\underline{\theta}_0$).

Se este critério é verificado para um $\underline{\theta}_0 \in \Xi$ arbitrário, fica definida uma condição necessária e suficiente de identificabilidade de Ξ quando a família é auto-dominada por qualquer dos seus membros.

De acordo com este critério, a relevância da identificabilidade paramétrica na averiguação da consistência do estimador de máxima verosimilhança do parâmetro pode ser vista em termos da condição de $\underline{\theta}_0$ ser o único minimizante de $I(\underline{\theta}_0, \underline{\theta}) = H(\underline{\theta}_0, \underline{\theta}_0) - H(\underline{\theta}_0, \underline{\theta})$. Quando \underline{y} representa um conjunto de n observações y_i i.i.d., a lei dos grandes números assegura que $n^{-1} \log f(\underline{y} | \underline{\theta})$ converge para $H(\underline{\theta}_0, \underline{\theta})$ para todo o $\underline{\theta}$ e quase todo o \underline{y} , na hipótese assumida no critério (ela garante que $H(\underline{\theta}_0, \underline{\theta}) > -\infty$). Como Silvey (1970) apontou, se esta convergência for uniforme com respeito a $\underline{\theta}$, $n^{-1} \log f(\underline{y} | \underline{\theta})$, para n grande e quase todo o \underline{y} , estará uniformemente perto de $H(\underline{\theta}_0, \underline{\theta})$ e, sob condições de regularidade adequadas, atingirá o seu supremo perto do ponto onde $I(\underline{\theta}_0, \underline{\theta})$ atinge o seu ínfimo (que é o ponto de supremo de $H(\underline{\theta}_0, \underline{\theta})$). Daí a não consistência do estimador de máxima verosimilhança para um $\underline{\theta}_0$ não identificável.

O critério I, quando aplicável, goza da vantagem de reunir a inspecção da identificabilidade global e local e de permitir definir a partição identificante. De facto, pelo teorema 2.5, $[\underline{\theta}_0] = \{\underline{\theta} \in \Xi : I(\underline{\theta}_0, \underline{\theta}) = 0\}$. Contudo, para muitos modelos não é fácil determinar directamente todas as soluções da equação $I(\underline{\theta}_0, \underline{\theta}) = 0$. O caso do modelo linear normal, ilustrado em Bowden, é um exemplo bem ilustrativo dessa situação, se bem que no caso particular em que a matriz de covariâncias dos erros é conhecida, aquela equação é já facilmente resolvida. Como as soluções da referida equação correspondem aos pontos de mínimo da função $I(\underline{\theta}_0, \cdot)$, esse problema pode ser desbloqueado nos casos em que $I(\underline{\theta}_0, \underline{\theta})$ é uma função contínua de $\underline{\theta}$, com Ξ um subconjunto aberto de um espaço euclídeano, mediante a aplicação dos métodos de optimização da análise matemática ou de programação não-linear. No caso não restringido, se $\frac{\partial}{\partial \underline{\theta}} I(\underline{\theta}_0, \underline{\theta}) = 0$ é somente satisfeita por $\underline{\theta} = \underline{\theta}_0$, então esta solu-

ção é necessariamente o único minimizante, o que garante a identificabilidade global (e local) de $\underline{\theta}_0$. Se há vários pontos críticos de $I(\underline{\theta}_0, \cdot)$ torna-se obviamente necessário indagar a sua natureza. Se $\underline{\theta}_1$ é um dos pontos críticos que anula $I(\underline{\theta}_0, \underline{\theta})$, então $\underline{\theta}_0$ não é globalmente identificável, podendo ser ou não localmente identificável. De qualquer modo, todos os pontos críticos que anulam $I(\underline{\theta}_0, \underline{\theta})$ definem a classe de valores observacionalmente equivalentes a $\underline{\theta}_0$.

Sobre a aplicabilidade deste critério, vale a pena sublinhar que a hipótese de continuidade absoluta nele assumida exclui os casos não regulares onde \mathcal{Y} depende de $\underline{\theta}$. Contudo, a análise da demonstração do teorema 2.5 permite-nos concluir que:

- . A condição $I(\underline{\theta}_0, \underline{\theta}) = 0$ é sempre necessária para a equivalência observacional de $\underline{\theta}$ e $\underline{\theta}_0$;
- . A condição $I(\underline{\theta}_0, \underline{\theta}) > 0$ é necessária para a não equivalência observacional de $\underline{\theta}$ e $\underline{\theta}_0$ desde que $I(\underline{\theta}_0, \underline{\theta})$ seja finita.

Em ordem a ilustrar a aplicabilidade destas conclusões nos referidos casos não regulares, consideremos que \mathcal{Y} é o intervalo da recta real $L(\theta) = [\phi_1(\theta), \phi_2(\theta)]$, onde $\phi_i(\theta)$, $i=1,2$ são funções contínuas de $\theta \in \mathbb{R}$ com $\phi_1(\theta) < \phi_2(\theta)$, $\forall \theta \in \mathbb{R}$, e seja \mathcal{P} a família de distribuições uniformes em $L(\theta)$. Da expressão de $F(y|\theta)$ é fácil concluir que para qualquer $\theta_0 \in \mathbb{R}$, $[\theta_0] = \{\theta \in \mathbb{R} : \phi_i(\theta) = \phi_i(\theta_0), i=1,2\}$, constatando-se assim que o modelo é identificável se e só se ambos os extremos de $L(\theta)$ são funções injectivas. Atendendo a (2.10), fica patente que $I(\theta_0, \theta) = 0$ para todo o $\theta \in [\theta_0]$, como era de esperar. Por outro lado, para todo o par (θ_0, θ) tal que $L(\theta_0)$ não está contido em $L(\theta)$, $I(\theta_0, \theta) = \infty$. Ou seja, um valor finito de $I(\theta_0, \theta)$ implica que $L(\theta_0) \subset L(\theta)$ (garantindo $P_{\theta_0} \ll P_\theta$) e, nestas circunstâncias, $I(\theta_0, \theta) = \log \frac{\phi_2(\theta) - \phi_1(\theta)}{\phi_2(\theta_0) - \phi_1(\theta_0)}$. É o que acontece, e.g., quando $\theta < \theta_0$ (resp. $\theta > \theta_0$) se $\phi_1(\theta)$ é uma função monótona crescente e $\phi_2(\theta)$ é uma função monótona decrescente (resp., $\phi_1(\theta)$ é uma função monótona decrescente e $\phi_2(\theta)$ uma função monótona crescente). Assim, nesses casos, $I(\theta_0, \theta) = 0$ implica que as amplitudes de $L(\theta)$ e de $L(\theta_0)$ são iguais, e portanto, $L(\theta) = L(\theta_0)$, pelo facto de $L(\theta_0) \subset L(\theta)$. Ou seja, fica confirmado que se $I(\theta_0, \theta)$ é finita, $I(\theta_0, \theta) = 0 \Rightarrow \theta \in [\theta_0]$. Um exemplo concreto é o da família de distribuições uniformes com

$$\phi_2(\theta) = -\phi_1(\theta) = \theta^2 \quad \text{e } \theta \in \mathbb{R}, \text{ onde } [\theta_0] = \{-\theta_0, \theta_0\}, \quad \forall \theta_0 \in \mathbb{R}.$$

Com o objectivo de ilustrar a aplicação deste critério, vamos considerar alguns exemplos da sub-secção 2.1.2 envolvendo dados categorizados.

Exemplo 1 : Consideremos o problema de observações multinomiais parcialmente categorizadas, genericamente descrito no final do exemplo IV de 2.1.2., com função de probabilidade

restringido pelas restrições naturais, não-lineares, $C'_k v(\beta) = N_k$, $k=1, \dots, s$. A

Estes dois exemplos, assim como o do modelo linear normal, ilustram a situação de continuidade da função $I(\theta_0, \cdot)$, que é verificada sob as usuais condições de regularidade na família $\mathcal{P} = \{F(\underline{y}|\underline{\theta}) : \underline{\theta} \in E\}$:

- \mathcal{P} é dominada por uma medida σ -finita m , sendo $\frac{dF(\underline{y}|\underline{\theta})}{dm(\underline{y})} = f(\underline{y}|\underline{\theta})$;
- o suporte \mathcal{Y} de $f(\underline{y}|\underline{\theta})$ é independente de $\underline{\theta} \in E$, com E um subconjunto aberto de \mathbb{R}^k ;
- As funções $f(\underline{y}|\underline{\theta})$ e $\log f(\underline{y}|\underline{\theta})$ são duas vezes diferenciáveis em $\underline{\theta}$ com derivadas contínuas para todo o $\underline{y} \in \mathcal{Y}$ e todo o $\underline{\theta} \in E$;
- Os elementos da função "score" $\underline{S}(\underline{y}|\underline{\theta}) = \frac{\partial}{\partial \underline{\theta}} \log f(\underline{y}|\underline{\theta})$ e das matrizes $\underline{S}(\underline{y}|\underline{\theta}) \underline{S}'(\underline{y}|\underline{\theta})$ e $\frac{\partial}{\partial \underline{\theta}'} \underline{S}(\underline{y}|\underline{\theta})$ são integráveis em relação a $dF(\underline{y}|\underline{\theta})$, para todo o $\underline{\theta}$.

Estas condições asseguram, como se sabe, que $E[\underline{S}(\underline{y}|\underline{\theta})|\underline{\theta}] = \underline{0}$ e $\underline{I}(\underline{\theta}) \equiv E[\underline{S}(\underline{y}|\underline{\theta}) \underline{S}'(\underline{y}|\underline{\theta})|\underline{\theta}] = -E[\frac{\partial}{\partial \underline{\theta}'} \underline{S}(\underline{y}|\underline{\theta})|\underline{\theta}]$, onde $\underline{I}(\underline{\theta})$ designa a matriz de informação de Fisher sobre $\underline{\theta}$ por observação de $F(\underline{y}|\underline{\theta})$. Como

$$I(\underline{\theta}_0, \underline{\theta}) = \int_{\mathcal{Y}} \log \left[\frac{f(\underline{y}|\underline{\theta}_0)}{f(\underline{y}|\underline{\theta})} \right] f(\underline{y}|\underline{\theta}_0) dm(\underline{y})$$

segue-se então que $\frac{\partial}{\partial \underline{\theta}'} I(\underline{\theta}_0, \underline{\theta})|_{\underline{\theta}_0} = -E[\underline{S}(\underline{y}|\underline{\theta}_0)|\underline{\theta}_0] = \underline{0}$ e $\frac{\partial^2 I(\underline{\theta}_0, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'}|_{\underline{\theta}_0} = \underline{I}(\underline{\theta}_0)$.

Consequentemente, a condição de $\underline{I}(\underline{\theta}_0)$ ser definida positiva é suficiente para a existência de um mínimo local único de $I(\underline{\theta}_0, \underline{\theta})$. Por outro lado, a expansão de $I(\underline{\theta}_0, \underline{\theta})$ em fórmula de Taylor em torno de $\underline{\theta}_0$ origina que

$$I(\underline{\theta}_0, \underline{\theta}) = \frac{1}{2} (\underline{\theta} - \underline{\theta}_0)' \underline{I}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}_0), \quad \underline{\theta}^* = t \underline{\theta}_0 + (1-t) \underline{\theta}, \quad 0 < t < 1$$

Assim, se $I(\underline{\theta}_0, \underline{\theta}) > 0$ na vizinhança de $\underline{\theta}_0$, $\underline{I}(\underline{\theta}^*)$ é definida positiva. Impondo o que Rothenberg apelidou de condição de regularidade de $\underline{I}(\underline{\theta})$ em $\underline{\theta}_0$, i.e., que o posto de $\underline{I}(\underline{\theta})$ não varia numa vizinhança aberta de $\underline{\theta}_0$, podemos concluir então que $\underline{I}(\underline{\theta}_0)$ é definida positiva. Atendendo ao critério I na sua versão local, fica demonstrado como seu corolário o critério de identificabilidade local de Rothenberg (1971) expresso em termos da matriz de informação de Fisher:

Critério II: Sob as condições de regularidade supra-citadas, a não singularidade de $\underline{I}(\underline{\theta}_0)$ implica a identificabilidade local de $\underline{\theta}_0$ que, por sua vez, implica a primeira condição se $\underline{I}(\underline{\theta})$ é regular numa vizinhança aberta de $\underline{\theta}_0$.

Nota-se assim que a singularidade de $\underline{I}(\underline{\theta}_0)$, por si só, não implica necessariamente a não identificabilidade local de $\underline{\theta}_0$. Este facto pode, adicionalmente, ser visto a partir da expansão em série de Taylor de $I(\underline{\theta}_0, \underline{\theta})$ para $\underline{\theta}$ próximo de $\underline{\theta}_0$.

$$f(\underline{y}|N, \underline{\theta}) = \frac{N!}{\prod_{i=1}^m y_i!} \exp[\underline{y}' \cdot \log(D\underline{\theta})], \quad \begin{aligned} \mathbf{1}'_q \underline{\theta} &= 1 \\ \mathbf{1}'_m \underline{y} &= N \end{aligned}$$

Fazendo $\underline{\psi} = (\psi_1 \dots \psi_m)'$ e $\underline{\psi}(\underline{\theta}) = D\underline{\theta}$ e $\underline{\psi}_0 = (\psi_{01} \dots \psi_{0m})'$ a imagem por $\underline{\psi}(\cdot)$ de um ponto $\underline{\theta}_0$ do simplex probabilístico de \mathbb{R}^q , Ξ , é fácil constatar que

$$I(\underline{\theta}_0, \underline{\theta}) = \underline{\psi}'_0 \cdot [\log \underline{\psi}_0 - \log \underline{\psi}(\underline{\theta})] = \sum_{i=1}^m \psi_{0i} \log \frac{\psi_{0i}}{\psi_i}$$

função que é identicamente nula sse $\psi_i = \psi_{0i}$, $i=1, \dots, m$. Assim,

$$[\underline{\theta}_0] = \{\underline{\theta} \in \Xi : D\underline{\theta} = D\underline{\theta}_0\} = \{\underline{\theta} \in \Xi : \underline{\theta} = \underline{\theta}_0 + \underline{\alpha}\}$$

onde o conjunto de valores possíveis de $\underline{\alpha}$ está incluído no subespaço de \mathbb{R}^q , $N(D)$, o espaço nulo de D .

Exemplo 2: Consideremos agora a distribuição de Poisson (2.8) sob o modelo log-linear para o vector média das q celas definido no exemplo II de 2.1.2. Fazendo $\underline{\mu} = (\mu_1 \dots \mu_q)'$ e $\underline{\mu}(\underline{\beta}) = \exp(\underline{X}_0 \underline{\beta})$ e denotando $\underline{\mu}_0 = (\mu_{01} \dots \mu_{0q})'$ a imagem por $\underline{\mu}(\cdot)$ de um determinado $\underline{\beta}_0 \in \mathbb{R}^p$, pode verificar-se que

$$I(\underline{\beta}_0, \underline{\beta}) = \underline{\mu}'_0 [\log \underline{\mu}_0 - \log \underline{\mu}(\underline{\beta})] - \mathbf{1}'_q [\underline{\mu}_0 - \underline{\mu}(\underline{\beta})]$$

Em ordem a detalhar a definição do conjunto de minimizantes globais de $I=I(\underline{\beta}_0, \underline{\beta})$ determinemos os seus pontos críticos, que denominaremos por $\underline{\beta}^*$.

Como

$$\frac{\partial I}{\partial \underline{\mu}} = -D_{\underline{\mu}}^{-1} \underline{\mu}_0 + \mathbf{1}_q \quad \text{e} \quad \frac{\partial \underline{\mu}}{\partial \underline{\beta}'} = D_{\underline{\mu}}(\underline{\beta}) \underline{X}_0$$

resulta

$$\frac{\partial I}{\partial \underline{\beta}} \Big|_{\underline{\beta}^*} \equiv \left(\frac{\partial \underline{\mu}}{\partial \underline{\beta}'} \right)' \frac{\partial I}{\partial \underline{\mu}} \Big|_{\underline{\beta}^*} = 0 \iff \underline{X}'_0 [\underline{\mu}(\underline{\beta}^*) - \underline{\mu}_0] = 0 \quad (2.11)$$

Nos modelos log-lineares usuais, $\mathbf{1}_q \in M(\underline{X}_0)$ e (2.11) implica então que $\mathbf{1}'_q [\underline{\mu}_0 - \underline{\mu}(\underline{\beta}^*)] = 0$. Por conseguinte,

$$\begin{aligned} [\underline{\beta}_0] &= \{\underline{\beta}^* \in \mathbb{R}^p : \underline{\mu}'_0 [\log \underline{\mu}_0 - \log \underline{\mu}(\underline{\beta}^*)] = 0\} \\ &= \{\underline{\beta}^* \in \mathbb{R}^p : \underline{\mu}(\underline{\beta}^*) = \underline{\mu}_0\} = \{\underline{\beta}^* \in \mathbb{R}^p : \underline{X}_0 \underline{\beta}^* = \underline{X}_0 \underline{\beta}_0\} \end{aligned}$$

Desta forma, o domínio de $\underline{\beta}$ fica decomposto em classes definidas pelos subespaços transladados $\{\underline{\beta}_0 + N(\underline{X}_0)\}$ de dimensão $p-r$, um resultado análogo ao obtido para o modelo linear normal. Usando o modelo multinomial (2.9) obter-se-á um resultado formalmente idêntico, com a diferença de já não se poder traduzir cada classe de equivalência em termos de um subespaço transladado, dado que o domínio de $\underline{\beta}$ fica

$$I(\underline{\theta}_0, \underline{\theta}) = \frac{1}{2} (\underline{\theta} - \underline{\theta}_0)' I(\underline{\theta}_0) (\underline{\theta} - \underline{\theta}_0) + o(\|\underline{\theta} - \underline{\theta}_0\|^3)$$

Os termos de ordem superior podem assegurar que $I(\underline{\theta}_0, \underline{\theta}) > 0$ para todo o $\underline{\theta} \neq \underline{\theta}_0$ na vizinhança de $\underline{\theta}_0$, i.e., a não identificabilidade local de $\underline{\theta}_0$, ainda que o 1º termo do 2º membro seja nulo.

Rothenberg estende ainda o critério II ao caso restringido em que $\Xi = \{\underline{\theta} \in \mathbb{R}^k : h_i(\underline{\theta}) = 0, i=1, \dots, m\}$, onde cada função h_i possui derivadas parciais contínuas. O critério generalizado relaciona a identificabilidade local de $\underline{\theta}_0$ com a não singularidade da matriz $V_0' V_0$, onde $V_0 = \begin{pmatrix} I(\underline{\theta}_0) \\ H(\underline{\theta}_0) \end{pmatrix}$, $H(\underline{\theta}) = \frac{\partial h(\underline{\theta})}{\partial \underline{\theta}}$ e $h(\underline{\theta}) = (h_1(\underline{\theta}) \dots h_m(\underline{\theta}))'$, no espírito do critério II.

Dadas as limitações da identificabilidade local, Rothenberg consegue ainda estabelecer a seguinte condição suficiente de identificabilidade global baseada apenas na matriz de informação de Fisher:

Critério III: Seja $f(\underline{y}|\underline{\theta})$ pertencente à família exponencial multiparamétrica, isto é, $f(\underline{y}|\underline{\theta}) = g(\underline{y})\lambda(\underline{\theta}) \exp[\underline{\phi}'(\underline{\theta}) \cdot T(\underline{y})]$ onde $\underline{\phi} = \underline{\phi}(\underline{\theta}) : \Xi \rightarrow \Phi$ é o respectivo parâmetro natural. Então, Φ é identificável se $I(\underline{\phi})$ é não singular em Φ .

A identificabilidade do parâmetro original $\underline{\theta}$ seguir-se-á da não singularidade de $I(\underline{\phi})$ se puder ser expresso como uma função de $\underline{\phi}$ pelo teorema 2.1.

Exemplo 3: Seja \underline{y} um vector multinomial $(q-1)$ -dimensional de parâmetros N (conhecido) e $\underline{\theta}^* = (\theta_1 \dots \theta_{q-1})'$ com $1 - \theta_{q-1} < 1$. A função de probabilidade pode ser expressa na forma indicada no critério III, onde $\lambda(\underline{\theta}^*) = (1 - 1 - \theta_{q-1}^*)^N$, $T(\underline{y}) = \underline{y}$ e $\underline{\phi}(\underline{\theta}^*) = \log[\underline{\theta}^*/(1 - 1 - \theta_{q-1}^*)]$. A matriz $I(\underline{\phi}) = N[D_{\exp(\underline{\phi})} - \exp(\underline{\phi})(\exp(\underline{\phi}))'] / (1 + 1 - \theta_{q-1} \exp(\underline{\phi}))$ é claramente não singular $\forall \underline{\phi} \in \Phi$ e, conseqüentemente, $\underline{\phi}$ é identificável. Como $1 - 1 - \theta_{q-1}^* = (1 + 1 - \theta_{q-1} \exp(\underline{\phi}))^{-1}$ e $\underline{\theta}^* = \exp(\underline{\phi}) / (1 + 1 - \theta_{q-1} \exp(\underline{\phi}))$ segue-se que $\underline{\theta}^*$ e $\underline{\theta} = (\underline{\theta}^* \cdot 1 - 1 - \theta_{q-1} \exp(\underline{\phi}))'$ são parâmetros igualmente identificáveis. Contudo, note-se que $I(\underline{\theta}) = N(D_{\underline{\theta} - \underline{\theta}\underline{\theta}'})$ é singular para todo o $\underline{\theta} \in \{\underline{\theta} \in \mathbb{R}^q : 0 < \theta_i < 1, i=1, \dots, q\}$, o que exemplifica que nem todo o parâmetro identificável corresponde a uma matriz de informação definida positiva. De qualquer modo, a existência da restrição $1 - \theta_{q-1} = 1$ leva a que sejam funcionalmente independentes apenas $q-1$ elementos de $\underline{\theta}$, por exemplo, os elementos de $\underline{\theta}^* = \underline{A}\underline{\theta}$, $\underline{A} = (I_{q-1} \ 0_{(q-1)})$, cuja matriz de informação é já não singular. A aplicação do critério II generalizado de Rothenberg garante a identificabilidade local de $\underline{\theta}$ quando considerado restringido pela condição linear referida, já que o posto de $\begin{pmatrix} I(\underline{\theta}) \\ 1 - \theta_{q-1} \end{pmatrix}$ é q .

Naturalmente que neste caso os critérios II e III não são os instrumentos mais adequados para apurar a identificabilidade de $\underline{\theta}^*$ (ou de $\underline{\theta}$). O primeiro porque se refere apenas à identificabilidade local e o segundo porque exige uma reparametrização prévia, que é desnecessária. O recurso à definição ou ao facto de $\underline{\theta}^*$ ser estimável por \underline{y}/N (veja-se secção seguinte) permite obter directamente tal conclusão.

Exemplo 4: Estas considerações aplicam-se nomeadamente ao 1º caso considerado no exemplo III da sub-secção 2.1.2. onde $q=2$ e $\underline{\theta}$ reflecte a reparametrização efectuada sobre o vector das médias das distribuições de Poisson originais. Consideremos por agora que o parâmetro de interesse é $\underline{\mu} = (\mu_1 \mu_2)' \in U$ e analisemos de novo a questão da identificabilidade à luz dos critérios supra mencionados. Pode constatar-se facilmente que

$$\tilde{I}(\underline{\mu}) = \frac{N}{\mu_1 + \mu_2} \begin{pmatrix} \mu_1^{-1} - (\mu_1 + \mu_2)^{-1} & -(\mu_1 + \mu_2)^{-1} \\ -(\mu_1 + \mu_2)^{-1} & \mu_2^{-1} - (\mu_1 + \mu_2)^{-1} \end{pmatrix}$$
 é singular, já que a 1ª coluna é igual à 2ª coluna multiplicada por $-\frac{\mu_2}{\mu_1}$. O critério II assegura então que $\underline{\mu}$ não é identificável localmente e, portanto, também globalmente. A aplicação do critério I permite adicionalmente definir as classes de equivalência em que U fica particionado. Denotando por $\underline{\mu}_0 = (\mu_1^0, \mu_2^0)'$ o verdadeiro valor de $\underline{\mu}$, é fácil verificar que

$$I(\underline{\mu}_0, \underline{\mu}) = \sum_{i=1}^2 N \frac{\mu_i^0}{\mu_1^0 + \mu_2^0} \log \left(\frac{\mu_i^0}{\mu_1^0 + \mu_2^0} \frac{\mu_1 + \mu_2}{\mu_i} \right)$$

e por derivação

$$\frac{\partial I}{\partial \mu_i} = 0 \Rightarrow \frac{\mu_i}{\mu_1 + \mu_2} = \frac{\mu_i^0}{\mu_1^0 + \mu_2^0}, \quad i = 1, 2$$

O conjunto dos pontos críticos não é assim singular e todos eles anulam $I(\underline{\mu}_0, \underline{\mu})$. Por conseguinte, $\underline{\mu}_0$ não é identificável e o conjunto dos pontos de mínimo de $I(\underline{\mu}_0, \underline{\mu})$ define a classe $[\underline{\mu}_0]$ determinada na sub-secção 2.1.2.

2.1.4. Identificabilidade versus estimabilidade

Já referimos que em muita da literatura estatística, o conceito de identificabilidade aparece camuflado pelo termo estimabilidade, um conceito que é, independentemente das diversas acepções em que é utilizado, estritamente inferencial. Na secção 2.1.1. mostrámos que a identificabilidade é uma condição necessária de estimabilidade, entendendo esta no sentido de unicidade de estimação em termos independentes dos dados, ou de existência de estimadores consistentes.

Em certas áreas, como na de Modelos Lineares, o termo estimabilidade é mais frequentemente usado no seguinte sentido:

Def. 2.7: A função $\phi = \phi(\theta)$ diz-se (linearmente) estimável se existe um estimador centrado, $T(\underline{y})$, de ϕ (que é função linear de \underline{y}).

O objectivo desta sub-secção reside precisamente em enunciar e discutir as relações existentes entre este conceito de estimabilidade e o de identificabilidade.

O teorema seguinte, que é uma generalização do resultado referido em 2.1.1 concernente à identificabilidade de um parâmetro indexante estimável, mostra que este conceito de estimabilidade continua sendo condição suficiente de identificabilidade.

Teor. 2.6: Se $\phi(\theta)$ é uma função estimável, então $\phi(\theta)$ é identificável.

Dem: Por hipótese, existe uma estatística $T = T(\underline{y}): (Y, \mathcal{A}) \rightarrow (T(Y), \mathcal{A}_1)$ P -integrável tal que $E[T(\underline{y}) | \theta] = \phi(\theta)$, $\forall \theta \in \Xi$, onde $E[T(\underline{y}) | \theta]$ é o vector de componentes $E[T_i(\underline{y}) | \theta] = \int_Y t_i(\underline{y}) dF(\underline{y} | \theta)$. Consideremos então dois pontos quaisquer de Ξ , θ_0 e θ_1 , observacionalmente equivalentes. Devido à identidade distribucional de $F(\underline{y} | \theta_1)$ e $F(\underline{y} | \theta_0)$

$$\phi_i(\theta_1) \equiv E[T_i(\underline{y}) | \theta_1] = E[T_i(\underline{y}) | \theta_0] \equiv \phi_i(\theta_0), \quad \forall i$$

o que traduz, pela definição 2.4, a identificabilidade de $\phi(\theta)$. Δ

Este teorema mostra assim que toda a função estimável (em particular, os momentos populacionais de \underline{y} quando existem) induz uma partição identificável, o que não significa que coincida com a partição identificante. Com efeito, suponhamos, para facilitar, o caso dominado unidimensional onde Y é o subconjunto discreto de \mathbb{R} , independente de Ξ , formado pelos pontos y_j , $j \in \Delta$, com $\Delta = \{1, 2, \dots, k\}$. Da definição de valor esperado da variável aleatória $T(\underline{y})$, de corre que

$$\begin{aligned} E[T(\underline{y}) | \theta] &= \sum_{j \in \Delta} T(y_j) [F(y_j | \theta) - F(y_{j-1} | \theta)] \\ &= T(y_0) + \sum_{j \in \Delta} [T(y_j) - T(y_{j-1})] (1 - F(y_{j-1} | \theta)) \end{aligned}$$

onde y_0 é qualquer ponto menor que y_1 (portanto, com função de distribuição nula) com $T(y_0)$ finito. Assim, para $\theta_1, \theta_0 \in \Xi$

$$E[T(\underline{y}) | \theta_1] - E[T(\underline{y}) | \theta_0] = \sum_{j \in \Delta} [T(y_j) - T(y_{j-1})] [F(y_{j-1} | \theta_0) - F(y_{j-1} | \theta_1)]$$

Admitamos que $\theta_1 \notin [\theta_0]$ e seja $\Delta_{1/0}$ o conjunto dos $j \in \Delta$ tais que $F(y_{j-1} | \theta_1) \neq F(y_{j-1} | \theta_0)$. Ora, se $T(y)$ for uma função tal que $T(y_{j-1}) = T(y_j)$ para $j \in \Delta_{1/0}$, tem-se $E[T(y) | \theta_1] = E[T(y) | \theta_0]$. Ou seja, a função identificável $\phi(\theta) = E[T(y) | \theta]$ não é suficiente.

Este exemplo mostra ainda que a partição identificante pode ser induzida por funções estimáveis, desde que a função estimadora seja estritamente monótona (crescente ou decrescente). É esta idéia que leva Tuncer (1985) a procurar estabelecer para parâmetros susceptíveis de estimação por funções desse tipo, uma ponte de ligação adicional entre identificabilidade e estimabilidade, nos casos dominados regulares usuais.

Em geral, a estimabilidade não é condição necessária de identificabilidade. Em 2.1.2 vimos, através de vários exemplos, como a perda de certa informação na obtenção dos dados originava um modelo probabilístico, com mais parâmetros que observações, sofrendo de falta de identificabilidade. Deve, contudo, salientar-se, dada a confusão de terminologia existente na literatura, que o facto da dimensionalidade de \mathcal{Y} ser inferior à de Ξ não implica necessariamente que estejamos em presença de um modelo não identificável. Várias dessas situações podem servir como contra-exemplos indicativos de que a identificabilidade não é condição suficiente de estimabilidade. Retrataremos duas situações reais, uma com dados contínuos com largas implicações em modelos ANOVA sem réplicas, e a outra envolvendo um modelo para uma experiência de captura-recaptura visando a estimação do tamanho de uma população animal.

Exemplo 1: Suponhamos que $\underline{y} = (y_1, \dots, y_n)'$ representam medições independentes do peso de n objectos diferentes, com médias μ_i , $i=1, \dots, n$ e variâncias supostas iguais ao valor comum σ^2 . Admitindo uma forma distribucional normal, a

função densidade de probabilidade é $f(\underline{y} | \underline{\theta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ - \frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2} \right\}$,

$\underline{y} \in \mathbb{R}^n$ onde $\underline{\theta} = (\mu_1 \dots \mu_n \sigma^2)'$ é tal que $\sigma^2 > 0$ e $\mu_i \in \mathbb{R}$, $i = 1, \dots, n$. O parâmetro $\underline{\theta}$ é claramente identificável mas não é estimável. Com efeito, σ^2 nem sequer é passível de ser estimado nos moldes usuais, uma vez que existe apenas uma observação para medir a dispersão em torno de cada média.

Exemplo 2: Suponhamos que a experiência consiste em capturar em duas fases animais de uma população de dimensão N (desconhecida). A 1ª fase resulta na captura de n_1 animais que são marcados, e então, libertados. Procede-se, em seguida, a outra experiência que resulta na captura de n_2 animais, x dos quais marcados. Admitamos que todo o animal tem probabilidade θ_1 de ser apanhado na 1ª experiência de captura e que a probabilidade de ser apanhado na experiência

de recaptura depende do facto de ter sido ou não capturado na primeira experiência. Denotemos por θ_2 esta probabilidade para qualquer animal que não tenha sido capturado previamente e por α a correspondente probabilidade quando tenha havido recaptura. Os dados observados são assim traduzidos por $\underline{y} = (n_1, x, n_2 - x)'$ dos quais se pretende estimar os parâmetros $(N, \underline{\theta}')'$, com $\underline{\theta}' = (\theta_1, \theta_2, \alpha)$. Como $n_1 | N, \theta_1 \rightsquigarrow \text{Bi}(N, \theta_1)$, $x | n_1, \alpha \rightsquigarrow \text{Bi}(n_1, \alpha)$ e $n_2 - x | n_1, x, N, \theta_2 \rightsquigarrow \text{Bi}(N - n_1, \theta_2)$, a função de probabilidade de \underline{y} é, com N inteiro positivo e $0 < \theta_1, \theta_2, \alpha < 1$,

$$f(\underline{y} | N, \underline{\theta}) = \frac{N! \theta_1^{n_1} (1 - \theta_1)^{N - n_1} \alpha^x (1 - \alpha)^{n_1 - x} \theta_2^{n_2 - x} (1 - \theta_2)^{N - n_1 - n_2 + x}}{x! (n_1 - x)! (n_2 - x)! (N - n_1 - n_2 + x)!}, \quad \begin{matrix} 0 \leq n_1 \leq N, & 0 \leq x \leq n_1, \\ & 0 \leq n_2 - x \leq N - n_1 \end{matrix}$$

É fácil constatar que N e $\underline{\theta}$ são parâmetros identificáveis. O significado destes e das estatísticas observadas deixa claro que a estimação de α não levanta problemas e que a de θ_1 e de θ_2 dependem da possibilidade de estimação de N . Consequentemente, o facto de se dispor de apenas 3 estatísticas para a estimação de 4 parâmetros repercute na impossibilidade de se proceder à estimação de N e, em decorrência, de θ_1 e θ_2 . Esta afirmação pode ser clarificada através da aplicação do método da máxima verosimilhança. As equações de verosimilhança só permitem obter as expressões $\hat{\alpha} = x/n_1$, $\hat{\theta}_1 = n_1/\hat{N}$ e $\hat{\theta}_2 = (n_2 - x)/(\hat{N} - n_1)$, já que a equação relativa a N é redundante. Δ

Na análise de modelos lineares, $\underline{y} = \underline{X}\underline{\beta} + \underline{u}$, definidos como no exemplo I, onde a forma matemática da distribuição de \underline{u} pode ser relaxada, o interesse concentra-se em funções lineares de $\underline{\theta} = (\underline{\beta}' \sigma^2)'$, concretamente em σ^2 e em $\underline{A}\underline{\beta}$, onde \underline{A} é uma matriz $s \times p$ de posto $s \leq p$, com os vectores linha designados por \underline{a}_i' , $i=1, \dots, s$. Concentremo-nos na função $\underline{A}\underline{\beta}$, já que σ^2 é identificável e também estimável (no sentido da def. 2.7), pois foi admitido que $n > p$. É perfeitamente conhecido o resultado estabelecido na seguinte proposição:

Prop. 2.1: A função linear $\underline{A}\underline{\beta}$ é linearmente estimável sse $\underline{a}_i \in M(\underline{X}')$, $i=1, \dots, s$, i.e., sse existe uma matriz $s \times n$ \underline{B} tal que $\underline{A} = \underline{B}\underline{X}$.

Reiersol (1963) provou que os conceitos de identificabilidade e de estimabilidade linear são equivalentes para funções lineares. Este resultado, que enunciamos no próximo teorema, será aqui demonstrado a partir da caracterização de funções identificáveis estabelecida no teorema 2.4.

Teor. 2.7: A função $\underline{A}\underline{\beta}$ é identificável sse é linearmente estimável.

Dem. Como $(\underline{X}\underline{\beta})' \sigma^2$ é uma função identificante e $\underline{A}\underline{\beta}$ é uma função independente de σ^2 , o teorema 2.4 diz-nos que $\underline{A}\underline{\beta}$ é identificável sse existe uma função f

Reiersol
Teor. 2

tal que $\underline{A}\underline{\beta} = \underline{f}(\underline{X}\underline{\beta})$, $\forall \underline{\beta} \in \mathbb{R}^p$. Neste sentido, a implicação linearmente estimável \longrightarrow identificável é imediata.

Para provar a proposição recíproca, desdobremos \mathbb{R}^p na soma directa do subespaço $N(\underline{X})$ com o subespaço G gerado por uma dada inversa generalizada de \underline{X} , i.e., $\mathbb{R}^p = N(\underline{X}) + G$. Como a transformação \underline{X} restringida, $\underline{X}: G \rightarrow M(\underline{X})$ é um isomorfismo (transformação linear bijectiva), designemos, para todo o $\underline{\alpha} \in M(\underline{X})$, como $\underline{T}\underline{\alpha}$ o elemento de G tal que $\underline{\alpha} = \underline{X}\underline{T}\underline{\alpha}$. Assim, $\underline{f}(\underline{\alpha}) = \underline{f}(\underline{X}\underline{T}\underline{\alpha}) = \underline{A}\underline{T}\underline{\alpha}$, por hipótese, o que prova a linearidade da restrição de \underline{f} a $M(\underline{X})$. Consequentemente

$$\underline{A}\underline{\beta} = \underline{f}(\underline{X}\underline{\beta}) = \underline{A}\underline{T}\underline{X}\underline{\beta} \quad , \quad \forall \underline{\beta} \in \mathbb{R}^p$$

o que traduz, pela proposição 2.1, a estimabilidade linear de $\underline{A}\underline{\beta}$. Δ

Este resultado permite, pois, mudar a linguagem usual da teoria dos mínimos quadrados. Em particular, a condição de $\underline{a}'_i \underline{\beta}$, $i=1, \dots, s$ ser identificável é necessária e suficiente para a unicidade do seu BLUE, $\underline{a}'_i \hat{\underline{\beta}}$, onde $\hat{\underline{\beta}}$ é qualquer solução das equações normais $\underline{X}'\underline{X}\hat{\underline{\beta}} = \underline{X}'\underline{y}$. Todos os critérios de estimabilidade linear são assim critérios de identificabilidade para funções lineares.

O resultado expresso no teor. 2.7 mostra ainda, na sua aplicação ao modelo linear com celas "missing" formulado em termos de restrições no vector média das celas, que o conceito geral de conectividade de Murray and Smith (1985) - definido em termos da estimabilidade linear de $\underline{\mu}$ na experiência envolvendo \underline{D} e o modelo associado $\underline{\mu} \in N(\underline{W}')$ - pode ser expresso em termos da identificabilidade de $\underline{\mu}$. A nossa demonstração da equivalência entre a identificabilidade de $\underline{\mu}$ e o carácter isomorfo de $M(\underline{V})$ e \mathbb{R}^r , feita ao longo do exemplo I de 2.1.2, pode assim ser usada como alternativa para provar o critério de conectividade de Murray and Smith. O conceito de conectividade, originalmente ligado à localização das celas observadas em tabelas bidimensionais e relacionado com a estimabilidade de certas combinações lineares dos parâmetros do modelo de efeitos principais, passa desta forma a ser sinónimo de identificabilidade do vector média de todas as celas, conceito este que, por sua vez, depende, como vimos, da configuração observada e do modelo linear postulado.

A equivalência mencionada no teorema 2.7 permite ainda compreender o porquê da restrição do conceito de estimabilidade de funções lineares ao de estimabilidade linear. Por conjugação dos teoremas 2.6 e 2.7 fica demonstrado que:

Teor. 2.8: A função $\underline{A}\underline{\beta}$ é identificável sse for estimável.

Este resultado evidencia que não pode haver nenhuma função linear que seja estimável sem ser linearmente estimável, conforme foi referido por Bunke and Bunke (1974) no seu estudo sobre identificabilidade de modelos lineares mul

Bunke
1974, p. 3

tivariados. A consideração de funções não lineares permite, no entanto, obter novos exemplos comprovativos de que a identificabilidade não implica estimabilidade (e.g. a função $(\eta + \alpha_1 + \gamma_1)^2$ citada no exemplo I onde σ^2 é desconhecido).

Em algumas situações, o domínio de $\underline{\beta}$ está restringido a um subconjunto de \mathbb{R}^p , frequentemente traduzido pelo subespaço $N(\underline{H})$ onde \underline{H} é uma matriz $t \times p$ de posto $t < p$. Em consonância com a análise do modelo não restringido, a identificabilidade de $\underline{\beta}$ é condição necessária e suficiente para que o sistema linear $\underline{X}\underline{\beta} = \bar{\underline{\mu}}$ com $\underline{\beta} \in N(\underline{H})$, para cada valor da média $\bar{\underline{\mu}}$ das observações, seja determinado. Ora, de acordo com o resultado algébrico contido em Scheffé (1959, sec. 1.4) - proposição A.3 do apêndice -, o referido sistema, definido por $\underline{G}\underline{\beta} = \bar{\underline{\mu}}^*$, onde $\underline{G} = (\underline{X}'\underline{H}')'$ e $\bar{\underline{\mu}}^* = (\bar{\underline{\mu}} \ 0'_{(t)})'$, é consistente e tem solução única para todo o $\bar{\underline{\mu}} \in M(\underline{X})$ sse $M(\underline{X}') \cap M(\underline{H}') = \{0\}$ e $r(\underline{G}) = p$. Deste modo, assegurada a consistência das $n+t$ equações lineares do sistema, $\underline{\beta}$ é identificável sse \underline{G} tem posto completo.

Quando alguma dessas condições não é verificada, a classe de pontos observacionalmente equivalentes é definida analogamente ao modelo não restringido, com a diferença de que Ξ é encarado como o produto cartesiano $N(\underline{H}) \times \mathbb{R}^+$. Como σ^2 é identificável, a classe de pontos observacionalmente equivalentes a $\underline{\beta}_0 \in N(\underline{H})$ é então definida por

$$[\underline{\beta}_0] = \{\underline{\beta}_1 \in N(\underline{H}) : \underline{\beta}_1 = \underline{\beta}_0 + \underline{\alpha}, \underline{\alpha} \in N(\underline{X})\}$$

onde o espaço nulo de \underline{X} é um subespaço $(p-r)$ -dimensional de \mathbb{R}^p .

Numa linguagem de espaços vectoriais (vide, e.g., Nering [1964, pgs. 183/4]), $[\underline{\beta}_0]$ é um subespaço transladado de dimensão inferior a $p-r$, a sua dimensão no modelo irrestrito. Com efeito, se $\underline{X}\underline{\beta}_0 = \bar{\underline{\mu}}_0$, então $[\underline{\beta}_0]$ é o conjunto das soluções do sistema linear não homogêneo em $\underline{\beta} \in \mathbb{R}^p$,

$\underline{G}\underline{\beta} = (\bar{\underline{\mu}}_0' \ 0')'$ e portanto, corresponde para $\bar{\underline{\mu}}_0$ fixo a um subespaço transladado com a dimensão de $N(\underline{G})$ que é $p-r(\underline{G}) < p-r$. O espaço $N(\underline{H})$ fica assim particionado no conjunto $N(\underline{H})/\sim$ de subespaços transladados paralelos de dimensão $p-r(\underline{G})$, constituindo ele próprio um espaço vectorial, relativamente a operações lineares apropriadamente definidas, de dimensão $r(\underline{G})$ - vide Halmos (1958, pgs 33/4). Note-se que a visualização acima de $[\underline{\beta}_0]$, com $\underline{\beta}_0 \in N(\underline{H})$, revela claramente que essa classe é formada apenas por $\underline{\beta}_0$ quando e só quando $r(\underline{G}) = p$.

Os resultados expressos nos teoremas 2.7 e 2.8 mantêm-se formalmente para o modelo restrito. A única diferença está em que as funções vectoriais em causa passam a ser encaradas como transformações lineares de \mathbb{R}^{p-t} dada a sua isomorfia com $N(\underline{H})$ (Halmos [1958, pg 15]), o que redundará em modificações no critério de estimabilidade linear da proposição 2.1. Reiersol (1963) estabelece es

se novo critério que reproduzimos em seguida para uma exposição mais completa. Com base nele provaremos a aplicabilidade das afirmações contidas nos teoremas 2.7 e 2.8 a este novo contexto, seguindo o mesmo tipo de argumento que foi usado para o modelo irrestrito.

Teor. 2.9: Sob a condição $\underline{\beta} \in N(\underline{H})$, a função $\underline{A}\underline{\beta}$ é linearmente estimável sse $\underline{a}_i \in M(\underline{G}')$, $i=1, \dots, s$.

Dem. A suposição de que as linhas de \underline{A} pertencem a $M(\underline{G}')$ significa que existe uma matriz \underline{B}^* , $s \times (n+t)$ tal que $\underline{A} = \underline{B}^* \underline{G}$. Particionando \underline{B}^* nas sub-matrizes \underline{B} , $s \times n$, e \underline{C} , $s \times t$, obtemos para todo o $\underline{\beta} \in N(\underline{H})$, $\underline{A}\underline{\beta} = \underline{B}\underline{X}\underline{\beta} = E(\underline{B}\underline{Y})$, o que traduz a estimabilidade linear de $\underline{A}\underline{\beta}$.

Reciprocamente, seja $\underline{A}\underline{\beta}$, $\underline{\beta} \in N(\underline{H})$, linearmente estimável e $\underline{B}\underline{Y}$ um seu estimador linear centrado. Então $(\underline{A} - \underline{B}\underline{X})\underline{\beta} = \underline{0}_{(s)}$, $\forall \underline{\beta} \in N(\underline{H})$. Designando por \underline{b}'_i os vectores linha de \underline{B} , esta última condição significa que $(\underline{a}'_i - \underline{b}'_i \underline{X})\underline{\beta} = 0$, $\forall \underline{\beta} \in N(\underline{H})$, $i=1, \dots, s$, o que implica que os vectores $\underline{a}_i - \underline{X}'\underline{b}_i$, $i=1, \dots, s$ pertencem ao complemento ortogonal de $N(\underline{H})$, i.e., a $M(\underline{H}')$ (vide proposição A.1 do apêndice). Assim, existem vectores $\underline{c}_i \in \mathbb{R}^t$ tais que $\underline{a}_i = \underline{X}'\underline{b}_i + \underline{H}'\underline{c}_i = \underline{G}'\underline{b}_i^*$, onde $\underline{b}_i^* = (\underline{b}'_i \underline{c}'_i)'$, $i=1, \dots, s$, o que revela que $\underline{a}_i \in M(\underline{G}')$, $i=1, \dots, s$. Δ

Teor. 2.10. Sob $\underline{\beta} \in N(\underline{H})$, a função $\underline{A}\underline{\beta}$ é identificável sse for estimável.

Dem. A implicação estimabilidade \Rightarrow identificabilidade segue-se do teorema 2.6. O recurso aos teoremas 2.4 e 2.9 prova a equivalência entre identificabilidade e estimabilidade linear, condição esta que implica obviamente a estimabilidade. Δ

Naturalmente que toda a função $\underline{A}\underline{\beta}$ identificável no modelo irrestrito é também identificável num modelo restrito, enquanto que a proposição recíproca não é necessariamente verdadeira. A própria função que resulta de $\underline{A}\underline{\beta}$, suposta identificável em \mathbb{R}^p , pela incorporação das restrições $\underline{H}\underline{\beta} = \underline{0}_{(t)}$ pode deixar de ser identificável no modelo irrestrito. De facto, pelo que vimos acima, a função identificável $\underline{A}\underline{\beta}$ quando restringida a $N(\underline{H})$ pode definir-se por $\underline{B}\underline{X}\underline{\beta} + \underline{C}\underline{H}\underline{\beta}$. Esta expressão revela que a função pode não ser identificável em \mathbb{R}^p se a função $\underline{H}\underline{\beta}$ não é identificável (sê-lo-á obrigatoriamente em caso de identificabilidade de $\underline{H}\underline{\beta}$). Para um exemplo, retomemos o modelo ANOVA II considerado no exemplo I de 2.1.2 e consideremos a função $\phi(\underline{\beta}) = \mu + \alpha_1 + \frac{1}{b} \sum_{j=1}^b \gamma_j$, claramente identificável. Impondo a restrição $\sum_{j=1}^b \gamma_j = 0$, cuja função definida é não identificável, a função $\phi(\underline{\beta})$ converte-se em $\mu + \alpha_1$. A função transformada é identificável no modelo restrito mas não é identificável no modelo irrestrito.

A relevância destas considerações não se confina naturalmente a modelos lineares para dados contínuos. Em muitas situações com dados categorizados tem interesse considerar modelos lineares para o vector das probabilidades multinomiais. Um exemplo é dado precisamente pelo problema de dados incompletos detalhado no cap. 3, onde certos dados são categorizados completamente e os outros apenas de forma parcial. No caso de estes se estruturarem em várias tabelas de frequências, fica estabelecido um modelo linear relacionando o vector paramétrico, digamos $\bar{\theta}$, de todas as frequências multinomiais com o vector $\underline{\theta}$ das probabilidades das celas da tabela original correspondente à classificação completa.

Se todas as tabelas forem parcialmente categorizadas, pode acontecer que o sistema linear $X\underline{\theta} = \bar{\theta}$, com $\mathbf{1}'\underline{\theta} = 1$, para cada $\bar{\theta}$ não tenha solução única, implicando assim a não identificabilidade e a não estimabilidade de $\underline{\theta}$. A mesma situação pode (ou não) ocorrer se $\underline{\theta}$ satisfaz adicionalmente uma hipótese linear, digamos $\underline{\theta} \in M(H)$. Por exemplo, consideremos o caso simples de uma tabela de contingência original 2×2 com a categorização observada definida pela matriz 3×4 D de linhas $(1 \ 0 \ 0 \ 0)$, $(0 \ 1 \ 0 \ 0)$ e $(0 \ 0 \ 1 \ 1)$, onde as frequências subjacentes estão ordenadas lexicograficamente. Representando $E = \{ \underline{\theta} = (\theta_{11} \ \theta_{12} \ \theta_{21} \ \theta_{22}) : 0 < \theta_{ij} < 1, i, j = 1, 2, \mathbf{1}'\underline{\theta} = 1 \}$ o domínio do parâmetro indexante do modelo, suposto multinomial, para as frequências observadas e $\underline{\psi}_0 = D\underline{\theta}_0$ a imagem de um determinado $\underline{\theta}_0 \in E$ pela transformação linear D , então

$$[\underline{\theta}_0] = \{ \underline{\theta} \in E : D\underline{\theta} = \underline{\psi}_0 \} \neq \{ \underline{\theta}_0 \}$$

Admitamos agora que o interesse reside no ajustamento do modelo de homogeneidade marginal, definido por $\underline{\lambda}'\underline{\theta} = 0$, onde $\underline{\lambda}' = (0 \ 1 \ -1 \ 0)$. É fácil constatar que o sistema linear $(D' \ \mathbf{1}'_4 \ \underline{\lambda})'\underline{\theta} = (\underline{\psi}'_0 \ 1 \ 0)$ tem uma solução única para cada $\underline{\psi}_0$ - basta verificar que cada θ_{ij} é estimável - e, por conseguinte, o parâmetro passa a ser identificável sob a hipótese de homogeneidade marginal.

2.1.5. Formas de identificação do modelo

Como já foi referido, nos modelos não identificados existem, em geral, funções identificáveis que podem caracterizar igualmente a estrutura do modelo. A reparametrização do modelo com base numa dessas funções que induzem a partição identificante constitui um meio de obter um modelo identificado. Por exemplo, no modelo log-linear ordinário (vide exemplo II de 2.1.2) para o vector probabilidade das q celas de uma tabela de observações multinomiais, \underline{y} , a função de probabilidade depende de $\underline{\beta}$ através da função identificante $X_0\underline{\beta} = \log\theta$. O modelo $\underline{y} | N, \underline{\beta} \rightsquigarrow M(N, \exp(X_0\underline{\beta}))$ com $\mathbf{1}'_q \exp(S_0\underline{\beta}) = 1$

pode então ser convertido para o modelo equivalente $y|N, \theta \rightsquigarrow M(N, \theta)$ com $1'_{q-r} \theta = 1$ e $\log \theta \in N(W')$, onde W' é uma matriz $q \times (q-r)$ base do complemento ortogonal de $M(X_0)$.

Esta reparametrização identificante afigura-se extremamente conveniente quando o objectivo inferencial se centra na função identificante ou em funções dela. Quando a não identificabilidade do parâmetro original deriva da sobreparametrização da especificação adoptada para o modelo, há sempre uma possibilidade de reparametrização por redução da dimensionalidade do espaço paramétrico. Por exemplo, no modelo linear normal (exemplo I de 2.1.2), a existência de apenas $r < p$ elementos de β funcionalmente independentes justifica que possamos particionar a matriz X de especificação do modelo em $(X_1 X_2)$ onde X_1 é $(n \times r)$ de posto r . A sub-matriz X_2 é então definida por $X_2 = X_1 E$, onde as colunas da matriz $r \times (p-r) E$ descrevem os coeficientes das combinações lineares das colunas de X_1 definidoras das $p-r$ colunas de X_2 . Desta forma, o modelo linear original é convertido no modelo de posto completo, $y = X_1 \delta + u$, onde δ representa a função identificável $\beta_1 + E\beta_2$. Este tipo de reparametrização defronta-se, todavia, com o problema de interpretação do novo parâmetro, o qual pode mesmo ser desprovido de significado no contexto do problema.

Naturalmente que este tipo de reparametrizações não esgota as formas de obtenção de um modelo identificado. Vimos já que a função de verosimilhança de um modelo não identificável é, essencialmente para quaisquer conjuntos de dados, uniforme em cada classe de equivalência. Assim, uma realização amostral dá apenas informação sobre a classe a que pertence o verdadeiro valor do parâmetro, não conseguindo atribuir a este um valor pertencente àquela classe de equivalência. Daí que a imposição de restrições no espaço paramétrico, visando a singularização de cada classe de equivalência, constitua um processo de identificação do modelo, aliás, o mais aplicado no contexto clássico.

Em ordem a clarificar as condições nas restrições exigidas pela identificação do modelo, suponhamos que ψ é uma função identificante de modo que $[\theta_0] = \{\theta \in \Xi : \psi(\theta) = \psi_0\}$, onde $\psi_0 = \psi(\theta_0)$ e definamos o espaço paramétrico restringido por $\bar{\Xi} = \{\theta \in \Xi : \lambda(\theta) = 0\}$. Nestas condições, como referimos em 2.1.1., $\bar{\Xi}$ é identificável sse cada $[\theta_0]$ é singularizado por $\lambda(\theta) = 0$ isto é, sse o sistema $\begin{cases} \psi(\theta) = \psi_0 \\ \lambda(\theta) = 0 \end{cases}$ tem solução única em cada $[\theta_0] \in \Xi/\sim$. É fácil constatar desde já que o objectivo pretendido não pode ser preenchido com uma função $\lambda(\theta)$ identificável. De facto, se $\lambda(\theta)$ for identificável, pela definição 2.4 haverá classes de equivalência incompatíveis com o valor 0 para $\lambda(\theta)$ e, por conseguinte, o referido sistema não será consistente para todas as classes de Ξ/\sim . Note-se ainda que as classes compatíveis, em número de pelo menos uma (haverá apenas uma se $\lambda(\theta)$ for adicionalmente suficiente) não terão a sua dimensionalidade reduzida, pelo que $\bar{\Xi}/\sim = \{[\theta] \in \Xi/\sim : \lambda([\theta]) = 0\}$, o que significa que o problema de não identificabilidade se manterá em $\bar{\Xi}$.

É um argumento deste tipo que justifica que a dimensionalidade de cada classe de equivalência correspondente ao modelo log-linear para dados Poisson (exemplo II de 2.1.2) se mantenha em cada classe de equivalência, compatível com as restrições naturais, relativa ao correspondente modelo log-linear para dados multinomiais. A não identificabilidade deste último modelo pode deduzir-se directamente em termos do sistema não linear

$$\begin{cases} \underline{X}_0 \underline{\beta} = \log \underline{\theta}_0 \\ \underline{1}' \exp(\underline{X}_0 \underline{\beta}) = 1 \end{cases}, \text{ equivalente a } \frac{\exp(\underline{X}_0 \underline{\alpha})}{\underline{1}' \exp(\underline{X}_0 \underline{\alpha})} = \underline{\theta}_0, \text{ onde } \underline{\beta} = (\eta \ \underline{\alpha}')' \text{ e } \underline{X}_0 = \begin{pmatrix} \underline{1} & \underline{X} \\ \underline{q} & \end{pmatrix},$$

cujas soluções não é claramente única, para cada $\underline{\theta}_0$.

O processo de identificação de \bar{E} através da imposição das restrições consistentes, $\underline{\lambda}(\underline{\theta}) = (\lambda_1(\underline{\theta}), \dots, \lambda_t(\underline{\theta}))' = \underline{0}_{(t)}$, com $\lambda_i(\underline{\theta})$, $i=1, \dots, t$ não identificáveis, pode ser visto em termos sequenciais. Suponhamos, sem quebra de generalidade, que o ponto escolhido para representar cada classe de equivalência pertence a \bar{E} e denotemos por E_j o espaço paramétrico restringido pelas condições $\lambda_i(\underline{\theta}) = 0$, $i=1, \dots, j$ ($E_t \equiv \bar{E}$). A imposição, em primeiro lugar, de $\lambda_1(\underline{\theta}) = 0$ implica que em cada $[\underline{\theta}_0] \in E/\sim$ haja apenas alguns pontos compatíveis com tal restrição. Designando $[\underline{\theta}_0]_1 = [\underline{\theta}_0] \cap E_1$, a afirmação anterior significa que a dimensionalidade de $[\underline{\theta}_0]_1$ é inferior à de $[\underline{\theta}_0]$. Impondo agora $\lambda_2(\underline{\theta}) = 0$, obtém-se $[\underline{\theta}_0]_2$ à custa de $[\underline{\theta}_0]_1$ por eliminação dos pontos incompatíveis com tal restrição e assim sucessivamente. Se a dimensionalidade de cada $[\underline{\theta}_0]_j \in E_j/\sim$ é superior a 1, impõe-se nova restrição, $\lambda_{j+1}(\underline{\theta}) = 0$, de modo a obter a classe de dimensionalidade inferior, $[\underline{\theta}_0]_{j+1}$. Este processo sequencial pára quando se conseguir a singularização de cada elemento de E/\sim . Se isto acontece após a adição de $\lambda_t(\underline{\theta}) = 0$, então $E_t/\sim = \{\{\underline{\theta}\}, \underline{\theta} \in E_t\}$, estabelecendo a desejada identificabilidade do modelo.

Em várias situações, $[\underline{\theta}_0]$ é definida por equações lineares e as restrições impostas são definidas em termos de funções lineares. Concretamente, sejam \underline{C} e $\underline{\Lambda} = (\lambda_1 \dots \lambda_t)'$ as matrizes representando as transformações lineares de $\underline{\theta}$ definidoras, respectivamente, das classes de equivalência e das restrições. Neste caso, os resultados matemáticos sobre sistemas lineares permitem-nos definir condições necessárias e suficientes para a unicidade da solução de

$$\begin{cases} \underline{C}\underline{\theta} = \underline{C}\underline{\theta}_0 \equiv \underline{\psi}_0 \\ \underline{\Lambda}\underline{\theta} = \underline{0}_{(t)} \end{cases}. \text{ Em particular, a não identificabilidade de } \underline{\Lambda}\underline{\theta} \text{ é condição necessária e suficiente de consistência do referido sistema.}$$

De facto, pelos resultados de 2.1.4, $\underline{\Lambda}\underline{\theta}$ é não identificável sse é não (linearmente) estimável sse $\lambda_i \notin M(\underline{C}')$, $i=1, \dots, t$. Isto é, $\underline{\Lambda}\underline{\theta}$ é não identificável sse nenhum vector não nulo de $M(\underline{\Lambda}')$ é vector de $M(\underline{C}')$, ou equivalentemente, sse $M(\underline{\Lambda}') \cap M(\underline{C}') = \{0\}$. Mas, pela proposição A.3 (apêndice), esta condição é necessária e suficiente para a consistência do referido sistema linear, para todo

$\underline{\psi}_0 \in M(\underline{C})$. Neste contexto, a condição de as linhas de $\underline{\Lambda}$ e de \underline{C} serem mutuamente linearmente independentes, citada já em 2.1.4 para o modelo linear normal, não é mais do que a condição de não identificabilidade da função $\underline{\Lambda}\underline{\theta}$, como foi destacado por Reiersol (1963). Uma vez assegurada a consistência, o sistema terá solução única se o posto de $(\underline{C}'\underline{\Lambda}')'$ é igual à dimensionalidade de $\underline{\theta}$, digamos p . Se o posto de \underline{C} é $r < p$, sendo assim suficiente definir $t = p-r$ restrições, o sistema é então consistente e de solução única sse $\underline{\Lambda}\underline{\theta}$ e não identificável e $\underline{\Lambda}$ tem posto completo. Neste pressuposto, $\underline{\theta}$ exprime-se como uma função linear da função identificante $\underline{\psi} = \underline{C}\underline{\theta}$, mais precisamente, como $\underline{\theta} = (\underline{C}'\underline{C} + \underline{\Lambda}'\underline{\Lambda})^{-1}\underline{C}'\underline{\psi}$ o que evidencia a identificabilidade de $\underline{\theta}$.

Para ilustrar o processo sequencial de identificação do espaço paramétrico, consideremos o modelo ANOVA de 2 factores, sem interacção e com uma observação em cada uma das ab celas. Suponhamos, para facilidade de exposição, que σ^2 é conhecido, o que não altera a questão de fundo, já que a não identificabilidade do modelo linear normal se prende com a não identificabilidade do parâmetro $\underline{\beta}$. O modelo é assim especificado pela matriz \underline{X} $n \times p$ com $n = ab$ e $p = a+b+1$, traduzindo as relações $E(y_{ij} | \underline{\beta}) \equiv \mu_{ij} = \eta + \alpha_i + \gamma_j$, $i=1, \dots, a$; $j = 1, \dots, b$, onde $\underline{\beta} = (\eta \ \alpha_1 \ \dots \ \alpha_a \ \gamma_1 \ \dots \ \gamma_b)'$ $\in \Xi \equiv \mathbb{R}^p$. O posto de \underline{X} é $r = p-2$ e para um dado $\underline{\beta}_0 = (\eta^0 \ \alpha_1^0 \ \dots \ \alpha_a^0 \ \gamma_1^0 \ \dots \ \gamma_b^0) \in \Xi$, é fácil verificar que

$$[\underline{\beta}_0] \equiv \{ \underline{\beta} \in \Xi : \underline{X}\underline{\beta} = \underline{X}\underline{\beta}_0 \} = \{ \underline{\beta} \in \Xi : \eta = \eta^0 + c + d; \alpha_i = \alpha_i^0 - c, \forall i; \gamma_j = \gamma_j^0 - d, \forall j; c, d \in \mathbb{R} \}.$$

Para uma simplificação adicional consideremos $a = b = 2$, pelo que $\underline{X} = (\underline{x}_1 \ \underline{x}_2 \ \dots \ \underline{x}_5)'$ com $\underline{x}_1 = \underline{1}_4$, $\underline{x}_2 = (1 \ 1 \ 0 \ 0)'$, $\underline{x}_3 = (0 \ 0 \ 1 \ 1)'$, $\underline{x}_4 = (1 \ 0 \ 1 \ 0)'$ e $\underline{x}_5 = (0 \ 1 \ 0 \ 1)'$. Imponhamos a restrição $\underline{\lambda}'_1 \underline{\beta} = 0$, onde $\underline{\lambda}'_1 = (0 \ 1 \ 1 \ 0 \ 0)$, que é apropriada, já que $\underline{\lambda}_1 \notin M(\underline{X})$ (teor. 2.6). Então Ξ_1 é o subespaço tetradiimensional de \mathbb{R}^5 , $N(\underline{\lambda}'_1)$ e $[\underline{\beta}_0]_1 = [\underline{\beta}_0] \cap N(\underline{\lambda}'_1)$. A dimensionalidade das classes de equivalência passa de $p-r = 2$, já que cada $[\underline{\beta}_0]$ é um subespaço transladado de dimensão igual à de $N(\underline{X})$, para $p-(r+1) = 1$, pelo facto do sistema linear $\underline{\beta} \in [\underline{\beta}_0] \cap N(\underline{\lambda}'_1)$ ser consistente e, em consequência, cada $[\underline{\beta}_0]_1$ corresponder a um subespaço transladado de dimensão igual à de $N(G_1)$, onde $G_1 = (\underline{X}'\underline{\lambda}_1)'$ tem posto $r+1$ (vide sub-secção 2.1.4).

A adição desta restrição conduz à eliminação dos pontos de $[\underline{\beta}_0]$ correspondentes a $c \neq 0$ e assim para $\underline{\beta}_0 \in \Xi_1$

$$[\underline{\beta}_0]_1 = \{ \underline{\beta} \in \Xi : G_1 \underline{\beta} = G_1 \underline{\beta}_0 \} = \{ \underline{\beta} \in \Xi : \eta = \eta^0 + d; \alpha_i = \alpha_i^0, i=1,2; \gamma_j = \gamma_j^0 - d, j=1,2, d \in \mathbb{R} \}$$

Para conseguir a identificação do modelo falta, pois, adicionar mais uma restrição, necessariamente independente linearmente da anterior - caso contrário, o conjunto Ξ_1/\sim permaneceria imutável. Suponhamos então que pretendíamos im-

por $\lambda_3' \beta = 0$ onde $\lambda_3' = (0 \ 0 \ 0 \ 1 \ -1)$. Como $\lambda_3 = G_1' \alpha$ com $\alpha = (1 \ -1 \ 0 \ 0 \ 0)'$, $\lambda_3' \beta$ é identificável (teors. 2.8 e 2.9 em 2.1.3) e, portanto, inadequada para o fim em vista. Note-se, a este respeito, que $\lambda_3' \beta = \mu_{11} - \mu_{12}$, o que mostra que seriam removidas de E_1/\sim todas as classes $[\beta_0]_1$ para as quais as duas primeiras componentes de $X\beta_0$ são diferentes. As restantes, onde se inclui a classe $N(G_1)$, seriam mantidas sem redução de dimensionalidade e assim, no subespaço tridimensional de \mathbb{R}^5 , $N(\Lambda)$ com $\Lambda = (\lambda_1 \ \lambda_3)'$, manter-se-ia o problema da não identificabilidade. A adição de $\lambda_2' \beta = 0$ com $\beta_2' = (0 \ 0 \ 0 \ 1 \ 1)$ já resolve o problema. De facto, a não identificabilidade de $\lambda_2' \beta$ segue-se de λ_2 não pertencer a $M(G_1')$ e, por outro lado, a imposição de $\lambda_2' \beta = 0$ remove de cada $[\beta_0]_1 \in E_1/\sim$ os pontos correspondentes a $d \neq 0$, pelo que

$$[\beta_0]_2 = [\beta_0]_1 \cap N(\lambda_2') = \{\beta \in E: G_2 \beta = G_2 \beta_0, G_2 = (G_1' \ \lambda_2)'\} = \{\beta_0\}.$$

A identificabilidade do espaço paramétrico restringido, $N(H)$, $H = (\lambda_1 \ \lambda_2)'$, pode verificar-se alternativamente em termos da formulação das restrições em equações livres. Designando W uma matriz $p \times r$ base do complemento ortogonal de $M(H')$, então $H\beta = 0_{(p-r)}$, $r(H) = p-r \iff \beta = W\delta$, $\delta \in \mathbb{R}^r$, o que permite funcionar directamente em termos do espaço paramétrico reduzido \mathbb{R}^r . Para um dado $\delta_0 \in \mathbb{R}^r$, temos então $[\delta_0] = \{\delta \in \mathbb{R}^r: Z\delta = Z\delta_0\} = \{\delta_0\}$ pois $Z = XW$ tem posto completo. No exemplo concreto analisado podemos tomar

$$W' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \text{ correspondente à escolha de } \delta = (\eta \ \alpha_1 \ \gamma_1)', \text{ e assim}$$

$$Z' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}.$$

Naturalmente que a identificação do modelo não exige que as 2 restrições linearmente independentes sejam aquelas que considerámos acima e que correspondem às restrições $\sum_i \alpha_i = \sum_j \gamma_j = 0$ habitualmente usadas na literatura dos modelos lineares, e igualmente, nos modelos log-lineares para dados categorizados, dada a semelhança estrutural com os modelos ANOVA. Essas restrições têm a vantagem de permitir interpretar os efeitos principais no modelo restringido como desvios dos efeitos correspondentes no modelo irrestrito de cada nível de cada factor relativamente ao efeito médio do respectivo factor. Ou equivalentemente, como desvios da média de cada nível de cada factor relativamente à média geral.

Obtido um modelo identificável, a arbitrariedade da estimação desaparece naturalmente. Por exemplo, no que concerne ao modelo linear normal identificado pelas $p-r$ restrições $H\beta = 0$, o sistema das equações normais restringidas

$$\begin{pmatrix} \underline{\underline{X}}' \underline{\underline{X}} & \underline{\underline{H}}' \\ \underline{\underline{H}} & \underline{\underline{0}} \end{pmatrix} \begin{pmatrix} \hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}} \\ \hat{\underline{\underline{\rho}}}_{\underline{\underline{H}}} \end{pmatrix} = \begin{pmatrix} \underline{\underline{X}}' \underline{\underline{y}} \\ \underline{\underline{0}} \end{pmatrix}$$

onde $\hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}}$ é o vector dos multiplicadores Lagrangianos, tem solução única devido à não singularidade da matriz dos coeficientes - note-se que $M(\underline{\underline{X}}' \underline{\underline{X}}) = M(\underline{\underline{X}}')$.

Neste ponto, torna-se importante compreender que as inferências sobre funções $\underline{\underline{A}}\underline{\underline{\beta}}$ identificáveis no modelo irrestrito não variam quando se funciona em termos do modelo identificável $N(\underline{\underline{H}})$. Com efeito, a não identificabilidade de $\underline{\underline{H}}\underline{\underline{\beta}}$ é equivalente a afirmar que $M(\underline{\underline{X}}' \underline{\underline{X}}) \cap M(\underline{\underline{H}}') = \{0\}$. Recorrendo à proposição A.3. do apêndice, esta condição é necessária e suficiente para a consistência do sistema

do sistema $\begin{pmatrix} \underline{\underline{X}}' \underline{\underline{X}} \\ \underline{\underline{H}}' \end{pmatrix} \hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}} = \begin{pmatrix} \underline{\underline{X}}' \underline{\underline{y}} \\ \underline{\underline{0}} \end{pmatrix}$, $\forall \underline{\underline{y}} \in \mathbb{R}^n$, o qual corresponde ao sistema

das equações normais restringidas com $\hat{\underline{\underline{\rho}}}_{\underline{\underline{H}}} = \underline{\underline{0}}$, $\forall \underline{\underline{y}} \in \mathbb{R}^n$. Isto significa que toda a solução $\hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}}$ das equações normais restringidas é solução das respectivas equações não restringidas sse as restrições são definidas por funções não identificáveis. Deste modo, a soma dos quadrados residual $(\underline{\underline{y}} - \underline{\underline{X}}\hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}})'(\underline{\underline{y}} - \underline{\underline{X}}\hat{\underline{\underline{\beta}}}_{\underline{\underline{H}}})$ coincide com a respectiva soma de quadrados para o modelo irrestrito, o que, aliás, deixa transparecer a impossibilidade das restrições $\underline{\underline{H}}\underline{\underline{\beta}} = \underline{\underline{0}}$ serem testadas.

Admitamos então que o objectivo é testar a hipótese $\underline{\underline{A}}\underline{\underline{\beta}} = \underline{\underline{0}}$, suposta consistente com o modelo $N(\underline{\underline{H}})$. Por aplicação ao novo sistema de equações normais

$$\begin{pmatrix} \underline{\underline{X}}' \underline{\underline{X}} & \underline{\underline{A}}' & \underline{\underline{H}}' \\ \underline{\underline{A}} & \underline{\underline{0}} & \underline{\underline{0}} \\ \underline{\underline{H}} & \underline{\underline{0}} & \underline{\underline{0}} \end{pmatrix} \begin{pmatrix} \hat{\underline{\underline{\beta}}}_{\underline{\underline{A}, \underline{\underline{H}}}} \\ \hat{\underline{\underline{\rho}}}_{\underline{\underline{A}}} \\ \hat{\underline{\underline{\rho}}}_{\underline{\underline{H}}} \end{pmatrix} = \begin{pmatrix} \underline{\underline{X}}' \underline{\underline{y}} \\ \underline{\underline{0}} \\ \underline{\underline{0}} \end{pmatrix}$$

de um argumento similar ao da proposição A.3 (vide, e.g. Reiersol [1963]), concluímos que toda a solução $\hat{\underline{\underline{\beta}}}_{\underline{\underline{A}, \underline{\underline{H}}}}$ é solução do sistema de equações normais relativo ao modelo $N(\underline{\underline{A}})$. Por conseguinte, a estatística F usual para testar a hipótese $\underline{\underline{\beta}} \in N(\underline{\underline{A}})$ não difere com a imposição de restrições visando a identificação do modelo. Deve, contudo, sublinhar-se que a hipótese a testar quando modificada pelas restrições $\underline{\underline{H}}\underline{\underline{\beta}} = \underline{\underline{0}}$ pode deixar de ser testável no modelo irrestrito, dada a não identificabilidade de $\underline{\underline{H}}\underline{\underline{\beta}}$.

Fica assim patente o duplo papel que as funções não identificáveis desempenham, o de assegurar a construção de um modelo identificável e o de não "perturbar" as inferências sobre as funções identificáveis. Estes aspectos permitem compreender os dois meios usuais de análise "clássica" do modelo linear normal quando formulado em termos de equações livres: conversão prévia para um modelo identificável, ou uso de uma estimativa do parâmetro original, obtida à cus

ta de uma inversa generalizada $X'X$, ou de imposição, unicamente ao sistema de equações normais, de restrições não identificáveis suficientes para o tornar determinado.

2.2. A identificabilidade no contexto Bayesiano

2.2.1 - Considerações e instrumentos preliminares

1. O conceito de identificabilidade na óptica Bayesiana não se tem mostrado pacífico, já que se podem encontrar na literatura estatística várias definições que não são equivalentes àquela mencionada na secção anterior e que é habitualmente usada pelos partidários da inferência clássica.

A razão da controvérsia materializada na existência de vários conceitos alternativos reside na própria estrutura do modelo Bayesiano, cuja formação exige, para além do modelo clássico, a consideração de uma estrutura mensurável \mathcal{B} e de uma medida ν no espaço paramétrico.

Como afirma Zellner (1971): "...there is a need to broaden the concept of identification to allow for the more general kind of prior information used in Bayesian analysis". Nesta óptica, Zellner define no caso dominado a não identificabilidade de um modelo paramétrico relativamente a outro modelo através da igualdade das respectivas probabilidades marginais das observações. Com a mesma preocupação e ligando o conceito a aspectos inferenciais, Leamer (1978) usa a expressão "identified in probability" para caracterizar um modelo para o qual a distribuição à posteriori é própria, independentemente do conjunto particular de dados obtido. Drèze (1972), na sua análise do modelo econométrico de equações simultâneas, apresenta uma outra visão com base na interpretação de que "From the Bayesian viewpoint, classical identification theory is really concerned with local uniqueness of posterior modes".

Uma outra contrapartida Bayesiana à definição usada na secção anterior (vide, e.g., Deistler and Seifert [1978]) consiste em considerar identificável todo o modelo estatístico para o qual a aplicação $\theta \rightarrow P_{\theta}(A)$, $\forall A \in \mathcal{A}$ é injectiva a menos de conjuntos ν -nulos. Neste sentido, uma função paramétrica \mathcal{B} -mensurável é identificável se é identificável no sentido usado em 2.1, ν - quase em toda parte. Outros autores, tais como Florens et al. (1983, cap. 4) utilizam o termo identificabilidade no sentido em que ele é utilizado para descrever uma relação entre duas sub- σ -álgebras em qualquer espaço de probabilidade. Mais concretamente, o modelo Bayesiano é considerado identificável se a σ -álgebra \mathcal{B} coincide essencialmente com a σ -álgebra gerada pelas funções \mathcal{B} -mensuráveis $P_{\theta}(A)$, $\forall A \in \mathcal{A}$.

Nesta controvérsia dentro da literatura Bayesiana sobre a definição do conceito de identificabilidade, destacam-se ainda os apologistas de que tal conceito é uma característica do modelo $(Y, \mathcal{A}\{P_\theta: \theta \in \Xi\})$, sendo pois independente da estrutura de medida definida sobre θ . Na linha de Barankin (1960), Kadane (1974), no âmbito do caso dominado, é categórico: "... identification is a property of the likelihood function and is the same whether considered classically or from the Bayesian approach".

A idéia de identificabilidade subjacente nesta secção está em consonância com essa última visão, partilhada ainda por Drèze (1974). Contudo, um dos principais objectivos desta secção está em pôr em relevo, num quadro relativamente geral (e não apenas em casos específicos como aquele em que o espaço amostral é discreto), a dualidade entre conceitos correspondentes da teoria da suficiência (amostral) Bayesiana e da teoria da identificabilidade. Com esse fim, generalizaremos em 2.2.2, na linha do tratamento de Picci (1974), os conceitos de identificabilidade e suficiência paramétricas de modo que as definições dadas em 2.1 sejam válidas quase sempre relativamente à medida à priori ν usada. Isto permitir-nos-á adicionalmente estabelecer relações entre tais conceitos e os dois de correspondentes conceitos da teoria da suficiência clássica.

Esses conceitos estendidos serão rotulados com o termo adicional "Bayesiano" em ordem a diferenciá-los dos conceitos usuais que não envolvem o tecnicismo do " ν -quase em toda a parte". Naturalmente que é fácil arranjar exemplos de modelos não identificáveis que são Bayesianamente identificáveis no sentido acima (vide, e.g., Florens et al. [1983], cap 4]). Para os modelos Bayesianos usuais, todavia, esses dois conceitos coincidem frequentemente, como o ilustra o seguinte modelo para o exemplo III de 2.1.2:

Exemplo: Vimos que o modelo Poisson condicional P_μ^* traduzido pela distribuição $Bi(N, \mu_1/(\mu_1+\mu_2))$ é claramente não identificável. Adoptando para $\underline{\mu} = (\mu_1, \mu_2)'$ uma medida de probabilidade à priori ν equivalente à medida de Lebesgue em $\mathbb{R}^+ \times \mathbb{R}^+$ (e.g., caracterizada pelo produto de distribuições gama) não é possível escolher um boreliano B_0 do 1º quadrante de \mathbb{R}^2 com $\nu(B_0) = 0$ tal que a função $\underline{\mu} \rightarrow P_\mu^*(.)$, $\underline{\mu} \in (\mathbb{R}^+)^2 - B_0$ seja injectiva. Por outras palavras, este modelo é não identificável em ambos os sentidos.

Observação 2.3 - O termo identificabilidade é ainda usado noutros sentidos na literatura das Probabilidades e Estatística, como no conceito de identificabilidade forte entre duas sub- σ -álgebras (vide apêndice B.3). Num contexto Bayesiano, a identificabilidade forte de \mathcal{B} por \mathcal{A} implica o conceito de identificabilidade de Florens et al. (1983) referido acima e pode relacionar-se ainda com um outro conceito da Teoria das Probabilidades, o de identificabilidade de misturas.

Observação 2.4 - Uma análise detalhada de vários destes conceitos de identificabilidade e das suas inter-relações, bem como da sua ligação com correspondentes conceitos de estimabilidade, encarados no sentido de existência de estimadores consistentes, pode encontrar-se em Florens et al. (1983, caps. 4, 5 e 7).

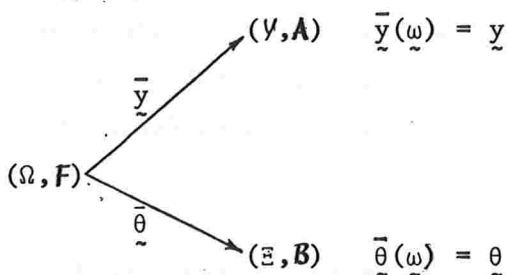
2. Em ordem a preparar o terreno para pôr em evidência a dualidade entre os aludidos conceitos Bayesianos da teoria da identificabilidade e conceitos correspondentes da teoria da suficiência amostral - objectivo da próxima sub-secção - torna-se necessário definir rigorosamente a estrutura em que se apoia a abordagem Bayesiana, o que é feito detalhadamente no apêndice B.2. Limitar-nos-emos agora a resumir os aspectos básicos e a descrever a notação que será utilizada.

Consideremos então o modelo estatístico "clássico" (Y, A, P) , $P = \{P_\theta : \theta \in \Xi\}$ e o espaço de probabilidade a priori (Ξ, \mathcal{B}, ν) . Será admitido que a σ -álgebra \mathcal{B} é seleccionada de modo a que as funções $\theta \rightarrow P_\theta(A)$ para todo o $A \in \mathcal{A}$, seja mensuráveis relativamente a ela. Desta forma, $\tilde{P}_\theta(A)$ representa uma função de transição de (Ξ, \mathcal{B}) para (Y, \mathcal{A}) .

O teorema generalizado da medida produto permite então definir no espaço (Ω, \mathcal{F}) , onde $\Omega = \Xi \times Y$ e $\mathcal{F} = \mathcal{B} \times \mathcal{A}$ representa a σ -álgebra produto, a medida produto $\mu = \nu \times P_\theta$ cuja restrição à classe de rectângulos mensuráveis $B \times A$, $B \in \mathcal{B}$, $A \in \mathcal{A}$ é definida por

$$\mu(B \times A) = \int_{\Xi} P_\theta(A) \nu(d\theta) \quad (2.12)$$

No modelo estatístico Bayesiano assim construído, $(\Omega, \mathcal{F}, \mu)$, as observações e o parâmetro podem ser visualizados através das funções de $\omega = (\theta', y')' \in \Omega$, \mathcal{F} -mensuráveis,



As sub-álgebras (abreviação de sub- σ -álgebras) de \mathcal{F} induzidas por $\bar{\theta}$ e por \bar{y} , denotadas por $\mathcal{F}_{\bar{\theta}}$ e por $\mathcal{F}_{\bar{y}}$ são as σ -álgebras geradas pelas classes dos cilindros $\{\bar{B} = B \times Y, B \in \mathcal{B}\}$ e $\{\bar{A} = \Xi \times A, A \in \mathcal{A}\}$ respectivamente, i.e.

$\mathcal{F}_{\bar{\theta}} = \mathcal{B} \times Y$ e $\mathcal{F}_{\bar{y}} = \Xi \times \mathcal{A}$. Assim, $\mathcal{F}_{\bar{\theta}} \vee \mathcal{F}_{\bar{y}} = \sigma\{\bar{B}\bar{A} : B \in \mathcal{B}, A \in \mathcal{A}\}$ corresponde a \mathcal{F} .

A expressão (2.12) traduz então que $\nu(B) \equiv \mu(\bar{B})$, $\forall B \in \mathcal{B}$, ou seja, a medida de probabilidade a priori pode ser identificada com a restrição de μ a $\mathcal{F}_{\bar{\theta}}$. Do mesmo modo, a medida de probabilidade marginal em (Y, \mathcal{A}) , $P(A) = \int_{\Xi} P_\theta(A) \nu(d\theta)$ traduz a restrição de μ a $\mathcal{F}_{\bar{y}}$.

Admitiremos ao longo de toda esta secção que (Y, \mathcal{A}) e (E, \mathcal{B}) são euclídeos, expressão esta usada para descrever um espaço mensurável onde o espaço é um boreliano em \mathbb{R}^n para algum n e a respectiva σ -álgebra é a classe de todos os borelianos desse espaço. Neste contexto, a função de transição $P_\theta(A): E \times \mathcal{A} \rightarrow [0,1]$ está associada a uma versão regular da probabilidade condicional $\mu(\bar{A} | \mathcal{F}_\theta)$. Do mesmo modo, podemos definir, associada a uma versão regular de $\mu(\bar{B} | \mathcal{F}_\theta)$, uma função de transição de (Y, \mathcal{A}) para (E, \mathcal{B}) $v_y(B): Y \times \mathcal{B} \rightarrow [0,1]$. Assim \underline{y} $\mathcal{Q} = \{v_y: y \in Y\}$ define uma família de medidas em (E, \mathcal{B}) denominadas medidas de probabilidade à posteriori. Nestas condições, (2.12) é equivalente a

$$\mu(B \times A) = \int_A v_y(B) P(dy) \quad (2.13)$$

o que revela que a medida produto $\mu = v \times P_\theta$ admite também a decomposição $\mu = P \times v_y$. Esta dupla decomposição de μ será designada, para usar a terminologia de Flörens et al. (1983), por regularidade do modelo Bayesiano. Sob as asas de um modelo Bayesiano regular fica portanto definido o modelo $(E, \mathcal{B}, \mathcal{Q})$, o denominado dual Bayesiano de (Y, \mathcal{A}, P) .

Sob a hipótese de (Y, \mathcal{A}) ser euclídeo e de P ser dominada por uma medida σ -finita m , a selecção de uma versão \mathcal{F} -mensurável, $f(y | \theta)$, de dP_θ/dm leva que o modelo Bayesiano seja dominado, entendendo-se esta dominação no sentido em que $\mu \ll v \times P$. Com base no uso de uma versão de $d\mu/d(v \times P)$, torna-se possível definir versões regulares das probabilidades condicionais $\mu(\bar{A} | \mathcal{F}_\theta)$, $A \in \mathcal{A}$ e $\mu(\bar{B} | \mathcal{F}_\theta)$, $B \in \mathcal{B}$ e assim assegurar a regularidade do modelo Bayesiano com a particularidade de $\mathcal{P} \ll \mathcal{Q}$ e de $\mathcal{Q} \ll \mathcal{P}$, onde \ll denota a relação de continuidade absoluta mútua.

O facto de E estar munido da estrutura mensurável \mathcal{B} leva-nos à consideração de funções de θ \mathcal{B} -mensuráveis. Por analogia com o significado de estatística no contexto clássico, utilizaremos doravante a expressão estatística paramétrica no seguinte sentido:

Def. 2.8: A função $\phi = \phi(\theta): E \rightarrow \Phi$ é uma estatística paramétrica (real) se ϕ é uma transformação \mathcal{B} -mensurável, de θ (com valores reais) relativamente a uma σ -álgebra \mathcal{B}_1 definida em Φ (\mathcal{B}_1 é a σ -álgebra dos borelianos de Φ), o que denotaremos por $\phi: (E, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_1)$.

Omitiremos, por vezes, a designação paramétrica quando nos reportamos a tais funções, desde que no contexto não surja qualquer possibilidade de confusão com estatísticas amostrais.

As estatísticas paramétrica $\phi: (\mathbb{E}, \mathcal{B}) \rightarrow (\mathbb{T}, \mathcal{B}_1)$ e amostral $\underline{t}: (\mathbb{Y}, \mathcal{A}) \rightarrow (\mathbb{T}, \mathcal{A}_1)$ serão identificadas pelas sub-álgebras induzidas, respectivamente $\mathcal{B}_{\phi} = \phi^{-1}(\mathcal{B}_1) \subset \mathcal{B}$ e $\mathcal{A}_{\underline{t}} = \underline{t}^{-1}(\mathcal{A}_1) \subset \mathcal{A}$. Estas definem em Ω as transformações compostas F -mensuráveis $\bar{\phi} = \phi \circ \bar{\theta}$ e $\bar{t} = \underline{t} \circ \bar{y}$, por sua vez indutoras das sub-álgebras de F , $\mathcal{F}_{\bar{\phi}} = \mathcal{B}_{\phi} \times \mathcal{Y}$ e $\mathcal{F}_{\bar{t}} = \mathbb{E} \times \mathcal{A}_{\underline{t}}$.

Em virtude das hipóteses assumidas sobre os espaços $(\mathbb{Y}, \mathcal{A})$ e $(\mathbb{E}, \mathcal{B})$ fica assegurada a existência de versões regulares das probabilidades condicionais em (Ω, \mathcal{F}) dado qualquer sub-álgebra de F . Adoptando a tática de identificar as sub-álgebras $\mathcal{C} \subset \mathcal{B}$ e $\mathcal{D} \subset \mathcal{A}$ com $\mathcal{C} \times \mathcal{Y} \subset \mathcal{F}_{\bar{\theta}}$ e $\mathbb{E} \times \mathcal{D} \subset \mathcal{F}_{\bar{y}}$, denotaremos a restrição a $(\mathbb{E}, \mathcal{B})$ duma versão regular das probabilidades μ condicionais em \mathcal{C} , $\mathcal{A} \vee \mathcal{C}$ (identificação de $\mathcal{F}_{\bar{y}} \vee (\mathcal{C} \times \mathcal{Y})$) e \mathcal{D} respectivamente por $\nu_{\mathcal{C}}$, $\nu_{\mathcal{A} \vee \mathcal{C}}$ e $\nu_{\mathcal{D}}$. Neste sentido, $\nu_{\mathcal{A}}(B)$ traduzirá a função de transição acima designada por $\nu_{\mathbb{Y}}(B)$, $B \in \mathcal{B}$. Do mesmo modo, a restrição a $(\mathbb{Y}, \mathcal{A})$ duma versão regular das probabilidades μ condicionais em \mathcal{D} , $\mathcal{B} \vee \mathcal{D}$ (identificação de $\mathcal{F}_{\bar{\theta}} \vee (\mathbb{E} \times \mathcal{D})$) e \mathcal{C} serão designadas respectivamente por $P_{\mathcal{D}}$, $P_{\mathcal{B} \vee \mathcal{D}}$ e $P_{\mathcal{C}}$. Assim, a função de transição $P_{\bar{\theta}}(A)$ passará a ser representada por $P_{\mathcal{B}}(A)$.

2.2.2 - A dualidade entre a teoria da identificabilidade paramétrica e a teoria da suficiência amostral

2.2.2.1 - A suficiência paramétrica Bayesiana e os conceitos Bayesiano e clássico de suficiência amostral

1. Em 2.1.1. vimos que a definição de suficiência paramétrica pode ser caracterizada em termos da factorização das medidas $P_{\bar{\theta}}$, $\bar{\theta} \in \mathbb{E}$, estabelecida no teorema 2.2. No actual contexto, esse teorema pode ser reformulado de modo a tomar em consideração a natureza de $P_{\bar{\theta}}(A)$ como função de transição em $\mathbb{E} \times \mathcal{A}$. Assim:

Teor. 2.11: A estatística $\underline{t}: (\mathbb{E}, \mathcal{B}) \rightarrow (\mathbb{T}, \mathcal{B}_1)$ com $\underline{t}(\mathbb{E}) \in \mathcal{B}_1$ é suficiente sse existe uma função de transição $P_{\underline{t}}^*(\cdot)$ em $\mathbb{T} \times \mathcal{A}$ tal que para todo o $A \in \mathcal{A}$, $P_{\bar{\theta}}(A) = P_{\underline{t}(\bar{\theta})}^*(A)$, $\forall \bar{\theta} \in \mathbb{E}$.

Dem. A implicação "se" está demonstrada no teorema 2.2. Provemos então a implicação "só se". A hipótese de \underline{t} ser suficiente implica que para cada $A \in \mathcal{A}$ a função $\gamma_A(\bar{\theta}) \equiv P_{\bar{\theta}}(A)$ depende de $\bar{\theta}$ apenas através de $\underline{t}(\bar{\theta})$. Ou seja, para cada $A \in \mathcal{A}$ a função $\gamma_A: (\mathbb{E}, \mathcal{B}) \rightarrow ([0,1], \mathcal{B}_0)$, onde \mathcal{B}_0 é a σ -álgebra dos borelianos de $[0,1]$, é uma estatística real $\mathcal{B}_{\underline{t}}$ -mensurável. Existe então (prop. B.1.3 e observação subsequente) uma transformação mensurável $\rho_A: (\mathbb{T}, \mathcal{B}_1) \rightarrow ([0,1], \mathcal{B}_0)$ tal que

$\gamma_A = \rho_A \circ \tau$. A função ρ_A está definida em T a menos de conjuntos de medida ν_τ^{-1} nula, já que qualquer conjunto \mathcal{B}_1 -mensurável incluído em $T - \tau(\mathcal{E})$ tem como imagem inversa por τ o conjunto vazio. Pelas hipóteses sobre $P_\theta(A)$, $\rho_A \circ \tau$ constitui uma função de transição em $\tau(\mathcal{E}) \times \mathcal{A}$. Como $\tau(\mathcal{E}) \in \mathcal{B}_1$, essa função de transição pode ser estendida a $T \times \mathcal{A}$, fixando-a para todo o $\tau \in T - \tau(\mathcal{E})$ igual a uma medida de probabilidade arbitrária em (Y, \mathcal{A}) . Denotando tal função de transição $T \times \mathcal{A} \rightarrow [0,1]$ por $P_\tau^*(A)$, fica assim demonstrado que para todo o $A \in \mathcal{A}$, $P_\theta(A) = P_{\tau(\theta)}^*(A)$, $\forall \theta \in \mathcal{E}$. \square

Dado que uma estatística suficiente não é, em geral, uma função injectiva de θ , este resultado indica-nos que é impossível extrair dos dados informações sobre θ tão detalhadas quanto se queira. Designadamente, torna-se impossível que a observação de y permita discriminar \mathcal{E} para além dos conjuntos que formam a partição nele induzida por τ . Estes conjuntos são os átomos (vide definição no apêndice B.1) de \mathcal{B}_τ , sendo elementos de \mathcal{B}_τ se os conjuntos unitários $\{\alpha\}$, $\alpha \in \tau(\mathcal{E})$, integram \mathcal{B}_1 , como acontece se τ é uma estatística real. Note-se que a função de transição $P_\tau^*(A)$ do teorema 2.11 corresponde, na notação definida em 2.2.1 e B.2, a $P_\tau(A)$ e a $\tilde{P}_{\mathcal{B}_\tau}(A)$.

2. Esta caracterização pode, pois, ser usada alternativamente como definição de uma estatística suficiente (como faz Picci [1977]) e, por extensão, de uma sub-álgebra suficiente (caso de Florens et al. [1983, sec. 2.3]). Em consonância com a linguagem mais geral e abstracta da teoria da suficiência clássica, optaremos aqui pelo tratamento dos conceitos da teoria da identificabilidade em termos de sub-álgebras paramétricas. A sua especialização a estatísticas é feita naturalmente pela consideração das respectivas sub-álgebras induzidas. A sua particularização a partições e funções arbitrárias é entendida através das sub-álgebras por elas induzidas (vide apêndice B.1).

Por outro lado, tendo em conta o referido em 2.2.1, estenderemos o conceito de suficiência paramétrica de modo a que a verificação da aludida factorização de ν -quase todas as medidas P_θ não invalide a caracterização de suficiência. Assim, numa extensão do teorema 2.11 diremos que

Def. 2.9: A σ -álgebra $\mathcal{C} \subset \mathcal{B}$ é suficiente Bayesiana se

$$\forall A \in \mathcal{A}, P_B(A) = P_C(A) [\nu]$$

Atendendo ao significado das funções no espaço produto subjacente, esta definição traduz que para cada $A \in \mathcal{A}$ a probabilidade condicional $\mu(\bar{A} | \mathcal{F}_\theta)$ tem uma versão $\mathcal{C} \times \mathcal{Y}$ -mensurável, igual μ -essencialmente a $\mu(\bar{A} | \mathcal{C} \times \mathcal{Y})$. Notando que

$\mathcal{C} \times \mathcal{Y} \subset \mathcal{F}_{\tilde{\theta}}$, fica patente pelo corolário B.3.4 (apêndice B.3) que esta definição é equivalente à independência condicional de $\mathcal{F}_{\tilde{y}}$ e $\mathcal{F}_{\tilde{\theta}}$ dado $\mathcal{C} \times \mathcal{Y}$, $\mathcal{F}_{\tilde{y}} \perp\!\!\!\perp \mathcal{F}_{\tilde{\theta}} \mid \mathcal{C} \times \mathcal{Y}$.

É neste sentido que deve ser entendida a notação mais simplificada, $\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C}$, que utilizaremos doravante. Pela simetria do conceito de independência condicional, a definição 2.9 torna-se equivalente a $\mathcal{B} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{C}$, o que, por sua vez, significa que as probabilidades condicionais $\mu[\bar{\mathcal{B}} \mid (\mathcal{C} \times \mathcal{Y}) \vee \mathcal{F}_{\tilde{y}}]$ e $\mu(\bar{\mathcal{B}} \mid \mathcal{C} \times \mathcal{Y})$ são μ -essencialmente iguais para todo o $B \in \mathcal{B}$ (proposição B.3.2 do apêndice B.3). Por conseguinte, fica estabelecido que

Prop. 2.2: A σ -álgebra $\mathcal{C} \subset \mathcal{B}$ é suficiente Bayesiana sse verificar uma das seguintes condições equivalentes:

- i) $\forall A \in \mathcal{A}, P_{\mathcal{B}}(A) = P_{\mathcal{C}}(A) \quad [v]$
- ii) $\mathcal{B} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{C}$
- iii) $\forall B \in \mathcal{B}, v_{\mathcal{A} \vee \mathcal{C}}(B) = v_{\mathcal{C}}(B) \quad [v]$

A condição i), ao indicar que cada elemento da família de funções $\{\gamma_A(\tilde{\theta}) = P_{\theta}(A); A \in \mathcal{A}\}$ tem uma versão \mathcal{C} -mensurável, revela que cada γ_A é essencialmente constante nos átomos de \mathcal{C} , ou, por outras palavras, que essencialmente todos os elementos de cada átomo de \mathcal{C} são observacionalmente equivalentes. Em termos inferenciais, torna-se impossível que o processo observacional opere qualquer refinamento dos átomos de qualquer sub-álgebra suficiente, na direcção dos átomos mais finos de \mathcal{B} . Esta asserção é inequivocamente evidenciada pela condição iii) ao traduzir que a amostra observada não contribui em essência para o aprofundamento do nosso conhecimento *à priori* sobre $\tilde{\theta}$ quando condicionado em \mathcal{C} . Essencialmente dentro de cada átomo de \mathcal{C} , a distribuição *à priori* não sofre qualquer actualização. Noutros termos, a experiência Bayesiana condicionada numa sub-álgebra paramétrica suficiente é, na prática, totalmente não informativa sobre $\tilde{\theta}$.

No contexto de uma estrutura Bayesiana, a suficiência da sub-álgebra $\mathcal{D} \subset \mathcal{A}$ pode ser analogamente definida por qualquer das condições equivalentes

- i') $\forall B \in \mathcal{B}, v_{\mathcal{A}}(B) = v_{\mathcal{D}}(B) \quad [P]$
- ii') $\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{D}$
- iii') $\forall A \in \mathcal{A}, P_{\mathcal{B} \vee \mathcal{D}}(A) = P_{\mathcal{D}}(A) \quad [P]$

A condição i'), traduzindo que a função $v_{\tilde{y}}(B)$, para cada $B \in \mathcal{B}$ apresenta uma versão \mathcal{D} -mensurável (ou equivalentemente, que quase todas as medidas $v_{\tilde{y}}$ de-

pendem de \underline{y} através dos átomos de \mathcal{D}) é vulgarmente usada como definição Bayesiana de suficiência amostral. A condição iii') está próxima da definição clássica, uma vez que, à parte a referência ao conjunto \mathcal{P} -nulo e nas condições de identificação de $P_{\mathcal{B}|\mathcal{D}}(A)$ com $P_{\theta}(A|\mathcal{D})$, revela que cada probabilidade amostral condicionada em \mathcal{D} não depende de θ . A comparação de cada uma destas condições com as correspondentes condições da proposição 2.2 deixa transparecer a dualidade entre os conceitos Bayesianos de suficiência no espaço paramétrico e no espaço amostral, primeiramente evidenciada por Barankin (1960) e posteriormente desenvolvida e generalizada por Picci (1974, 1977) e Florens et al (1983).

3. Em virtude desta dualidade, é de esperar que o conceito de suficiência paramétrica Bayesiana num modelo regular esteja associado a uma caracterização de suficiência com respeito ao modelo $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$ formalmente análoga à definição clássica (vide Halmos and Savage [1949]) de suficiência amostral com respeito a $(\mathcal{Y}, \mathcal{A}, \mathcal{P})$.

Com efeito, suponhamos que $\mathcal{C} \subseteq \mathcal{B}$ é uma σ -álgebra tal que para todo o $B \in \mathcal{B}$ existe uma função real ψ_B \mathcal{C} -mensurável satisfazendo

$$\nu_{\underline{y}}(B \cap C) = \int_C \psi_B d\nu_{\underline{y}}, \quad \forall C \in \mathcal{C}, \underline{y} \in \mathcal{Y}$$

onde $\mathcal{Q} = \{\nu_{\underline{y}} : \underline{y} \in \mathcal{Y}\}$ é a família de medidas de probabilidade à posteriori. Ou seja, ψ_B é uma versão comum de $\nu_{\underline{y}}(B|\mathcal{C})$ para todo o $\nu_{\underline{y}} \in \mathcal{Q}$. Como mostramos em B.2, se \mathcal{C} for tal que $\nu_{\underline{y}}(B|\mathcal{C})$ admite para todo o $\underline{y} \in \mathcal{Y}$ uma versão $\mathcal{A} \vee \mathcal{C}$ -mensurável (condição assegurada pela separabilidade de \mathcal{C}) então essa versão, em (Ω, \mathcal{F}) , coincide μ -essencialmente com $\nu_{\mathcal{A} \vee \mathcal{C}}(B)$. Consequentemente, para toda a probabilidade preditiva P e fazendo $\bar{\psi}_B = \psi_B \circ \bar{\theta}$, temos

$$\forall B \in \mathcal{B}, \quad \mu[\bar{B} | \mathcal{F}_{\bar{\underline{y}}} \nu(\mathcal{C} \times \mathcal{Y})] = \bar{\psi}_B \quad [\mu] \quad (2.14)$$

Do desenvolvimento de ambos os membros da relação definidora de $\nu_{\mathcal{A} \vee \mathcal{C}}$ (relação (19) em B.2 com $\mathcal{F}_{\bar{\underline{y}}}$ substituído por $\mathcal{C} \times \mathcal{Y}$), traduzido pelas relações correspondentes a (21) e (23) de B.2, obtemos fazendo $A_1 = \mathcal{Y}$

$$\nu(B \cap B_1) \equiv \int_{\mathcal{Y}} \nu_{\underline{y}}(B \cap B_1) P(d\underline{y}) = \int_{\mathcal{Y}} \left[\int_{B_1} \nu_{\mathcal{A} \vee \mathcal{C}}(B) \tilde{\nu}_{\underline{y}}(d\theta) \right] P(d\underline{y})$$

onde $\tilde{\nu}_{\underline{y}}$ é agora a restrição de $\nu_{\underline{y}}$ a $(\mathcal{E}, \mathcal{C})$. Considerando (2.14) resulta

$$\nu(B \cap B_1) = \int_{B_1} \psi_B(\theta) \int_{\mathcal{Y}} \tilde{\nu}_{\underline{y}}(d\theta) P(d\underline{y}) = \int_{B_1} \psi_B(\theta) \tilde{\nu}(d\theta)$$

o que significa que a versão comum \mathcal{C} -mensurável das probabilidades à posteriori condicionais é uma versão da probabilidade à priori condicional $\nu(B|\mathcal{C}) = \nu_{\mathcal{C}}(B)$. Deste modo, (2.14) traduz a suficiência Bayesiana de \mathcal{C} pela condição iii) da proposição 2.2.

Reciprocamente suponhamos que \mathcal{C} é suficiente Bayesiana. Da relação (24) de B.2 correspondente a ν_{AVC} obtemos para todo o $B \in \mathcal{B}$

$$\forall B_1 \in \mathcal{C}, \nu_{\underline{y}}(B \cap B_1) = \int_{B_1} \nu_{\mathcal{C}(B)} \tilde{\nu}_{\underline{y}}(d\theta) \quad [P]$$

isto é, $E_{\nu_{\underline{y}}} \{I_{B_1} I_B - I_{B_1} \nu_{\mathcal{C}(B)}\} = 0 \quad [P]$. Isto revela que para todo o $B \in \mathcal{B}$, $\nu_{\mathcal{C}(B)} = \nu_{\underline{y}}(B|\mathcal{C}) [\tilde{\nu}_{\underline{y}}]$ para P -essencialmente todo o $\underline{y} \in \mathcal{Y}$. Esta conclusão é válida sem conjuntos P -excepcionais se P for regular para \mathcal{Q} (vide 4) de B.2), o que significa então que $\nu_{\mathcal{C}}(\cdot)$ é uma versão comum \mathcal{C} -mensurável de $\nu_{\underline{y}}(\cdot|\mathcal{C})$ para todo o $\underline{y} \in \mathcal{Y}$. Fica assim provado que:

Teor. 2.12: Se P é regular para $\mathcal{Q} = \{\nu_{\underline{y}} : \underline{y} \in \mathcal{Y}\}$ e $\mathcal{C} \subset \mathcal{B}$ é separável, \mathcal{C} é suficiente Bayesiana sse para todo o $B \in \mathcal{B}$ existe uma função ψ_B \mathcal{C} -mensurável tal que, para todo o membro de \mathcal{Q} , $\psi_B = \nu_{\underline{y}}(B|\mathcal{C}) [\nu_{\underline{y}}]$.

4. Este resultado traduz assim o dual da conhecida equivalência entre os conceitos de suficiência Fisheriana e Bayesiana de uma σ -álgebra $\mathcal{D} \subset \mathcal{A}$ com respeito ao modelo $(\mathcal{Y}, \mathcal{A}, P)$.

No contexto de uma estrutura Bayesiana com \mathcal{Q} dominada por uma medida σ -finita e com a respectiva densidade F -mensurável (em 2) de B.2 vimos que ν é um candidato natural para a dominação de \mathcal{Q} sob a dominação de P , podemos então utilizar, sob as hipóteses do teorema anterior, o conceito dual da suficiência clássica para $\mathcal{C} \subset \mathcal{B}$ com respeito ao modelo $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$ - que denotaremos por $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$ - para chegarmos ao dual do teorema de factorização clássica. Dada a semelhança formal com o procedimento de Halmos - Savage - Bahadur, os resultados são apenas enunciados e comentados. No entanto, em ordem a obter uma descrição mais detalhada do assunto, as respectivas demonstrações podem encontrar-se no apêndice B.4.

Suponhamos em primeiro lugar que a família de medidas de probabilidade à posteriori é autodominada por $\nu_{\underline{y}_0}$ e seja $\phi_{\underline{y}}$, para cada $\underline{y} \in \mathcal{Y}$, uma versão fixa de $d\nu_{\underline{y}}/d\nu_{\underline{y}_0}$. Esta situação ocorre muitas vezes quando P e ν são dominadas e ν é um membro da família de distribuições conjugadas de P .

Teor. 2.13: Uma condição necessária e suficiente para $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$ é que $\phi_{\underline{y}}$ seja \mathcal{Q} -essencialmente \mathcal{C} -mensurável para cada $\underline{y} \in \mathcal{Y}$, i.e., $\sigma\{\phi_{\underline{y}} : \underline{y} \in \mathcal{Y}\} \subset \mathcal{C}[\tilde{\mathcal{Q}}]$.

Lema 2.1: $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$ sse $\mathcal{C} \text{ suf } (\mathcal{B}, \tilde{\mathcal{Q}})$, onde $\tilde{\mathcal{Q}}$ é o invólucro convexo de \mathcal{Q} , i.e.,

$$\tilde{\mathcal{Q}} = \{\nu^* : \nu^* = \sum_{i=1}^{\infty} a_i \nu_{\underline{y}_i}, \nu_{\underline{y}_i} \in \mathcal{Q}, a_i > 0, i=1, \dots, \infty, \sum_{i=1}^{\infty} a_i = 1\}.$$

Suponhamos agora que \mathcal{Q} é dominada por alguma medida σ -finita λ em $(\mathcal{E}, \mathcal{B})$. De acordo com Halmos and Savage (1949) existe uma medida de probabilidade

$\nu^* = \sum_{i=1}^{\infty} a_i \nu_{\underline{y}_i} \in \tilde{Q}$, para alguma sub-família contável $Q_0 = \{\nu_{\underline{y}_1}, \nu_{\underline{y}_2}, \dots\}$, tal que

$Q \ll \nu^*$. Note-se que sob a dominação de P , definida num espaço euclídeo, a medida à priori ν , que é uma mistura dos membros de Q , torna-se equivalente, com Q apropriadamente definida, à mistura discreta ν^* citada.

Nestas condições de dominação de Q por λ resulta, tendo em conta o lema e teorema anteriores:

Teor. 2.14: Sendo $Q \ll \lambda$ e ν^* a combinação convexa estabelecida acima, $C \text{ suf}(\mathcal{B}, Q)$ sse para cada $\underline{y} \in \mathcal{Y}$, $d\nu_{\underline{y}}/d\nu^*$ é Q -essencialmente C -mensurável, i.e., sse $\sigma\{d\nu_{\underline{y}}/d\nu^*: \underline{y} \in \mathcal{Y}\} \subset C[\tilde{Q}]$.

Neste momento estão reunidos os ingredientes necessários para a demonstração do teorema de factorização da função densidade à posteriori na forma correspondente à do teorema de factorização clássico de Bahadur (1954).

Teor. 2.15: Seja $Q \ll \lambda$ e $q_{\underline{y}}$ uma versão fixa de $d\nu_{\underline{y}}/d\lambda$. Então $C \text{ suf}(\mathcal{B}, Q)$ sse existem uma função não negativa \mathcal{B} -mensurável, h , e uma família de funções não negativas C -mensuráveis, $\{g_{\underline{y}}: \underline{y} \in \mathcal{Y}\}$, tal que para cada $\underline{y} \in \mathcal{Y}$, $q_{\underline{y}} = g_{\underline{y}} h[\lambda]$.

Este teorema proporciona assim um critério de suficiência paramétrica Bayesiana de uma σ -álgebra separável, caso P seja regular para Q . Contudo, a demonstração de que a suficiência Bayesiana de C implica que $C \text{ suf}(\mathcal{B}, Q)$ revela que a não verificação dessa condição de regularidade implica que a factorização do teorema 2.15 seja válida para P -essencialmente todo o \underline{y} .

Especializando o teorema anterior para a σ -álgebra induzida por uma estatística e tendo em conta a proposição B.1.3 obtemos:

Corolário: Seja P regular para a família Q dominada pela medida σ -finita λ e seja $q_{\underline{y}}$ uma versão F -mensurável da respectiva função densidade de $\nu_{\underline{y}}$. A estatística $\underline{\psi}: (\mathcal{E}, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1)$ com \mathcal{B}_1 separável é suficiente sse existem uma função h não negativa e \mathcal{B} -mensurável e um conjunto $\{g_{\underline{y}}: \underline{y} \in \mathcal{Y}\}$ de funções mensuráveis e não negativas de $\underline{\psi}$ tal que para todo o $\underline{y} \in \mathcal{Y}$

$$q_{\underline{y}}(\underline{\theta}) = g_{\underline{y}}[\underline{\psi}(\underline{\theta})] h(\underline{\theta}) \quad [\lambda] \quad \Delta \quad (2.15)$$

Como vimos em 2) de B.2, parte dos requisitos deste corolário são satisfeitos sob a dominação de ν por λ e de P por m tal que Q é seleccionada de modo que $\nu_{\underline{y}}(B) = \int_B g(\theta, \underline{y}) \nu(d\theta)$ com $g(\theta, \underline{y})$ uma versão F -mensurável de $d\nu/d(\nu \times P)$.

O recurso à teoria da suficiência clássica proporciona-nos assim uma derivação do critério de factorização para uma estatística suficiente diferente da

de Picci (1974). Para fins ilustrativos tomemos o seguinte exemplo:

Exemplo 1: Consideremos de novo a situação focada em 2.2.1 relativa ao exemplo III de 2.1.2. Neste caso, $Y = \{0, 1, \dots, N\}$, \mathcal{A} é o conjunto das partes de Y , $\Xi = \{\underline{\mu} = (\mu_1, \mu_2) : \mu_i \in \mathbb{R}^+, i=1,2\}$ e \mathcal{P} é definida em relação à medida de contagem m em Y por

$$dP_{\underline{\mu}}(y) = \binom{N}{y} \left(\frac{\mu_1}{\mu_1 + \mu_2} \right)^y \left(1 - \frac{\mu_1}{\mu_1 + \mu_2} \right)^{N-y} dm(y)$$

Tomemos $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_+$, com \mathcal{B}_+ a σ -álgebra de Borel em \mathbb{R}^+ e ν o produto das medidas gama (ou de Euler), $\mu_i | b_i \rightsquigarrow \Gamma(b_i, 1)$, $b_i > 0$, $i=1,2$, i.e.

$$d\nu(\underline{\mu}) = \prod_{i=1}^2 b_i^{-1} e^{-\mu_i} [\Gamma(b_i)]^{-1} d\lambda(\underline{\mu})$$

onde λ é a medida de Lebesgue em (Ξ, \mathcal{B}) .

É fácil constatar que a medida P é caracterizada pela distribuição beta-binomial de parâmetros N , b_1 e b_2 , sendo pois regular para \mathcal{Q} . Esta família é caracterizada pelas funções densidade $\{q_y(\underline{\mu}), y \in Y\}$, as quais se podem pôr na forma

$$q_y(\underline{\mu}) = [B(a_1, a_2)]^{-1} \left(\frac{\mu_1}{\mu_1 + \mu_2} \right)^{a_1 - 1} \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^{a_2 - 1} [\Gamma(b_i)]^{-1} (\mu_1 + \mu_2)^{b_i - 2} e^{-(\mu_1 + \mu_2)}$$

onde $b_i = b_1 + b_2$, $a_1 = a_1(y) = b_1 + y$ e $a_2 = a_2(y) = b_2 + N - y$. A inspeção desta fatorização revela, de acordo com o corolário anterior, que a estatística

$\psi = \mu_1 / (\mu_1 + \mu_2) : (\Xi, \mathcal{B}) \rightarrow ([0, 1], \mathcal{B}_+ \cap [0, 1])$ é suficiente enquanto que a estatística $\phi = \mu_1 + \mu_2 : (\Xi, \mathcal{B}) \rightarrow (\mathbb{R}^+, \mathcal{B}_+)$ já não satisfaz tal característica.

Estes resultados estão em consonância com a aplicação do critério i) da proposição 2.2 e podem igualmente ser obtidos à luz do critério iii) dessa proposição. A distribuição (à priori ou à posteriori) de $\underline{\mu}$ condicional a ψ é degenerada ao estar concentrada na parte da recta $\mu_1 / \mu_2 = \psi / (1 - \psi)$ contida em Ξ . Para a sua caracterização, podemos notar que a distribuição à posteriori (resp. à priori) de $\underline{\mu}$ é equivalente a assumir que ψ e ϕ são independentemente distribuídos segundo as distribuições $\psi | y \rightsquigarrow \text{Be}(a_1, a_2)$ (resp. $\psi \rightsquigarrow \text{Be}(b_1, b_2)$) e $\phi | y \rightsquigarrow \phi \rightsquigarrow \Gamma(b_i, 1)$. Atendendo a que ϕ e μ_1 , dado ψ , estão bijectivamente relacionados, a distribuição à posteriori de $\underline{\mu}$ condicional em ψ pode ser caracterizada pela distribuição $\mu_1 | \psi, y \rightsquigarrow \Gamma(b_i, 1/\psi)$ e por $\mu_2 = (1 - \psi)\mu_1 / \psi$ com probabilidade 1, coincidindo portanto com a correspondente distribuição condicional à priori. Esta identidade distribucional já não se verifica quando o condicionamento é em ϕ . As observações interferem apenas com a distribuição de ψ , ficando ϕ imutável distribucionalmente. Usando a reparametrização $\psi = \mu_1 / \mu_2$ e $\phi = \mu_2$ obtêm-se resultados semelhantes.

2.2.2.2 - A suficiência paramétrica Bayesiana minimal

1. Na linha do que foi referido em 2.1 em termos de partições (ou funções) paramétricas, várias sub-álgebras de \mathcal{B} satisfazem a definição de suficiência. Por exemplo, \mathcal{B} , que podemos associar à estatística $\tau(\theta) = \theta: (\mathcal{E}, \mathcal{B}) \rightarrow (\mathcal{E}, \mathcal{B})$, é trivialmente suficiente. A própria definição de suficiência implica, independentemente da dominação do modelo Bayesiano, a sua monotonicidade no sentido de a suficiência de \mathcal{C} implicar a suficiência de qualquer sub-álgebra de \mathcal{B} que contém \mathcal{C} (basta ter em conta a relação ii) da prop. 2.2 e a prop. B.3.5 do apêndice).

Interessa pois pesquisar a existência do elemento mínimo dentro da classe de σ -álgebras suficientes, à semelhança do que é feito para a suficiência amostral clássica. Contudo, dada a estrutura do modelo Bayesiano, tal elemento mínimo deve ser definido com respeito à relação de ordem parcial de inclusão μ -essencial entre as sub-álgebras de $F_{\bar{\theta}}$. Devido ao modo como é definido o completamento das sub-álgebras de $F_{\bar{\theta}}$ (vide B.1), $\overline{\mathcal{C} \times \mathcal{Y}}$ com $\mathcal{C} \sqsubset \mathcal{B}$ não está incluída necessariamente em $F_{\bar{\theta}}$, e $\overline{\mathcal{C} \times \mathcal{Y}} \cap F_{\bar{\theta}}$ (identificada por $\overline{\mathcal{C}} \cap \mathcal{B}$) representa a menor sub-álgebra de $F_{\bar{\theta}}$ contendo $\mathcal{C} \times \mathcal{Y}$ e todos os conjuntos μ -nulos de $F_{\bar{\theta}}$. Por isso, Florens et al. (1983), definem a suficiência Bayesiana minimal do seguinte modo, onde mantemos o procedimento de identificar as sub-álgebras de $F_{\bar{\theta}}$ pelas correspondentes sub-álgebras de \mathcal{B} :

Def. 2.10: $\mathcal{C}^* \sqsubset \mathcal{B}$ é suficiente Bayesiana minimal se $\mathcal{B} \perp\!\!\!\perp \mathcal{A} | \mathcal{C}^*$ e $\mathcal{C}^* \sqsubset \overline{\mathcal{C}} \cap \mathcal{B}$, $\forall \mathcal{C} \sqsubset \mathcal{B}$ tal que $\mathcal{B} \perp\!\!\!\perp \mathcal{A} | \mathcal{C}$.

A prop. B.3.6 do apêndice comprova que este conceito não é vazio. Concretamente, a projecção de $F_{\bar{y}}$ sobre $F_{\bar{\theta}}$ definida por

$$F_{\bar{\theta}}[F_{\bar{y}}] = \sigma\{E[\bar{t} | F_{\bar{\theta}}], \forall \bar{t} \in L_{\infty}(F_{\bar{y}})\}$$

onde $L_{\infty}(F_{\bar{y}})$ é o conjunto das funções reais $F_{\bar{y}}$ -mensuráveis μ -essencialmente limitadas, e identificada doravante por $\mathcal{B}[\mathcal{A}]$, é a menor das sub-álgebras de \mathcal{B} que são suficientes, no sentido em que $\mathcal{B}[\mathcal{A}] \sqsubset \overline{\mathcal{C}}$, para todo \mathcal{C} suficiente. Atendendo à definição do \mathcal{B} -completamento (apêndice B.1) e ao facto de $\mathcal{B}[\mathcal{A}]$ ser \mathcal{B} -completa (Florens et al [1983, p. IV.11], conclui-se então que $\mathcal{B}[\mathcal{A}]$ é uma sub-álgebra suficiente minimal.

A definição 2.10 implica que a suficiência Bayesiana minimal é unicamente definida dentro da classe das sub-álgebras \mathcal{B} -completas, pelo que $\mathcal{B}[\mathcal{A}]$ é a única sub-álgebra suficiente minimal \mathcal{B} -completa.

2. Em B.3 definiu-se o conceito de identificabilidade forte entre duas sub-álgebras de \mathcal{F} . Em particular, $\mathcal{C} \subset \mathcal{B}$ é fortemente identificável por \mathbf{A} se

$$\forall \bar{\phi} \in L_{\infty}(\mathcal{C} \times Y), E(\bar{\phi} | \mathcal{F}_{\bar{y}}) = 0 [\mu] \implies \bar{\phi} = \underline{0} [\mu] \quad (2.16)$$

No contexto em causa de um modelo Bayesiano regular, isto significa que não existe basicamente nenhuma estatística não trivial, essencialmente limitada e \mathcal{C} -mensurável, cuja esperança a posteriori seja nula para quase todas as realizações amostrais. Atendendo à definição de suficiência Bayesiana, a proposição B.3.7 particulariza-se então em:

Teor. 2.16. Se $\mathcal{C} \subset \mathcal{B}$ é suficiente e fortemente identificável por \mathbf{A} , então \mathcal{C} é essencialmente igual à projecção de \mathbf{A} sobre \mathcal{B} .

Considerando para sub-álgebras de \mathcal{B} o conceito de σ -álgebra completa limitada como dual do correspondente conceito clássico (vide definição em B.3), a proposição B.3.8 permite-nos então concluir do teorema anterior que

Corolário: Se P é regular para \mathcal{Q} e \mathcal{C} é uma σ -álgebra suficiente e completa limitada, então \mathcal{C} é suficiente minimal.

Este corolário, aplicado à σ -álgebra \mathcal{B}_{ψ} induzida pela estatística ψ , representa, pois, a versão Bayesiana do dual do conhecido resultado de Lehmann-Scheffé de que uma estatística amostral suficiente e completa limitada é suficiente minimal (numa acepção Fisheriana).

3. Designemos $P_{\theta}(A)$ como função real \mathcal{B} -mensurável por $\gamma_A(\theta)$, i.e.

$\gamma_A: (\mathbb{E}, \mathcal{B}) \rightarrow ([0,1], \mathcal{B}_0)$ onde \mathcal{B}_0 é a σ -álgebra dos borelianos de $[0,1]$. Seja \mathcal{C}_0 a σ -álgebra gerada pela família de funções $\{\gamma_A: A \in \mathcal{A}\}$, i.e., $\mathcal{C}_0 = \bigvee_{A \in \mathcal{A}} \gamma_A^{-1}(\mathcal{B}_0)$.

Recordando a noção de equivalência observacional e a definição de partição induzida por uma σ -álgebra (apêndice B.1), temos

$$\begin{aligned} \theta_{\sim 1} \sim \theta_{\sim 0} &\iff \gamma_A(\theta_{\sim 1}) = \gamma_A(\theta_{\sim 0}), \forall A \in \mathcal{A} \iff \forall B_a \in \gamma_A^{-1}(\mathcal{B}_0), I_{B_a}(\theta_{\sim 1}) = I_{B_a}(\theta_{\sim 0}), \forall A \in \mathcal{A} \\ &\iff \theta_{\sim 1} \gamma_A^{-1}(\mathcal{B}_0) \theta_{\sim 0}, \forall A \in \mathcal{A} \iff \theta_{\sim 1} \tilde{\mathcal{C}}_0 \theta_{\sim 0} \end{aligned}$$

ou seja, as classes de equivalência observacional correspondem aos átomos de \mathcal{C}_0 , i.e., $\mathbb{E}/\sim = \Pi(\mathcal{C}_0)$.

A natureza euclideana de (Y, \mathcal{A}) implica que a partição induzida por \mathcal{C}_0 , $\Pi_0 = \mathbb{E}/\sim$, seja mensurável. De facto, seja S uma classe- π contável de conjuntos A_n gerando \mathcal{A} .

Por um argumento em termos do teorema de Dynkin (apêndice B.3) prova-se que

$\mathcal{C}_0 = \bigvee_{n=1}^{\infty} \gamma_{A_n}^{-1}(\mathcal{B}_0)$, o que ilustra a separabilidade de \mathcal{C}_0 , dada a separabilidade de cada $\gamma_{A_n}^{-1}(\mathcal{B}_0)$. Sendo então $\{C_n, n \geq 1\}$ uma classe geradora de \mathcal{C}_0 , a partição induzida por \mathcal{C}_0 é então definida por

$$\Pi_0 = \left\{ \bigcap_{n=1}^{\infty} B_n : B_n = C_n \text{ ou } C_n^c \right\}, \quad C_n^c = \Xi - C_n$$

Mais concretamente

$$[\theta_0] = \left[\bigcap_{n: \theta_0 \in C_n} C_n \right] \cap \left[\bigcap_{n: \theta_0 \notin C_n} C_n^c \right].$$

Os átomos de \mathcal{C}_0 são pois elementos de $\mathcal{C}_0 \subset \mathcal{B}$, o que implica nomeadamente que Π_0 seja também induzida pela σ -álgebra $\mathcal{B}_{\Pi_0} \supset \mathcal{C}_0$ induzida por Π_0 (prop. B.1.6). Enquanto \mathcal{B}_{Π_0} , por definição, é induzida pela função $\gamma_0: \Xi \rightarrow \Pi_0$ encarada como estatística no sentido de Bahadur (i.e. pela transformação $(\Xi, \mathcal{B}) \rightarrow (\Pi_0, \mathcal{B}_1^*)$, com $\mathcal{B}_1^* = \{C \subset \Pi_0 : \gamma_0^{-1}(C) \in \mathcal{B}\}$, ou seja, $\mathcal{B}_{\Pi_0} = \mathcal{B} \cap \gamma_0^{-1}[P(\Pi_0)]$ - vide apêndice B.1 - \mathcal{C}_0 pode ser vista (Picci [1977]) como induzida por γ_0 entendida como a estatística $(\Xi, \mathcal{B}) \rightarrow (\Pi_0, \mathcal{B}_0^*)$, $\mathcal{B}_0^* = \{C \subset \Pi_0 : \gamma_0^{-1}(C) \in \mathcal{C}_0\}$. Contudo, dada a sua separabilidade, \mathcal{C}_0 pode ser visualizada como induzida por uma estatística real, necessariamente ligada à estatística γ_0 por uma transformação bijectiva mensurável. No exemplo referido em 2.2.2.1, \mathcal{C}_0 é, dada a forma de $P_{\underline{\mu}}$, claramente gerada pela estatística $\psi(\underline{\mu}) = \mu_1 / (\mu_1 + \mu_2): (\Xi, \mathcal{B}) \rightarrow ([0, 1], \mathcal{B}_0)$ ou ainda por $\eta(\underline{\mu}) = \mu_1 / \mu_2: (\Xi, \mathcal{B}) \rightarrow (\mathbb{R}^+, \mathcal{B}_+)$.

A própria definição de \mathcal{C}_0 revela a sua suficiência (def. 2.9). Além disso, o uso do mesmo argumento que conduziu em B.2 à identificação de $E_{\mu}(\bar{I}_A | \bar{F}_{\theta})$ com $E_{P_{\theta}}(\bar{I}_A)$ leva a que $E_{\mu}(\bar{t} | \bar{F}_{\theta}) = E_{P_{\theta}}(t) [\mu]$, para $\bar{t} = t\bar{y}$ com $t \in L_{\infty}(\mathcal{A})$. Assim,

$\mathcal{C}_0 \subset \overline{\mathcal{B}[A]} \cap \mathcal{B} = \mathcal{B}[A]$, o que ilustra a suficiência minimal de \mathcal{C}_0 , que é a σ -álgebra suficiente minimal usada por Picci (1977). As sub-álgebras \mathcal{C}_0 e $\mathcal{B}[A]$ exemplificam, deste modo, uma situação de compatibilidade da definição 2.10 com a existência de duas sub-álgebras suficientes minimais com uma eventualmente incluída estritamente na outra - basta que \mathcal{C}_0 não seja \mathcal{B} -completa. Por outro lado, como $E_{\mu}(\bar{t} | \bar{F}_{\theta})$ é um elemento de $\overline{\mathcal{C}_0} \times \mathcal{V} \cap \bar{F}_{\theta}$, resulta $\mathcal{B}[A] \subset \overline{\mathcal{C}_0} \cap \mathcal{B} = \overline{\mathcal{C}_0}^{\vee}$, onde " \vee " indica o completamento com os conjuntos de \mathcal{B} \vee -nulos. Este resultado de Florens et al. (1983) estabelece portanto que:

Teor. 2.17: Para toda a medida de probabilidade $\tilde{\nu}$ a priori ν , a menor sub-álgebra de \mathcal{B} que torna as probabilidades amostrais $P_{\theta}(A)$, $A \in \mathcal{A}$, mensuráveis é suficiente minimal e o seu completamento com os conjuntos ν -nulos de \mathcal{B} coincide com a projecção de \mathcal{A} sobre \mathcal{B} .

4. Em 2.2.2.1 definiu-se um outro conceito de suficiência paramétrica que é o análogo dual do conceito Fisheriano de suficiência amostral. Por analogia com o problema da suficiência minimal clássica, o elemento mínimo da classe de sub-álgebras que são suf $(\mathcal{B}, \mathcal{Q})$ pode não existir em geral, contrariamente ao que se passa com o conceito de suficiência Bayesiana. Contudo, no caso de dominação de \mathcal{Q} , o teorema 2.14 evidencia que $\mathcal{C}_* = \sigma\{d\nu_y/d\nu^* : y \in Y\}$ é uma σ -álgebra suficiente minimal no sentido em que $\mathcal{C}_* \subset \overline{\mathcal{C}}^{\nu^*}$, $\forall \mathcal{C} \text{ suf}(\mathcal{B}, \mathcal{Q})$, onde " $\overline{\cdot}^{\nu^*}$ " representa o completamento com os conjuntos ν^* -nulos (ou equivalentemente, \mathcal{Q} -nulos) de \mathcal{B} . É fácil verificar que, nas condições do teorema 2.12, \mathcal{C}_* é também suficiente minimal no sentido da definição 2.10. Mais concretamente, seguindo a argumentação de Florens et al. (1983), podemos provar por dualidade que:

Teor. 2.18: Se \mathcal{C}_* é separável e se P é regular para \mathcal{Q} , então o completamento de \mathcal{C}_* com os conjuntos ν^* -nulos de \mathcal{B} é igual à projecção de \mathcal{A} sobre \mathcal{B} .

Dem. A hipótese de regularidade de P para \mathcal{Q} , implica, por um lado, $\nu^* \ll \nu$, pois $\mathcal{Q} \ll \nu$, e por outro, que $\mathcal{B}[\mathcal{A}]$ é também suf $(\mathcal{B}, \mathcal{Q})$. Assim, $\mathcal{C}_* \subset \overline{\mathcal{B}[\mathcal{A}]}^{\nu^*} = \overline{\mathcal{B}[\mathcal{A}]}^{\nu} = \overline{\mathcal{B}[\mathcal{A}]} \cap \mathcal{B} = \mathcal{B}[\mathcal{A}]$, o que implica a suficiência Bayesiana minimal de \mathcal{C}_* e $\overline{\mathcal{C}_*}^{\nu^*} \subset \mathcal{B}[\mathcal{A}]$. A separabilidade de \mathcal{C}_* implica pelo teorema 2.12 a sua suficiência Bayesiana e como $\mathcal{B}[\mathcal{A}]$ é suficiente Bayesiana minimal, $\mathcal{B}[\mathcal{A}] \subset \overline{\mathcal{C}_*}^{\nu} = \overline{\mathcal{C}_*}^{\nu^*}$, c.q.d. Δ

No caso de dominação de \mathcal{P} pela medida σ -finita m , vimos em B.2 que para qualquer versão de $d\nu/d(\nu \times P) = g(\underline{\theta}, \underline{y})$ existem conjuntos N_1 , ν -nulo, e N_2 , P -nulo, tal que $dP_{\theta}/dP = g(\underline{\theta}, \underline{y})$ em $E - N_1$ e $d\nu_y/d\nu = g(\underline{\theta}, \underline{y})$ em $Y - N_2$. Em particular, pode ser escolhida uma versão $h(\underline{\theta}, \underline{y})$ tal que $\mathcal{P} \ll \mathcal{P} \ll \nu$ e $\mathcal{Q} \ll \nu$. Assim, ν pode tomar o papel de ν^* e como tal, de acordo com os dois teoremas anteriores, $\sigma\{h(\underline{\theta}, \underline{y}) : \underline{y} \in Y\}$ será ν -equivalente a \mathcal{C}_0 . Com base essencialmente neste tipo de raciocínio aplicado sobre ambos os espaços paramétrico e amostral, Picci (1974, 1977) obtém um teorema de factorização simultâneo de cada versão de $g(\underline{\theta}, \underline{y})$ em termos da estatística paramétrica γ_0 e da estatística amostral suficiente Bayesiana minimal dual. A sua aplicação à família exponencial de distribuições permite-lhe obter simultaneamente uma caracterização da suficiência minimal Bayesiana amostral e paramétrica.

5. A existência de uma sub-álgebra suficiente Bayesiana minimal, como por exemplo \mathcal{C}_0 , permite concluir dada a monotonicidade da suficiência Bayesiana e a definição de suficiência minimal que

Prop 2.3: $\mathcal{C} \subset \mathcal{B}$ é suficiente Bayesiana sse $\mathcal{C}_0 \subset \mathcal{C}[\nu]$.

Esta caracterização de suficiência Bayesiana para σ -álgebras está assim para a definição 2.5 em termos de partições como a definição 2.9 está para a caracterização estabelecida no teorema 2.2 relativa às funções associadas às partições. Como a suficiência de uma partição é definida através da suficiência da respectiva σ -álgebra induzida, a generalização da definição 2.5 que corresponde à proposição 2.3 para sub-álgebras é então:

Def. 2.11: A partição Π é suficiente Bayesiana se $\Pi_0 \equiv \Pi(\mathcal{C}_0) < \mathbb{H}[\nu]$.

Note-se que esta definição é uma consequência do dual da definição clássica de partição amostral suficiente minimal, já que $\nu \ll \mathcal{Q}$, tornando-se as duas definições equivalentes no caso de dominação usual do modelo clássico, nos termos já anteriormente referidos. Por outro lado, ela é equivalente à relação de inclusão ν -essencial entre as σ -álgebras induzidas na classe das partições induzidas por estatísticas reais (prop. B.1.8).

2.2.2.3 - A identificabilidade Bayesiana e o conceito amostral de "necessidade"

Pelo mesmo argumento que conduziu à definição 2.11, generalizaremos a definição 2.4 para

Def. 2.12: A partição Π é identificável no sentido Bayesiano se $\Pi < \Pi_0[\nu]$

As definições 2.11 e 2.12 implicam assim que Π é identificante (i.e. identificável e suficiente) no sentido Bayesiano se e só se $\Pi = \Pi_0[\nu]$. Em termos de σ -álgebras, diremos analogamente

Def. 2.13: A σ -álgebra $\mathcal{C} \subset \mathcal{B}$ é identificável no sentido Bayesiano se $\mathcal{C} \subset \mathcal{C}_0[\nu]$.

Patenteia-se assim a dualidade com o conceito clássico de σ -álgebra necessária quando $\nu \equiv \mathcal{Q}$. Com esta definição, todo o elemento minimal da classe de sub-álgebras suficientes é um elemento maximal da classe de sub-álgebras identificáveis e vice-versa. Daí a designação identificante para caracterizar o conceito que, nesta secção, começou por ser denominado de suficiência minimal. Os átomos de uma σ -álgebra identificante são simultaneamente os átomos mais grossos (essencialmente falando) que são deixados invariantes pelos dados - no sentido de iii) da proposição 2.2 - e os átomos mais finos que são detectáveis pelo processo observacional.

A definição 2.13 mostra ainda que toda a sub-álgebra é identificável quando \mathcal{B} coincide com \mathcal{C}_0 . Dada a natureza euclideana de $(\mathcal{E}, \mathcal{B})$, $\Pi(\mathcal{B}) = \{\{\emptyset\}, \emptyset \in \mathcal{E}\}$, e assim $\mathcal{B} = \mathcal{C}_0$ implica também a identificabilidade de qualquer partição ou função.

Facilmente se obtêm versões dos teoremas enunciados em 2.1.1 correspondentes a este conceito estendido de identificabilidade. No que concerne à caracterização de funções identificáveis estabelecida no teorema 2.4, podemos então afirmar que a função $\phi: \mathcal{E} \rightarrow \Phi$ é identificável no sentido Bayesiano sse é ν -essencialmente igual a uma função de ψ , onde $\psi: \mathcal{E} \rightarrow \Psi$ é uma função que induz uma partição identificante. Sendo ϕ e ψ estatísticas reais, indutoras das σ -álgebras \mathcal{B}_ϕ e \mathcal{B}_ψ , respectivamente, e lembrando que $(\mathcal{E}, \mathcal{B})$ é euclideano, podemos concluir dos resultados de Bahadur (1955) que a afirmação anterior é equivalente a $\mathcal{B}_\phi \subset \mathcal{B}_\psi[\nu]$. Assim fica demonstrado que:

Prop. 2.4: Sendo ψ uma estatística real identificante, a estatística real ϕ é identificável no sentido Bayesiano sse ϕ é ν -essencialmente \mathcal{B}_ψ -mensurável.

A restrição aos conceitos de suficiência e de identificabilidade de 2.1, independentes de qualquer medida a priori, permite-nos "simplificar" a caracterização estabelecida nesta proposição. Com efeito, sendo ψ uma função sobrejectiva identificante arbitrária indutora da σ -álgebra \mathcal{B}_ψ^* (no sentido de Bahadur - vide B.1), e tomando em consideração a proposição B.1.2. e a observação ii) subsequente à proposição B.1.3, o teorema 2.4 permite-nos afirmar que

Prop. 2.5: Sendo ψ uma função sobrejectiva identificante, a estatística $\phi: (\mathcal{E}, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_2)$, com \mathcal{B}_2 incluindo os conjuntos singulares de ϕ , é identificável sse ϕ é \mathcal{B}_ψ^* -mensurável.

O teorema 1 de Kadane (1974) é um caso especial deste resultado, dele obtido considerando ϕ uma estatística real.

2.2.3 - A identificabilidade e o conceito de invariância paramétrica

O objectivo desta sub-secção consiste em evidenciar certas analogias entre os conceitos da teoria da identificabilidade, tal como foram definidos em 2.1 e conceitos correspondentes da teoria da invariância face a um determinado grupo de transformações mensuráveis em \mathcal{E} . Isto permitirá que a caracterização de funções identificáveis estabelecida no teorema 2.4 e na proposição 2.5 possa ser formulada na linguagem da teoria da invariância paramétrica.

Seja J a classe de transformações bijectivas \mathcal{B} -mensuráveis de \mathcal{E} e seja "o" a operação composição nela definida. É fácil constatar que (J, o) é um grupo de transformações. Seja H uma sub-classe de J tal que (H, o) seja um sub-grupo de

transformações e definamos em Ξ a relação $\theta_1 H \theta_0 \iff \exists \zeta \in H: \theta_1 = \zeta(\theta_0)$. A natureza de (H, \circ) como grupo implica que esta relação binária seja uma relação de equivalência. O conjunto quociente Ξ/H (na terminologia de Lehmann [1986, cap. 6], o conjunto das órbitas de H) é formado então pelos elementos

$$[\theta]_H = \{\theta \in \Xi: \theta = \zeta(\theta_0) \text{ para algum } \zeta \in H\}$$

Recordando os conceitos da teoria da invariância, temos

Def. 2.14: i) A função $\phi: \Xi \rightarrow \Phi$ é invariante sob H se é constante em cada órbita de H , i.e., se

$$\forall \theta_1, \theta_0 \in \Xi, \theta_1 H \theta_0 \implies \phi(\theta_1) = \phi(\theta_0)$$

ii) A função $\phi: \Xi \rightarrow \Phi$ é invariante maximal sob H se é invariante e assume valores diferentes de órbita para órbita, i.e., se além de i) verifica

$$\forall \theta_1, \theta_0 \in \Xi, \phi(\theta_1) = \phi(\theta_0) \implies \theta_1 H \theta_0$$

Esta definição põe em relevo o paralelismo formal que existe entre os conceitos de invariância e de identificabilidade. Uma função invariante (invariante maximal) sob H actua perante a partição Ξ/H do mesmo modo que uma função identificável (identificante ou identificável maximal) perante Ξ/\sim . Com base nesta relação, torna-se imediato, por analogia com o teorema 2.4, que uma função é invariante sse é função de uma função invariante maximal (análogo paramétrico do teorema 1 de Lehmann [1986, p. 285]).

Esta constatação permite compreender que o conceito de identificabilidade pode ser traduzido em termos de invariância relativamente a um grupo de transformações que induza a partição identificante Ξ/\sim . É nesta linha de raciocínio que se insere a ideia de Kadane (1972) ao considerar o sub-grupo (G, \circ) onde $G = \{\zeta \in J: \zeta(\theta) \sim \theta, \forall \theta \in \Xi\}$ é a classe de transformações bijectivas mensuráveis de Ξ observacionalmente equivalentes à função identidade.

É fácil verificar que $\Xi/G = \Xi/\sim$. De facto, sejam θ_1 e θ_0 dois pontos arbitrários de Ξ tais que $\theta_1 G \theta_0$. Então, por definição, $\exists \zeta \in G: \theta_1 = \zeta(\theta_0)$, o que implica que $\theta_1 \sim \theta_0$ por definição de G . Por outro lado, se $\theta_1 \sim \theta_0$ definamos a função ζ tal que $\zeta(\theta_1) = \theta_0$, $\zeta(\theta_0) = \theta_1$ e $\zeta(\theta) = \theta$, $\forall \theta \neq \theta_1, \theta_0$. É fácil constatar que ζ é uma transformação bijectiva $(\Xi, \mathcal{B}) \rightarrow (\Xi, \mathcal{B})$ pertencente a G . Assim, $\theta_1 \sim \theta_0 \implies \theta_1 G \theta_0$. Por conseguinte, fica demonstrado que

Teor. 2.19: A função $\phi: \Xi \rightarrow \Phi$ é identificável se e só se ϕ é invariante face ao grupo G .

Deve notar-se que este teorema é precisamente o teorema 2 de Kadane (1972), uma vez que a definição 2.14 i) é equivalente a afirmar que $(\phi \circ \zeta)(\underline{\theta}) = \phi(\underline{\theta})$, $\forall \underline{\theta} \in \Xi$, para toda a função ζ de G . O recurso à partição induzida por um grupo de transformações e ao conceito de invariância possibilitam assim uma derivação e percepção mais claras da correspondência enunciada no teorema 2.19.

Observação 2.5: Este teorema permite comprovar imediatamente o caso trivial de que qualquer função constante é sempre identificável, bem como o facto de que só as funções constantes são identificáveis quando todos os pontos de Ξ são observacionalmente quivalentes. De facto, se $\Xi/\sim = \{\Xi\}$, por definição $G=J$ e assim, $\forall A \in \mathcal{A}$, $\forall \underline{\theta} \in \Xi$, $P_{\underline{\theta}}(A) = P_{\zeta(\underline{\theta})}(A)$ para toda a função $\zeta \in J$, o que significa que a medida $P_{\underline{\theta}}$ não depende efectivamente de $\underline{\theta}$.

Consideremos agora o outro caso extremo, no qual Ξ é identificável, ou, nouros termos, $\underline{I}(\underline{\theta})$ (função identidade) é identificável. Por aplicação do teorema 2.19 e atendendo à definição de $\underline{I}(\cdot)$, temos para todo o $\underline{\theta} \in \Xi$

$$\forall \zeta \in G: \underline{I}[\zeta(\underline{\theta})] = \underline{I}(\underline{\theta}) \iff \forall \zeta \in G: \zeta(\underline{\theta}) = \underline{\theta} \iff G = \{\underline{I}\}$$

o que ilustra de outro modo o resultado de que toda a função definida num espaço identificável é identificável. Δ

O teorema 2.19 aliado ao análogo invariante do teorema 2.4 possibilita-nos afirmar que a função ϕ é identificável sse é uma função de uma invariante maximal sob G . Por analogia com a proposição 2.5, podemos então estabelecer que:

Teor. 2.20: A estatística real $\phi: (\Xi, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_2)$ é identificável sse ϕ é \mathcal{B}^* -mensurável, onde \mathcal{B}^* é a σ -álgebra induzida no sentido de Bahadur, pela função $\psi: \Xi \rightarrow \Psi$ invariante maximal relativamente a G .

Definamos a invariância de uma partição Π através da invariância da função associada $\gamma_{\Pi}: \Xi \rightarrow \Pi$. Pode então obter-se um enunciado correspondente ao do teorema anterior para a caracterização da identificabilidade de uma partição \mathcal{B} -mensurável, em termos da relação de redução com uma partição invariante maximal arbitrária, mediante o emprego das funções associadas e da proposição B.1.4.

As caracterizações definidas nos teoremas anteriores constituem, pois, um método de determinação da identificabilidade através de conceitos da teoria da invariância. A sua aplicação exige a especificação completa de G e é aqui que reside a principal dificuldade, já que não é difícil, em geral, definir um sub-grupo de G . O caso do modelo linear normal é um exemplo em que não é complicado definir-se completamente G . Para ilustração, consideremos a situação particular do modelo ANOVA de 2 factores, cada um em 2 níveis, sem interacção e com uma observação em cada cela. Da caracterização das classes observacionalmente equivalentes

feita em 2.1 conclui-se facilmente, fazendo $\underline{\theta} = (\eta \ \alpha_1 \ \alpha_2 \ \gamma_1 \ \gamma_2 \ \sigma^2)'$, que G é um grupo aditivo definido por $G = \{\underline{z} : \underline{z}(\underline{\theta}) = \underline{\theta} + \underline{\delta}, \ \underline{\delta} \in M(\underline{\delta}_1, \underline{\delta}_2)\}$, onde $M(\underline{\delta}_1, \underline{\delta}_2)$ designa o subespaço de $\mathbb{R}^5 \times \mathbb{R}_+$ gerado pelos vectores $\underline{\delta}_1 = (1 \ -1 \ -1 \ 0 \ 0 \ 0)'$ e $\underline{\delta}_2 = (1 \ 0 \ 0 \ -1 \ -1 \ 0)'$, i.e., $\underline{\delta} = c\underline{\delta}_1 + d\underline{\delta}_2$, $c, d \in \mathbb{R}$. O facto da gama de valores de $\underline{\delta}$ estar contida estritamente em $\mathbb{R}^5 \times \mathbb{R}_+$ implica que G não seja um grupo transitivo em Ξ , por outras palavras, que Ξ/G seja formado por mais do que uma órbita. Por outro lado, constata-se igualmente que

$\underline{\psi}(\underline{\theta}) = (\mu + \alpha_1 + \gamma_1 \ \mu + \alpha_1 + \gamma_2 \ \mu + \alpha_2 + \gamma_1 \ \mu + \alpha_2 + \gamma_2 \ \sigma^2)'$ é um exemplo de uma função invariante maximal relativamente a G . A averiguação da identificabilidade de estatísticas reais mediante funções identificantes (prop. 2.5) pode ser equivalentemente feita através de funções invariantes maximais face ao grupo G (teor. 2.20).

2.2.4 - A não identificabilidade e o tratamento Bayesiano

A relevância da identificabilidade para os procedimentos inferenciais apresenta graduações distintas nas abordagens clássica e Bayesiana. No que concerne aos métodos clássicos de estimação, vimos na secção 2.1 que a identificabilidade (local) de um parâmetro é uma condição necessária para a unicidade (local) da sua estimação. Não faz assim sentido proceder-se à estimação de um modelo não identificável. Além disso, certas propriedades de optimalidade assintótica do respectivo estimador ficam comprometidas, como é o caso da consistência. Estes dois aspectos conduzem à inviabilização de testes assintóticos de hipóteses, ainda que o modelo seja identificável sob a hipótese nula. Pense-se, por exemplo no teste de Wald quando o espaço paramétrico restringido $\Xi = \{\underline{\theta} \in \Xi : \underline{\lambda}(\underline{\theta}) = (\lambda_1(\underline{\theta}) \dots \lambda_t(\underline{\theta}))' = \underline{0}_{(t)}\}$ é identificável. Como o estimador de máxima verosimilhança irrestrito, $\hat{\underline{\theta}}$, não é único, o facto de $\underline{\lambda}(\underline{\theta})$ não ser identificável (vide 2.1.5) leva por si só a que a estatística de Wald não seja bem definida. Outro factor que justifica a inviabilidade da expressão da estatística do teste reside no facto de esta envolver $\underline{I}^{-1}(\hat{\underline{\theta}})$, já que $\hat{\underline{\theta}}$ não é consistente e $\underline{I}(\hat{\underline{\theta}})$ pode ser singular. De facto, nas situações usuais, a não identificabilidade de $\underline{\theta} \in \Xi$ implica a singularidade de $\underline{I}(\underline{\theta})$ (vide critério II em 2.1.3). Os problemas com o teste da razão das verosimilhanças surgem sobretudo na determinação da distribuição assintótica da respectiva estatística, em face da inconsistência de $\hat{\underline{\theta}}$, já que a identificabilidade da função de verosimilhança, quaisquer que sejam os dados, permite calcular o valor da estatística do teste, desde que se consiga determinar um valor para $\hat{\underline{\theta}}$. O facto de o teste de Rao envolver apenas o estimador de máxima verosimilhança restrito, $\tilde{\underline{\theta}}$, possibilita em princípio as inferências. O facto de a correspondente estatística envolver a inversa da matriz $\underline{I}(\tilde{\underline{\theta}})$, que pode ser singular, é um problema que pode ser sempre resolvido por reparametrização. É o que acontece com a sua aplicação à família de distribuições multinomiais - correspondendo à forma conhecida do teste do χ^2 para dados categorizados - na presença de dados incompletos, como nas situações descritas no exemplo IV

de 2.1.2. Na notação aí utilizada, a reparametrização para o parâmetro λ , identificável por hipótese, elimina qualquer problema na definição da estatística de ajustamento e na determinação da sua distribuição nula assintótica.

São problemas deste tipo que, numa óptica clássica, levam a uma tentativa de eliminação da não identificabilidade assente geralmente na imposição de restrições exactas. No quadro acima descrito, se $r < t$ das restrições $\lambda_{\perp}(\theta) = 0$, $i=1, \dots, t$ são suficientes para assegurar a identificabilidade, então essas r restrições são às vezes usadas para definir o novo espaço paramétrico, sob o qual vão ser testadas as $t-r$ restrições restantes. O problema original do teste de hipóteses é abandonado e substituído por um teste de natureza condicional, o teste das $t-r$ restrições condicionado na verificação das r restrições de identificabilidade. É neste quadro que se inserem, por exemplo, os testes assintóticos construídos por Silvey (1959) e Aitchison e Silvey (1960) de forma idêntica à dos correspondentes testes sob um modelo originalmente identificável.

A ânsia de identificação do modelo original com vista a uma viabilização inferencial pode levar à prática condenável de adopção de restrições de natureza nada menos que arbitrária. Parafraseando Koopmans e Reiersol (1950) "Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested". A existência de funções paramétricas identificáveis aponta claramente uma via inferencial alternativa capaz de aumentar o nosso conhecimento sobre os aspectos desconhecidos do modelo. Quando as próprias restrições emanam de uma análise sobre o problema em causa baseada em considerações teóricas e/ou em observações anteriores, a forma categórica, incisiva, exacta que reveste a informação à priori materializada nas restrições levanta novas dificuldades dada a natureza geralmente pouco precisa desse tipo de informação.

Este tipo de dificuldades atenuar-se-á com uma especificação estocástica de toda ou de parte da informação à priori, dada a sua natureza mais flexível e, portanto, mais justificável, o que não só é permitido mas também exigido pelo processo inferencial Bayesiano, seja o modelo identificável ou não.

A conjugação da informação à priori estocástica com a informação amostral através do teorema de Bayes implica que a distribuição à posteriori seja própria sempre que a distribuição à priori o for, quer se verifique ou não a identificabilidade do modelo paramétrico. Em particular, a existência dos momentos à posteriori não exige (nem decorre da) identificabilidade. É certamente isto que está na base da lacónica afirmação de Lindley (1971, pg 46): "In passing it might be noted that unidentifiability causes no real difficulty in the Bayesian approach".

Contudo, a observação anterior implica que as dificuldades decorrentes da não identificabilidade passam a estar ligadas à definição de uma distribuição à priori própria para todos os parâmetros do modelo, o que está longe de ser pacífico. Por exemplo, dada a ligação entre as distribuições informativas conjugadas e a materialização da experiência acumulada, torna-se legítimo indagar como definir adequadamente a distribuição à priori, condicional em σ^2 , para o parâmetro β do modelo linear normal de posto incompleto (vide exemplo I em 2.1.2). Dado o tipo de especificação do modelo, torna-se mais simples exigir uma distribuição à priori informativa para $p-r$ parâmetros, determinados por uma reparametrização não singular do modelo - vejam-se os exemplos de Paulino (1985, cap. 1) e de Leamer (1978, pg 188). Com efeito, essa distribuição à priori própria para um subconjunto dos parâmetros é suficiente para originar uma distribuição à posteriori própria para todos os elementos de β (vide Paulino 1985, pg 9). Naturalmente que este aspecto simplifica mas não elimina o problema de especificar, o menos arbitrariamente possível, uma distribuição à priori que assegure uma distribuição à posteriori própria. No entanto, outras situações existem, como aquelas em que a não identificabilidade do modelo resulta da natureza incompleta dos dados correntes, determinada por delineamento ou por intervenção de factores aleatórios, onde a definição de distribuições à priori informativas já não se afigura tão problemática.

A especificação de uma distribuição à priori ν para θ leva à definição em cada classe de equivalência observacional, definida por uma estatística identificante $\psi: (\mathcal{E}, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1)$, de uma distribuição à priori condicional sobre os elementos dessa classe, que denotamos por $\nu(\cdot | \psi)$. O processo clássico de identificação, assente na especificação de restrições exactas, pode ser visto neste contexto através da definição de $\nu(\cdot | \psi)$, para cada ψ em $\Psi(\mathcal{E})$, como uma distribuição 0-1, i.e., tomando o valor 1 apenas no ponto θ de cada classe que satisfaz as restrições. No quadro Bayesiano, o carácter de generador desta distribuição, quando bem definida, não é tão pronunciado, uma vez que ela apresenta uma função densidade num hiperplano de dimensão inferior à de \mathcal{E} .

A condição iii) da proposição 2.2. revela que essencialmente $\nu_y(\cdot | \psi) = \nu(\cdot | \psi)$, $\forall y \in \mathcal{Y}$, ou seja, dentro de cada classe de equivalência, a distribuição à priori de θ não é actualizada pelas observações. A razão das chances à posteriori de dois pontos observacionalmente equivalentes é assim independente de y e igual à correspondente razão das chances à priori. Vimos ainda (corolário ao teorema 2.15) que no caso dominado e sob a regularidade de P para \mathcal{Q} , a suficiência Bayesiana de ψ implica uma factorização da função densidade à posteriori de θ em dois factores, um dependente de θ , através de ψ , e de y e outro independente de y . Tal factorização traduz que, uma vez actualizada a distribuição de ψ , o papel das observações fica esgotado ao não exercerem mais nenhum efeito sobre o que resta. Isto é, sobre qualquer outra função paramétrica $\gamma(\theta)$ que se pode definir em ordem a haver uma correspondência biunívoca entre (ψ, γ) e θ (e.g., no exemplo 1 desta secção com

$\psi = \mu_1 / (\mu_1 + \mu_2)$, as funções $1_{\underline{y}}^{\mu}$ ou $(1 \ 0)_{\underline{y}}$ podem descrever a função, agora escalar, γ).

Este resultado conduz a que as análises Bayesianas possam ser feitas no espaço mensurável de uma estatística identificante no sentido expresso por Kadane (1974, teor. 5). Seguindo os seus passos, demonstraremos aqui uma extensão linear desse resultado e comentaremos ilustrativamente o seu significado.

Mantendo a notação de B.2, seja $g(\underline{\theta}|\underline{y}) = f(\underline{y}|\underline{\theta})/f(\underline{y})$, com $f(\underline{y}) = E_{\nu}[f(\underline{y}|\underline{\theta})]$, "a" derivada de Radon-Nikodym de $\nu_{\underline{y}}$ em ordem a ν . Seja ainda $q_{\underline{y}}(\underline{\psi})$ a função densidade à posteriori de $\underline{\psi}$ relativamente a alguma medida σ -finita $\tilde{\lambda}$.

Teor. 2.21: Num modelo não identificável com $\underline{\psi}: (\Xi, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1)$ identificante, a esperança à posteriori de qualquer estatística real $\phi: (\Xi, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_2)$ integrável pode ser calculada em (Ψ, \mathcal{B}_1) através da expressão

$$E_{\nu_{\underline{y}}}[\phi(\underline{\theta})] = \int_{\Psi} E(\phi|\underline{\psi})q_{\underline{y}}(\underline{\psi})d\tilde{\lambda}(\underline{\psi}) \quad (2.17)$$

Dem. Atendendo à definição da esperança condicional $E(\cdot|\mathcal{B}_{\underline{\psi}})$ e ao facto de $g(\underline{\theta}|\underline{y}) = f(\underline{y}|\underline{\psi}(\underline{\theta}))/f(\underline{y})$ ser $\mathcal{B}_{\underline{\psi}}$ -mensurável, temos

$$\begin{aligned} E_{\nu_{\underline{y}}}[\phi(\underline{\theta})] &= \int_{\Xi} \phi(\underline{\theta})g(\underline{\theta}|\underline{y})d\nu(\underline{\theta}) \\ &= \int_{\Xi} E\{\phi(\underline{\theta})g(\underline{\theta}|\underline{y})|\mathcal{B}_{\underline{\psi}}\}d\nu(\underline{\theta}) \\ &= \int_{\Xi} \frac{f(\underline{y}|\underline{\psi}(\underline{\theta}))}{f(\underline{y})} E\{\phi(\underline{\theta})|\mathcal{B}_{\underline{\psi}}\}d\nu(\underline{\theta}) \end{aligned}$$

Designando $E(\phi|\underline{\psi})$ a esperança condicional de ϕ em (Ψ, \mathcal{B}_1) e $\tilde{\nu}$ a medida à priori nesse espaço induzida por ν , e fazendo $h(\underline{y}|\underline{\psi}) = f(\underline{y}|\underline{\psi}(\underline{\theta}))$, para $\underline{\psi} = \underline{\psi}(\underline{\theta})$ resulta

$$E_{\nu_{\underline{y}}}[\phi(\underline{\theta})] = \int_{\Psi} \frac{h(\underline{y}|\underline{\psi})}{f(\underline{y})} E(\phi|\underline{\psi})d\tilde{\nu}(\underline{\psi}) = \int_{\Psi} E(\phi|\underline{\psi})q_{\underline{y}}(\underline{\psi})d\tilde{\lambda}(\underline{\psi}), \text{ c.q.d.} \quad \Delta$$

Este resultado exprime assim que a determinação de características à posteriori relativas a um parâmetro não identificável pode ser feita no espaço induzido por uma estatística identificante. A análise de (2.17) revela, nomeadamente, dois aspectos:

- Quando a distribuição à priori de ϕ é independente de $\underline{\psi}$, os momentos à posteriori de ϕ não diferem dos seus momentos à priori - é o que acontece, no exemplo 1, com a função não suficiente e não identificável $\phi = \mu_1 + \mu_2$, distribuída segundo $\Gamma(b, 1)$.

- Quando ϕ é identificável, e portanto \mathcal{B}_ψ -mensurável, os seus momentos à posteriori são calculáveis como tal em (Ψ, \mathcal{B}_1) , e.g., sendo $\phi = \rho(\psi)$, $E_{\tilde{\nu}_y}[\phi(\theta)] = E_{\tilde{\nu}_y}[\rho(\psi)]$, onde $\tilde{\nu}_y$ é a medida à posteriori induzida por ψ .

Para ilustração deste último ponto, voltemos ao exemplo 1 e seja $\phi = \mu_1/\mu_2$. O seu carácter identificador é evidenciado pelo facto de

$f(y|\mu) = \binom{N}{y} (\mu_1/\mu_2)^y [1 + (\mu_1/\mu_2)]^{-N}$. Atendendo a que $\psi|y \rightsquigarrow \text{Be}(a_1, a_2)$ e ao facto de $\phi = \psi/(1-\psi)$ resulta pela distribuição beta

$$E_{\tilde{\nu}_y}[\psi/(1-\psi)] = \frac{B(a_1+1, a_2-1)}{B(a_1, a_2)} = \frac{a_1}{a_2-1} \quad \text{se } a_2 > 1$$

Este resultado traduz a média de uma distribuição beta de 2ª espécie de parâmetros a_1 e a_2 , caracterizada pela função densidade

$$g(\alpha|a_1, a_2) = [B(a_1, a_2)]^{-1} \frac{\alpha^{a_1-1}}{(1+\alpha)^{a_1+a_2}}, \quad a_1, a_2 > 0, \alpha > 0$$

Esta distribuição é precisamente a distribuição à posteriori de $\phi = \mu_1/\mu_2$, conforme se prova facilmente a partir de $q_y(\underline{\mu})$ pela transformação de $\underline{\mu}$ para (ϕ, μ_2) .

Antes de prosseguirmos com a análise das implicações inferenciais de um modelo não identificável, vamos ilustrar com mais um exemplo, os aspectos essenciais da secção 2.2. Embora a análise feita seja idêntica à do exemplo 1, a sua escolha deve-se ao facto de ele traduzir uma situação de dados incompletos e ilustrar alguns traços essenciais do nosso tratamento Bayesiano do problema geral de dados categorizados incompletos descrito nos capítulos que se seguem.

Exemplo 2: Suponhamos que de uma população de 4 categorias com probabilidades θ_i , $i=1,2,3,4$, se obteve uma amostra aleatória de tamanho N . Em vez do resultado usual de conhecimento das frequências x_i de cada categoria, os registos traduziram apenas a informação sobre o nº de elementos classificados ou não nas categorias 1 e 4 juntas, i.e., sobre $x_1 + x_4 = y$ e $x_2 + x_3 = N-y$.

Admitamos que y pode ser modelado pela distribuição binomial de parâmetros N e $\theta_1 + \theta_4$ e adoptemos para $\underline{\theta} = (\theta_1 \theta_2 \theta_3 \theta_4)'$, $\sum_i \theta_i = 1$, uma distribuição à priori que é conjugada da verosimilhança multinomial dos dados completos, i.e., $\underline{\theta}|\underline{b} \rightsquigarrow D_4(\underline{b})$, $\underline{b} = (b_1 b_2 b_3 b_4)' \in (\mathbb{R}^+)^4$.

É fácil constatar que a densidade à posteriori de $\underline{\theta}$, parâmetro claramente não identificável, se pode pôr na forma (onde $a_1 = b_1 + b_4 + y$ e $a_2 = b_2 + b_3 + N - y$)

$$q_{y}(\underline{\theta}) = \frac{(\theta_1 + \theta_4)^{a_1 - 1} (\theta_2 + \theta_3)^{a_2 - 1}}{B(a_1, a_2)} \times \left\{ \frac{\left(\frac{\theta_1}{\theta_1 + \theta_4}\right)^{b_1 - 1} \left(\frac{\theta_4}{\theta_1 + \theta_4}\right)^{b_4 - 1}}{B(b_1, b_4)} \frac{1}{\theta_1} \times \frac{\left(\frac{\theta_2}{\theta_2 + \theta_3}\right)^{b_2 - 1} \left(\frac{\theta_3}{\theta_2 + \theta_3}\right)^{b_3 - 1}}{B(b_2, b_3)} \frac{1}{\theta_2} \right\}$$

a qual revela a suficiência de $\psi = \theta_1 + \theta_4$ (que é identificável) e a falta de suficiência de $\phi_1 = \theta_1 / (\theta_1 + \theta_4)$ e de $\phi_2 = \theta_2 / (\theta_2 + \theta_3)$ (funções que também não são identificáveis).

Operando a transformação $(\theta_1, \theta_2, \theta_4) \rightarrow (\theta_1, \theta_2, \psi)$ podemos ver que o 2º factor de $q_y(\underline{\theta})$ (entre chavetas) representa a distribuição \tilde{a} posteriori, bem definida, de (θ_1, θ_2) dado ψ . Ela reflecte que $(\theta_1 / \psi) \perp\!\!\!\perp (\theta_2 / (1 - \psi)) \mid \psi, y$ e $(\theta_1 / \psi) \mid \psi, y \rightsquigarrow \text{Be}(b_1, b_4)$ e $(\theta_2 / (1 - \psi)) \mid \psi, y \rightsquigarrow \text{Be}(b_2, b_3)$, coincidindo, pois, com os correspondentes resultados distribucionais \tilde{a} priori (veja-se, e.g., a caracterização da distribuição de Dirichlet da prop. C.1.1 do apêndice C.1). Como era de esperar, a distribuição \tilde{a} posteriori (degenerada) de $\underline{\theta}$ dado ψ , caracterizada pelas relações distribucionais acima e por $\theta_4 = \psi - \theta_1$ com probabilidade 1, não sofre o mínimo abalo com a observação de y .

De acordo com o teorema 2.21, os momentos \tilde{a} posteriori de $\underline{\theta}$ podem ser calculados por (2.17). Assim, como

$$\begin{aligned} E \left(\prod_{i=1}^4 \theta_i^{r_i} \mid \psi \right) &= E \left\{ \psi^{r_1 + r_4} \prod_{i=1,4} \left(\frac{\theta_i}{\psi} \right)^{r_i} (1 - \psi)^{r_2 + r_3} \prod_{i=2,3} \left(\frac{\theta_i}{1 - \psi} \right)^{r_i} \mid \psi \right\} \\ &= \psi^{r_1 + r_4} (1 - \psi)^{r_2 + r_3} E \left(\prod_{i=1,4} \left(\frac{\theta_i}{\psi} \right)^{r_i} \mid \psi \right) E \left(\prod_{i=2,3} \left(\frac{\theta_i}{1 - \psi} \right)^{r_i} \mid \psi \right) \end{aligned}$$

e $\psi \mid y \rightsquigarrow \text{Be}(a_1, a_2)$, segue-se que

$$E_{\psi \mid y} \left(\prod_{i=1}^4 \theta_i^{r_i} \right) = \frac{B(a_1 + r_1 + r_4, a_2 + r_2 + r_3)}{B(a_1, a_2)} \frac{B(b_1 + r_1, b_4 + r_4)}{B(b_1, b_4)} \frac{B(b_2 + r_2, b_3 + r_3)}{B(b_2, b_3)}$$

Notando que $q_y(\underline{\theta})$ está na forma de uma distribuição Dirichlet generalizada (vide C.3), a expressão (67) do apêndice C comprova que esse é o resultado obtido directamente da distribuição \tilde{a} posteriori de $\underline{\theta}$. Δ

A argumentação atrás desenvolvida manifestou a importância de um dado tipo de distribuição \tilde{a} priori para possibilitar a efectivação de uma análise \tilde{a} posteriori sobre um modelo não identificável. É conveniente salientar que as condições impostas em ν não asseguram a unicidade da moda \tilde{a} posteriori - o correspondente numérico do estimador de máxima verosimilhança para uma medida ν localmente uniforme. Com efeito, denotando $\underline{\theta}^*$ um valor que maximiza $q(\underline{\theta} \mid y)$, a plurimodalidade da distribuição condicional em $[\underline{\theta}^*]$ implica a não unicidade de $\underline{\theta}^*$. Ou seja, a unicidade da moda \tilde{a} posteriori exige a condição mais restritiva de unimodalidade da distribuição \tilde{a} priori condicional \tilde{a} classe de equivalência a que pertence essa moda. De qualquer

modo, como salienta Drêze (1974), a unicidade da moda $\hat{\theta}$ a posteriori não reveste, de um ponto de vista inferencial, a importância que uma estimação bem definida tem nos métodos clássicos.

Os aspectos focados nesta sub-seção parecem assim conferir à abordagem Bayesiana uma maior maleabilidade e naturalidade no desbloqueamento do processo inferencial sobre um parâmetro não identificável. A menor relevância da identificabilidade no contexto Bayesiano não significa que a sua ausência possa ser subestimada. A especificação de uma distribuição $\hat{\theta}$ a priori que garanta uma distribuição $\hat{\theta}$ a posteriori bem definida pode ser problemática. As distribuições $\hat{\theta}$ a priori e $\hat{\theta}$ a posteriori condicionais numa estatística paramétrica suficiente coincidem, independentemente dos dados. Para uma dada distribuição $\hat{\theta}$ a priori existem funções sobre as quais a experiência é não informativa.

O interesse na identificabilidade pode mesmo visualizar-se em termos de próprio conceito de experiência informativa. Com efeito, Kadane (1974) provou que a estatística $\hat{\theta}$ é identificável sse a experiência é informativa para $\hat{\theta}$ relativamente a qualquer distribuição $\hat{\theta}$ a priori concentrada em dois pontos de Ξ discriminados por $\hat{\theta}$. Num contexto de teoria da decisão, Kadane (1974) apresenta ainda uma outra caracterização de identificabilidade através do relacionamento com o conceito de valor esperado da experiência.

O que nos parece importante realçar em jeito de conclusão, é a sensatez e a honestidade que devem nortear o estatístico quando, numa situação concreta, se defronta com um modelo postulado padecendo de falta de identificabilidade. Quaisquer tendências imediatistas visando o abandono puro e simples do modelo ou a sua desfiguração através de operações de estética inadequadas por arbitrárias devem ser combatidas. Mesmo que muitas características realmente típicas do modelo não sejam acessíveis, analisar o que é analisável é uma alternativa mais credível. E neste quadro, a opção por caminhos Bayesianos pode levar-nos mais longe, em função dos escolhos menos acentuados que a presença de um modelo "doente" com falta de identificabilidade cria. Esta possibilidade é claramente exemplificada pela análise do problema geral de dados categorizados incompletos feita nos capítulos seguintes.

3. FORMULAÇÃO E DISCUSSÃO DO PROBLEMA DE DADOS CATEGORIZADOS INCOMPLETOS

3.1. Introdução

Consideremos uma população de dimensão suficientemente grande com s categorias da qual é extraída uma amostra aleatória de tamanho N . Seja $\underline{\theta} = (\theta_1, \dots, \theta_s)'$ com $\mathbf{1}'\underline{\theta} = 1$ onde θ_i representa a probabilidade (positiva) de qualquer indivíduo aleatoriamente seleccionado pertencer à categoria i , $i=1, \dots, s$. Representemos por $\{e_i, i=1, \dots, s\}$ os vectores da base canónica de \mathbb{R}^s e definamos os vectores aleatórios $\underline{W}_k = (W_{k1}, \dots, W_{ks})'$, $k=1, \dots, N$ de tal modo que $\underline{W}_k = e_i$ indica que o k -ésimo indivíduo seleccionado pertence à categoria i , $i=1, \dots, s$. Deste modo, temos que $\underline{W} = \{\underline{W}_k, k=1, \dots, N\} \mid \underline{\theta} \rightsquigarrow M_s(1, \underline{\theta})$.
iid

Frequentemente ocorre que o processo de amostragem \underline{W} é completamente observado, dele resultando as frequências amostrais de cada categoria. Nesta situação de dados completos, a análise estatística usual toma em consideração apenas o processo de amostragem, pressupondo a impossibilidade deste ser incompletamente observado. Contudo, em certos casos, o facto de ter sido obtida uma amostra completamente categorizada não significa que não seja possível a ocorrência noutra experiência de dados incompletos. Em muitas situações, os próprios dados obtidos revelam que para vários subconjuntos, próprios ou não, da amostra, certas categorias surgem como indistinguíveis no sentido em que não se consegue discriminar umas das outras. Este facto pode ser interpretado considerando a presença de um processo observável de registo (ou de resposta), geralmente aleatório, que actua no sentido de poder censurar de algum modo o resultado, suposto não observado, do processo de amostragem.

A ocorrência deste quadro mais complexo é muitas vezes enfrentada ignorando os dados incompletamente categorizados e traçando as inferências de interesse como se a amostra fosse constituída apenas pelos elementos completamente categorizados. O nosso propósito consiste precisamente em analisar o problema considerando, à partida e explicitamente, um mecanismo de registo que seja o responsável pelo padrão observável de censura. Assumiremos, desde já, que o mecanismo de registo não contradiz os resultados do processo não observado de categorização amostral. Os registos ou respostas correspondem sempre a classificações correctas. Esta suposição básica exclui assim do problema os denominados erros de classificação.

Neste contexto, o padrão mais geral de censura envolve todas as $\sum_{m=1}^s \binom{s}{m} = 2^s - 1$ possibilidades de classificação. Cada indivíduo é classificado em um e um só subconjunto entre todos os subconjuntos de categorias compatíveis com a categoria a que realmente pertence. Desta forma, os dados obtidos consistem nas fre-

quências respeitantes:

- aos conjuntos singulares (unitários) de categorias, correspondentes aos elementos classificados completamente (observações não censuradas);
- aos conjuntos de pelo menos 2 e no máximo $s-1$ categorias relativos aos elementos classificados parcialmente (observações parcialmente censuradas); e
- ao conjunto englobando as s categorias, representando o número de elementos não classificados (ou que sofreram censura total).

Em ordem a explicitar o mecanismo estocástico de registo, denotaremos por λ_i^δ a probabilidade de um indivíduo pertencendo à categoria i (suposta constante para todo o indivíduo nessas condições) ser classificado no subconjunto de categorias δ . Assumiremos que os parâmetros $\{\lambda_i^\delta\}$ são distintos de θ - ou, por outras palavras, que esses dois tipos de parâmetros são de variação independente - no sentido em que o seu espaço paramétrico conjunto é o produto cartesiano dos espaços relativos aos $\{\lambda_i^\delta\}$ e a θ .

A suposição básica de correcção dos registos implica assim que $\lambda_i^\delta = 0$ sempre que $i \notin \delta$. Desta forma, a definição do processo de registo acarreta que os $\{\lambda_i^\delta\}$ satisfaçam as restrições $\sum_{\delta: i \in \delta} \lambda_i^\delta = 1, i=1, \dots, s$.

Os parâmetros $\{\lambda_i^\delta\}$ podem representar-se numa tabela de dupla entrada com as linhas e colunas indicando, respectivamente, as s categorias e os 2^s-1 subconjuntos possíveis de categorias. Nas celas (i, δ) com $i \notin \delta$ surgem zeros com o sentido dos zeros estruturais das tabelas de contingência. Com fins ilustrativos, apresenta-se em seguida esse arranjo das probabilidades condicionais de registo para o caso $s=4$. Por conveniência, os subconjuntos δ serão sempre representados sem chavetas e sem vírgulas separando os seus elementos.

$i \setminus \delta$	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234	Σ
1	λ_1^1	0	0	0	λ_1^{12}	λ_1^{13}	λ_1^{14}	0	0	0	λ_1^{123}	λ_1^{124}	λ_1^{134}	0	λ_1^{1234}	1
2	0	λ_2^2	0	0	λ_2^{12}	0	0	λ_2^{23}	λ_2^{24}	0	λ_2^{123}	λ_2^{124}	0	λ_2^{234}	λ_2^{1234}	1
3	0	0	λ_3^3	0	0	λ_3^{13}	0	λ_3^{23}	0	λ_3^{34}	λ_3^{123}	0	λ_3^{134}	λ_3^{234}	λ_3^{1234}	1
4	0	0	0	λ_4^4	0	0	λ_4^{14}	0	λ_4^{24}	λ_4^{34}	0	λ_4^{124}	λ_4^{134}	λ_4^{234}	λ_4^{1234}	1

Neste quadro, o padrão de dados frequentemente obtido revela a existência de muitas frequências nulas. As frequências positivas podem, às vezes, organizar-se em tabelas completa e/ou parcialmente categorizadas. Por exemplo, suponhamos que no caso anterior as categorias traduzem as 4 celas, ordenadas lexicograficamente, de uma tabela 2×2 . Se houve classificação apenas em $\{1,4\}$ e $\{2,3\}$, as frequências podem representar-se numa tabela 2×2 com as celas diagonais com-

binadas. Se se registaram observações nos subconjuntos $\{i\}$, $i=1,2,3,4,\{1,2\}$, $\{3,4\}$, $\{1,3\}$ e $\{2,4\}$, tais frequências podem representar-se numa tabela 2×2 completa e nas duas tabelas marginais suplementares.

As frequências nulas podem interpretar-se de diferentes modos consoante o padrão de censura que for admitido. Constituirão necessariamente zeros amotrais no quadro do mecanismo mais geral de censura. Na presença de padrões específicos de censura que estipulem a impossibilidade de observação de certos δ 's - o que será traduzido afirmando que $\lambda_i^\delta = 0$ com certeza - algumas dessas frequências (eventualmente todas) significarão zeros estruturais. No exemplo anterior, o primeiro padrão observado citado poderá ser interpretado como consequência de se ter admitido que todos os λ_i^δ são nulos à excepção dos respeitantes a $\delta = \{1,4\}$, $\{2,3\}$ e $i \notin \delta$. Devido às restrições naturais, esta suposição equivale a admitir um mecanismo de censura determinístico. É o que acontece quando, devido à natureza da amostragem ou do próprio fenómeno observável, as celas 1 e 4, por um lado, e 2 e 3, por outro, são indistinguíveis para todo indivíduo. Na mesma linha de raciocínio, a ocorrência apenas da tabela 2×2 completamente categorizada corresponderá ao problema usual de dados completos se for admitido que apenas os λ_i^i , $i=1,2,3,4$, são diferentes de zero. De modo análogo, a observação da tabela completa acompanhada das sub-tabelas marginais poderá ser encarada, como acontece em grande parte da literatura concernente a esse problema, como resultado de um mecanismo aleatório de classificação estabelecendo apenas 3 possibilidades de censura para todo o elemento: não censura e censura do valor de cada uma das duas variáveis categorizadas em separado.

Em vez de formularmos o problema com base no padrão mais geral de censura, o que poderá levar à inclusão de alguns δ 's impossíveis de ocorrer, consideraremos em alternativa um padrão genérico de censura com inclusão apenas dos δ 's admissíveis (isto é, daqueles para os quais se tem, com certeza a priori, $\lambda_i^\delta > 0$ para pelo menos um $i=1,\dots,s$), cujo conjunto é denotado por Δ .

Assumiremos que $\mathbb{T} = \{\{i\}, i=1,\dots,s\}$ e $\{\delta_0\}$, $\delta_0 = \{1,2,\dots,s\}$ são subconjuntos de Δ , ou seja, que todo o elemento amostral tem uma probabilidade positiva de não ser censurado e que a censura total é uma possibilidade efectiva para pelo menos alguns elementos. Denotaremos ainda por \mathbb{T} o conjunto dos δ 's admissíveis constituídos por pelo menos 2 e no máximo $s-1$ categorias (em número de ℓ , $0 \leq \ell \leq 2^s - s - 2$). Assim, $\Delta = \mathbb{T} \cup \mathbb{T} \cup \{\delta_0\}$.

O objectivo da análise deste problema de dados incompletos está em traçar inferências sobre θ . Nesse sentido, as probabilidades condicionais de registo serão vistas como parâmetros perturbadores (ou excedentes).

3.2. Caso de um padrão genérico de censura

Para a definição do processo gerador do tipo de classificação, podemos considerar para cada $k=1, \dots, N$ o vector aleatório $\underline{m}_k = (m_k^\delta, \delta \in \Delta)'$, cujos valores possíveis são os vectores da base canônica de \mathbb{R}^{s+l+1} , designados por $\underline{\varepsilon}_\delta$. Isto é, $\underline{\varepsilon}_\delta$ tem todas as coordenadas nulas à excepção daquela correspondente à posição de m_k^δ em \underline{m}_k que é 1. Assim, $\underline{m}_k = \underline{\varepsilon}_\delta$ indica que o indivíduo k é classificado em δ (não censurado se $\delta \in \overline{\Delta}$, censurado parcialmente se $\delta \in T$ e censurado totalmente se $\delta = \delta_0$). Seja ainda $\underline{m} = (m_k', k=1, \dots, N)'$ e $\Delta_i = \{ \delta \in \Delta : i \in \delta \}$, cuja cardinalidade é representada por t_i , $i=1, \dots, s$.

Assumindo que (W_k, \underline{m}_k) , $k=1, \dots, N$ são i.i.d., a função de probabilidade conjunta de $(\underline{W}, \underline{m})$ é

$$f(\underline{W}, \underline{m} | \underline{\theta}, \{\lambda_i^\delta\}) = \prod_{k=1}^N f(W_k | \underline{\theta}) f(\underline{m}_k | W_k, \{\lambda_i^\delta\}) = \prod_{i=1}^s \prod_{\delta \in \Delta_i} (\theta_i \lambda_i^\delta)^{y_{i\delta}} \quad (3.1)$$

onde $y_{i\delta} = \sum_{k: W_k = e_i} m_k^\delta = \sum_{k: m_k = \underline{\varepsilon}_\delta} W_{ki}$ representa a frequência amostral (não observada para $\delta \in T \cup \{\delta_0\}$) dos indivíduos que pertencem à categoria i ($i=1, \dots, s$) e são registados no subconjunto $\delta \in \Delta_i$. Estamos admitindo que $y_{i\delta} = 0$ sempre que $\lambda_i^\delta = 0$ com certeza a priori e que $0^0 = 1$, como é usual.

Definamos

$$\underline{y}_c = (y_{11}, \dots, y_{ss})' \quad ; \quad \underline{y}_0 = (y_{1\delta_0}, \dots, y_{s\delta_0})'$$

$$\underline{y}_p = (\underline{y}^\delta, \delta \in T) \quad ; \quad \underline{y}^\delta = (y_{i\delta}, i \in \delta)'$$

Notando que $\sum_{i=1}^s \sum_{\delta \in \Delta_i} \theta_i \lambda_i^\delta = 1$ e $\sum_{i=1}^s \sum_{\delta \in \Delta_i} y_{i\delta} = N$, a expressão (3.1) revela que $\underline{y} = (\underline{y}_c' \underline{y}_p' \underline{y}_0')$ tem, condicionalmente a $(\underline{\theta}, \{\lambda_i^\delta\})$, uma distribuição multinomial de dimensão $d-1$, onde $d=2s + \sum_{\delta \in T} d_\delta$ com d_δ indicando o cardinal de δ .

Tendo em vista o modo como foi definido o processo de registo, os valores de \underline{m} definem cabalmente os dados observados a partir dos quais se obtêm as frequências observadas $n_\delta = \sum_{k=1}^N m_k^\delta$, $\delta \in \Delta$.

O processo mais expedito de derivar a função de verosimilhança da amostra consiste em determinar a distribuição marginal de cada vector \underline{m}_k . Como para $\underline{m}_k = \underline{\varepsilon}_\delta$

$$\begin{aligned} f(\underline{\varepsilon}_\delta | \underline{\theta}, \{\lambda_i^\delta\}) &\equiv P(\underline{m}_k = \underline{\varepsilon}_\delta | \underline{\theta}, \{\lambda_i^\delta\}) \\ &= \sum_{i \in \delta} P(W_k = e_i | \underline{\theta}) P(\underline{m}_k = \underline{\varepsilon}_\delta | W_k = e_i ; \{\lambda_i^\delta\}) = \sum_{i \in \delta} \theta_i \lambda_i^\delta \end{aligned}$$

segue-se imediatamente que a distribuição do padrão de dados \underline{m} , numericamente idêntica à função de verosimilhança, é

$$f(\underline{m}|\underline{\theta}, \{\lambda_i^\delta\}) = \prod_{k=1}^N f(m_k|\underline{\theta}, \{\lambda_i^\delta\}) = \prod_{\delta \in \Delta} \left(\sum_{i \in \delta} \theta_i \lambda_i^\delta \right)^{n_\delta}$$

da qual se pode obter a distribuição da estatística suficiente para $(\theta, \{\lambda_i^\delta\})$, $\underline{n} = (n_\delta, \delta \in \Delta)'$.

Esta distribuição pode ser alternativamente derivada através de (3.1), da do que \underline{n} se pode exprimir como um agrupamento dos elementos de \underline{y} . Com efeito, particionando as frequências observadas em $\underline{n} = (\underline{N}' \underline{N}' \underline{n}_o)'$ onde $\underline{N}'_c = (n_\delta, \delta \in \overline{11})$, $\underline{N}'_p = (n_\delta, \delta \in T)$ e $n_o = n_\delta$, temos $\underline{N}'_c = \underline{y}_c$, $n_\delta = \underline{1}'_d \underline{y}_\delta$, $\delta \in T$ e $n_o = \underline{1}'_s \underline{y}_o$. Da distribuição de \underline{y} segue-se então que \underline{n} tem a distribuição multinomial $(s+l)$ -dimensional definida por

$$\begin{aligned} f(\underline{n}|\underline{\theta}, \{\lambda_i^\delta\}) &= (N! / \prod_{\delta \in \Delta} n_\delta!) f(\underline{m}|\underline{\theta}, \{\lambda_i^\delta\}) \\ &= (N! / \prod_{\delta \in \Delta} n_\delta!) \left\{ \prod_{i=1}^s (\theta_i \lambda_i^i)^{n_i} \prod_{\delta \in T} \left(\sum_{i \in \delta} \theta_i \lambda_i^\delta \right)^{n_\delta} \left(\sum_{i=1}^s \theta_i \lambda_i^o \right)^{n_o} \right\} \end{aligned} \quad (3.2)$$

Se numa amostra concreta não se registaram observações em certos δ 's admissíveis, os factores correspondentes em (3.2), vista como função de verosimilhança, desaparecem.

Em ordem a organizar os $\{\lambda_i^\delta\}$ num vector, denotemos

$$\begin{aligned} \underline{\lambda}_c &= (\lambda_1^1, \dots, \lambda_s^s)' ; \quad \underline{\lambda}_o = (\lambda_1^o, \dots, \lambda_s^o)' \\ \underline{\lambda}_p &= (\lambda^{\delta'}, \delta \in T)' ; \quad \underline{\lambda}^\delta = (\lambda_i^\delta, i \in \delta)' \\ \underline{\beta}_i &= (\lambda_i^\delta, \delta \in \Delta_i)' , \quad i=1, \dots, s \end{aligned}$$

O vector coluna dos d parâmetros λ_i^δ pode então ser representado por $\underline{\lambda} = (\underline{\lambda}'_c \underline{\lambda}'_p \underline{\lambda}'_o)'$ ou pela permutação deste vector definida por $\underline{\beta} = (\underline{\beta}'_1 \dots \underline{\beta}'_s)'$, onde cada subvector $\underline{\beta}_i$ é $(t_i \times 1)$.

Operando uma reparametrização para as probabilidades conjuntas $\mu_{i\delta} = \theta_i \lambda_i^\delta$, $i=1, \dots, s$; $\delta \in \Delta_i$, seja

$$\begin{aligned} \underline{\mu}_c &= (\mu_{ii}, i=1, \dots, s)' ; \quad \underline{\mu}_o = (\mu_{i\delta}, i=1, \dots, s)' \\ \underline{\mu}_p &= (\underline{\mu}^{\delta'}, \delta \in T)' , \quad \underline{\mu}^\delta = (\mu_{i\delta}, i \in \delta)' \end{aligned}$$

O vector coluna de todas as d probabilidades $\mu_{i\delta}$ pode analogamente ser representado por $\underline{\mu} = (\mu'_{c_p \mu'_o})'$ ou pela permutação deste último vector definida por $\underline{\mu}^* = \underline{\theta} \otimes \underline{\beta} \equiv (\theta_1 \beta'_1, \dots, \theta_s \beta'_s)'$.

A função de verosimilhança $L(\underline{\theta}, \underline{\lambda} | \underline{n}) \equiv f(\underline{m} | \underline{\theta}, \underline{\lambda})$ definida acima converte-se assim em

$$L(\underline{\mu} | \underline{n}) = \prod_{i=1}^s \mu_{ii}^{n_i} \times \prod_{\delta \in T} (1'_{-d} \mu^\delta)^{n_\delta} \times (1'_{-s} \mu_o)^{n_o} \quad (3.3)$$

onde $1'_{-d} \mu = 1$.

Esta expressão reflecte que $\underline{\mu}$ é um parâmetro não identificável, uma vez que $L(\underline{\mu} | \underline{n})$ depende de $\underline{\mu}$ apenas através da função $Q' \underline{\mu}$, onde Q' é a matriz $(s+l+1) \times d$

$$Q' = \left(\begin{array}{c|c} I_s & \underline{0} \\ \hline 0 & Q'_I \end{array} \right)$$

com Q'_I a matriz $(l+1) \times (d-s)$ diagonal em blocos de blocos diagonais $1'_{-d\delta}$, $\delta \in T \cup \{\delta_o\}$. Como $1'_{-t_i} \beta_i = 1$, $i=1, \dots, s$, segue-se que $\underline{\theta} = \underline{S}^* \underline{\mu}^*$, onde \underline{S}^* é a matriz $s \times d$ diagonal em blocos com blocos diagonais $1'_{-t_1}, \dots, 1'_{-t_s}$. Torna-se assim evidente que a não identificabilidade de $\underline{\mu}$ impossibilita a realização de inferências clássicas sobre $\underline{\theta}$ sem a imposição de hipóteses adicionais sobre $\underline{\lambda}$.

Uma hipótese extremamente forte é a que considera todos os λ_i^δ conhecidos. Neste caso, recorrendo por exemplo à maximização do logaritmo da função de verosimilhança, restringida por $1'_{-s} \underline{\theta} = 1$, obtemos facilmente as equações de verosimilhança

$$N \hat{\theta}_i = n_i + \sum_{\delta \in T \cup \Delta} n_\delta \frac{\hat{\theta}_i \lambda_i^\delta}{\sum_{j \in \delta} \hat{\theta}_j \lambda_j^\delta} + n_o \frac{\hat{\theta}_i \lambda_i^{\delta_o}}{\sum_{j=1}^s \hat{\theta}_j \lambda_j^{\delta_o}}, \quad i=1, \dots, s \quad (3.4)$$

revelando que apenas as componentes de $\underline{\lambda}$ relacionadas com a censura parcial e total entram no cálculo da sua solução.

Neste ponto, é interessante notar que, se o processo de amostragem fosse completamente observado, os dados completos conduziriam ao vector de frequências $\underline{x} = (x_1 \dots x_s)'$ com $x_i = \sum_{\delta \in \Delta_i} y_{i\delta}$, $i=1, \dots, s$. Sendo \underline{S} a matriz $d \times s$ obtida de \underline{S}^* por permutação dos elementos de cada uma das suas colunas de modo que $\underline{S}' \underline{\mu} = \underline{S}^* \underline{\mu}^* = \underline{\theta}$, então $\underline{x} = \underline{S}' \underline{y}$. Dada a distribuição de \underline{y} , teremos

$$\underline{x} | N, \underline{\theta} \rightsquigarrow M_s(N, \underline{\theta}) \Rightarrow E(x_i | N, \underline{\theta}) = N \theta_i, \quad i=1, \dots, s$$

Por outro lado, a distribuição condicional de \underline{y} dado \underline{n} é o produto das $l+1$ distribuições condicionais

$$\underline{y}^\delta | \underline{n}_\delta; \underline{\mu} \rightsquigarrow M_{d_\delta}(n_\delta, \underline{\mu}^\delta / 1'_{d_\delta} \underline{\mu}^\delta), \delta \in T \quad (3.5)$$

$$\underline{y}_0 | \underline{n}_0; \underline{\mu} \rightsquigarrow M_s(n_0, \underline{\mu}_0 / 1'_s \underline{\mu}_0)$$

pelo que
$$E(x_i | \underline{n}, \underline{\mu}) = n_i + \sum_{\delta \in T \cap \Delta_i} n_\delta \frac{\mu_{i\delta}}{\sum_{j \in \delta} \mu_{j\delta}} + n_0 \frac{\mu_{i\delta_0}}{\sum_{j=1}^s \mu_{j\delta_0}}, \quad i=1, \dots, s$$

Fica assim evidenciado que as equações de verosimilhança acima correspondem à igualdade das esperanças não condicional e condicional (a \underline{n}) da estatística suficiente para os dados completos quando $\underline{\theta} = \hat{\underline{\theta}}$ é um ponto estacionário de $\log L(\underline{\theta} | \underline{\lambda}; \underline{n})$.

Essas equações podem, pois, ser resolvidas iterativamente pelo popular algoritmo EM (vide Dempster et al. [1977]), esperando-se a sua convergência para o estimador de máxima verosimilhança $\hat{\underline{\theta}}$ nos casos em que a função de verosimilhança é bem comportada (vide Wu [1983]), o que depende certamente dos valores de $\underline{\lambda}$. O passo E do algoritmo estima, para o valor corrente $\underline{\theta}^{(v)}$ do parâmetro (usa-se frequentemente como valor inicial $\underline{\theta}^{(0)} = \underline{N}_c / 1'_s \underline{N}_c$), \underline{x} pela média condicional $E(\underline{x} | \underline{n}; \underline{\theta}^{(v)}, \underline{\lambda})$ definida acima. O passo M calcula a nova estimativa de $\underline{\theta}$ para o modelo saturado por $\underline{\theta}^{(v+1)} = \underline{N}^{-1} E(\underline{x} | \underline{n}; \underline{\theta}^{(v)}, \underline{\lambda})$.

A análise do passo E revela que cada unidade parcialmente classificada em $\delta \in T \cup \{\delta_0\}$ é distribuída pelas categorias $i \in \delta$ em frações definidas por $\theta_i^{(v)} \lambda_i^\delta / \sum_{j \in \delta} \theta_j^{(v)} \lambda_j^\delta$. Isto significa que a chance de uma unidade registada em δ pertencer a i relativamente a pertencer a i' , $i, i' \in \delta$, é

$$[\theta_i^{(v)} / \theta_{i'}^{(v)}] [\lambda_i^\delta / \lambda_{i'}^\delta]$$

O conhecimento apenas das razões $\lambda_i^\delta / \lambda_{i'}^\delta$, para todos os pares (i, i') de δ é assim suficiente para a distribuição de n_δ pelas várias categorias de δ . Por outras palavras, a estimação por máxima verosimilhança de $\underline{\theta}$ condicional em $\underline{\lambda}$ prescinde do conhecimento exacto de $\underline{\lambda}$, usando apenas a informação a priori corporizada nos valores atribuídos às chances $\lambda_i^\delta / \lambda_{i'}^\delta$, $i, i' \in \delta \in T \cup \{\delta_0\}$ (para uma ilustração numérica vide Little [1980, pg. 405]).

A análise Bayesiana do modelo (3.3) de Dickey et al. (1987) considera também como conhecido o parâmetro $\underline{\lambda}$. É possível, no entanto, traçar inferências Bayesianas sobre $\underline{\theta}$ sem condicionamento nos parâmetros de um mecanismo geral de registo. A nossa solução desenvolvida no capítulo 5 (sec. 5.2), baseada no uso

de distribuições *a priori* Dirichlet para $(\underline{\theta}, \underline{\beta})$ comprova-o indubitavelmente.

Este problema geral de dados incompletos ilustra, pois, a capacidade que a ferramenta Bayesiana tem, na nossa opinião, de tornar obstáculos derivados de certas situações de não identificabilidade paramétrica. E inclusivamente, fá-lo com o recurso a uma informação *a priori* mais flexível e mais "vaga" do que aquela que permite o desbloqueamento inferencial clássico.

3.2.1. Processo de registo Bayesianamente ignorável

Em ordem a eliminar o parâmetro perturbador $\underline{\lambda}$, a hipótese sobre o mecanismo de registo mais vezes admitida, implícita ou explicitamente, na literatura pode exprimir-se por

$$M_1: \underline{\lambda}^\delta = \underline{1}_{-d_\delta} \lambda_\delta^\delta, \quad \forall \delta \in T$$

$$\underline{\lambda}^0 = \underline{1}_{-s} \lambda_0^\delta$$

onde o escalar λ_δ^δ , $\delta \in T \cup \{0\}$, representa a probabilidade condicional de classificação em δ comum a todos os indivíduos pertencendo a categorias compatíveis com δ .

Este modelo estipula que a probabilidade de se registar a classificação em δ , para todo o δ , não depende dos valores que poderiam ter sido registados mas apenas dos valores efectivamente observados. Ou seja, essa probabilidade só depende do resultado do processo de amostragem para todo o indivíduo k através da informação de que $\underline{W}_k \in \bigcup_{i \in \delta} \{e_i\}$. Na terminologia de Dawid and Dickey (1977), o modelo M_1 traduz um processo de registo baseado no próprio registo ("report-based reporting process").

Note-se que a distribuição do padrão observável \underline{m} condicional em \underline{W} é, por (3.1) e pela definição de \underline{n} , caracterizada sob M_1 pela função de probabilidade

$$f_1(\underline{m} | \underline{W}, \underline{\lambda}) = \prod_{i=1}^s (\lambda_1^i)^{n_i} \prod_{\delta \in T \cup \{0\}} (\lambda_\delta^\delta)^{n_\delta} \quad (3.6)$$

Esta função é assim, dado $\underline{\lambda}$, apenas dependente das frequências observadas, indicando que o modelo M_1 assegura que os dados "missing" são "missing" aleatoriamente (MAR em abreviatura), para empregar a linguagem de Rubin (1976). Atendendo a que o conceito MAR de Rubin se reporta, não a $f_1(\underline{m} | \underline{W}; \underline{\lambda})$, mas apenas à probabilidade condicional do padrão \underline{m} observado, é fácil entender que podemos ter processos MAR sem satisfazerem M_1 . Basta pensar no caso em que as condições M_1 são apenas violadas por um δ_* para o qual se observou $m_k^{\delta_*} = 0$, $\forall k$. A função de probabilidade condicional de \underline{m} já não é dada por (3.6) mas o seu valor no pa-

drão \underline{m} observado é definido por uma expressão que só difere de (3.6) por não incluir o factor relativo a δ_* .

Seja $\underline{Z} = [z_{\delta}, \delta \in \mathbb{T}]$ a matriz particionada $s \times l$ cujos vectores coluna têm coordenadas nulas à excepção daquelas correspondentes a $i \in \delta$ e que são iguais a 1. A função de verosimilhança sob M_1 converte-se então em

$$L_1(\underline{\theta}, \underline{\lambda} | \underline{n}) = \left[\prod_{i=1}^s (\lambda_i^i)^{n_i} \prod_{\delta \in \mathbb{T}} (\lambda_{\delta}^{\delta})^{n_{\delta}} (\lambda_{\delta_0}^{\delta_0})^{n_0} \right] \left[\prod_{i=1}^s \theta_i^{n_i} \prod_{\delta \in \mathbb{T}} (z_{\delta}^i \theta)^{n_{\delta}} \right] \quad (3.7)$$

$$= L_1(\underline{\lambda} | \underline{n}) L_1(\underline{\theta} | \underline{N}_c, \underline{N}_p)$$

A redução do número de parâmetros perturbadores e a ultrapassagem do problema de identificabilidade presente em (3.2), operadas por M_1 , tornam possível, com as frequências observadas positivas, a estimação dos $2s+l+1$ parâmetros $(\underline{\theta}, \underline{\lambda}, \lambda_{\delta_0}^{\delta_0}, \delta \in \mathbb{T} \cup \{\delta_0\})$ sujeitos às $s+1$ restrições naturais

$$\sum_{i=1}^s \theta_i = 1; \quad \lambda_i^i + \sum_{\delta \in \mathbb{T} \cup \{\delta_0\}} \lambda_{\delta}^{\delta} + \lambda_{\delta_0}^{\delta_0} = 1, \quad i=1, \dots, s \quad (3.8)$$

Deve observar-se, no entanto, que o modelo M_1 não garante sempre a identificabilidade paramétrica. Basta atentar no caso extremo em que \prod não é considerado admissível.

Por outro lado, a factorização (3.7) da função de verosimilhança implica que as inferências verosimilhançista e Bayesiana sobre $\underline{\theta}$ (neste último caso com a condição adicional de $\underline{\theta}$ ser distribuído à priori independentemente de $\underline{\lambda}$) podem realizar-se apenas com base em $L_1(\underline{\theta} | \underline{N}_c, \underline{N}_p)$. Deve notar-se que este aspecto é válido para qualquer mecanismo MAR, dado que a verosimilhança admite uma factorização do tipo de (3.7). A suposição MAR conduz assim a que n_0 só exerça influência nas inferências sobre as probabilidades condicionais de registo.

A fixação do padrão de dados observado, digamos \underline{m}_0 , leva a que os dados observados se possam definir como $\underline{W}_0 = \{W_k^{\delta}, k=1, \dots, N, \delta \in \Delta\}$ onde W_k^{δ} traduz a classificação do indivíduo k em δ . Isto é, os dados $\{W_k^{\delta}\}$ são encarados como resultados da observação das variáveis aleatórias Bernoullianas independentes (mas não identicamente distribuídas)

$$W_k^{\delta} = \begin{cases} 1 & \text{se } W_k \in \cup_{i \in \delta} \{e_i\} \\ 0 & \text{c.c.} \end{cases}$$

devendo notar-se que $W_k^{\delta_0}$ é degenerada.

A probabilidade marginal de \underline{W}_0 calculada sem condicionamento explícito no padrão observado \underline{m}_0 , i.e., a partir da distribuição de \underline{W} , é

$$f(\underline{W}_0 | \underline{\theta}) = \prod_{i=1}^s \theta_i^{n_i} \prod_{\delta \in T} (z_{\delta}' \cdot \underline{\theta})^{n_{\delta}} \quad (3.9)$$

A realização de inferências sobre $\underline{\theta}$ baseadas em (3.9) corresponde a ignorar o processo de registo. Face ao exposto acima, este procedimento afigura-se correcto para as abordagens baseadas na verosimilhança sob a validade de qualquer processo de registo MAR. Para uma discussão elucidativa do conceito de ignorabilidade do processo gerador de dados incompletos em geral, veja-se Rubin (1976, 1978) e Little (1980, 1982).

Uma implicação importante do modelo M_1 é que as probabilidades das categorias condicionais aos registos para os indivíduos incompletamente categorizados coincidem com as correspondentes probabilidades condicionais amostrais para todo o indivíduo com as mesmas características observadas. Ou seja, para qualquer k tal que $\underline{m}_k = \underline{\varepsilon}_{\delta}$, $\delta \in TU\{\delta_0\}$

$$P(\underline{W}_k = \underline{e}_i | \underline{m}_k = \underline{\varepsilon}_{\delta}; \underline{\theta}, \underline{\lambda}) = \begin{cases} \mu_{i\delta} / \sum_{j \in \delta} \mu_{j\delta} = \theta_i / (z_{\delta}' \cdot \underline{\theta}), & i \in \delta \\ 0, & i \notin \delta \end{cases} \quad (3.10)$$

$$= P(\underline{W}_k' = \underline{e}_i | \underline{W}_k' \in \cup_{j \in \delta} \{e_j\}; \underline{\theta})$$

No que concerne à determinação do estimador de máxima verosimilhança de $\underline{\theta}$, as equações derivadas da maximização restringida de $\log L_1(\underline{\theta} | N_c, N_p)$

$$\hat{\theta}_i = [1/(N-n_0)] \{n_i + \sum_{\delta \in TU\Delta_i} n_{\delta} \hat{\theta}_i / (z_{\delta}' \cdot \hat{\theta})\}, \quad i=1, \dots, s \quad (3.11)$$

podem ser vistas como resultantes das equações de verosimilhança (3.4), devido a (3.10), e ser interpretadas do mesmo modo. A forma de (3.11) sugere o procedimento iterativo em 2 passos

$$x_i^{(v)} = n_i + \sum_{\delta \in TU\Delta_i} n_{\delta} \theta_i^{(v)} / (z_{\delta}' \cdot \theta^{(v)}), \quad v = 0, 1, 2, \dots \quad (3.12)$$

$$\theta_i^{(v+1)} = x_i^{(v)} / (N-n_0), \quad i = 1, \dots, s$$

o qual pode ser encarado como o algoritmo EM quando os dados observados se resumem apenas a N_c e N_p . Os dados incompletos multinomiais apresentam usualmente uma verosimilhança cujo logaritmo é uma função côncava com um único ponto de máximo correspondente ao único ponto de estacionaridade. Deste modo, é de esperar

que o algoritmo iterativo acima definido convirja normalmente para o estimador de máxima verosimilhança.

Em ordem a clarificar o sentido de um processo de registo MAR consideremos o seguinte exemplo ilustrativo:

EXEMPLO 1: Suponhamos que de uma população de crianças na faixa etária de 3-4 anos sem um grau avançado de cárie, com dimensão considerada suficientemente grande, se colheu uma amostra aleatória de N crianças com o fim de indagar a acção de um tratamento na evolução da cárie de um determinado dente de leite molar.

Para o efeito seja X_1 (respectivamente X_2) a variável medindo o grau de cárie nesse dente antes da administração do tratamento (resp. após um período durante o qual for administrado o tratamento), com os níveis hígido ($i=0$), incipiente ($i=1$) e moderado ($i=2$) (resp. com os níveis hígido ($j=0$), incipiente ($j=1$), moderado ($j=2$) e avançado ($j=3$)). A tabela bidimensional (X_1, X_2) apresenta assim como possíveis apenas as 9 celas (i, j) , $i=0, 1, 2$, $j \geq i$.

Suponhamos que, por força da acção de um determinado processo de registo, a amostra concreta produziu as frequências $\{n_{\delta_j}, j=0, 1, \dots, 5\}$ onde $\delta_j = \{(0, j)\}$, $j = 0, 1, 2, 3$, $\delta_4 = \{(1, j), j=1, 2, 3\}$ e $\delta_5 = \{(2, j), j=2, 3\}$.

Um modelo possível para o processo de registo é definido considerando X_1 sempre observada e X_2 observada se e só se $X_1 = 0$. Neste caso, as configurações admissíveis resumem-se às configurações observadas, constituindo pois uma partição do conjunto das 9 celas. As probabilidades condicionais de classificação são assim conhecidas e tal que

$$\lambda_{0j}^{\delta_j} = 1, \quad j = 0, 1, 2, 3$$

$$\lambda_{11}^{\delta_4} = \lambda_{12}^{\delta_4} = \lambda_{13}^{\delta_4} = 1$$

$$\lambda_{22}^{\delta_5} = \lambda_{23}^{\delta_5} = 1$$

Fazendo $\underline{m}'_k = (m_k^\ell, \ell = 0, 1, \dots, 5)$, onde $m_k^\ell = 0$ ou 1 , $\sum_{\ell=0}^5 m_k^\ell = 1$, com $m_k^\ell = 1$ traduzindo que a criança k foi classificada em δ_ℓ , a distribuição condicional de $\underline{m} = (\underline{m}'_k, k=1, \dots, N)'$ dado \underline{W} pode exprimir-se por

$$f(\underline{m} | \underline{W}) = \prod_{\ell=0}^5 \prod_{k: m_k^\ell=1} \gamma [I_{E_\ell}(\underline{W}_k) - m_k^\ell]$$

onde

$$E_\ell = \bigcup_{(i,j) \in \delta_\ell} \{e_{ij}\}, \quad I_{E_\ell}(\underline{W}_k) = \begin{cases} 1, & \underline{W}_k \in E_\ell \\ 0, & \text{c.c.} \end{cases} \quad \text{e} \quad \gamma(g_k) = \begin{cases} 1, & g_k = 0 \\ 0, & \text{c.c.} \end{cases}$$

O facto de se observar sempre X_1 e, portanto, de se saber se $\underline{W}_k \in E_\ell$ ou não,

implica que a distribuição de \underline{m} dado \underline{W} para este modelo não depende dos valores "missing". Pela sua definição este processo MAR tem ainda a particularidade da distribuição condicional do padrão de registo ser 1. Isto conduz a que a distribuição das frequências observáveis só dependa de θ , coincidindo o seu núcleo com (3.9).

Apesar do padrão observado de frequências, podemos admitir que $\lambda_{ij}^{\delta_6} > 0$, $(i,j) \in \delta_6 = \{(0,j), j=1,2,3\}$, ou seja, que n_{δ_6} é observável. Um modelo que contempla esta possibilidade pode ser definido afirmando que, quando $X_1 = 0$, X_2 é observada completamente se $X_2 = 0$ e pode ser observada parcialmente em δ_6 ou completamente se $X_2 = j, \forall j \neq 0$. Considera-se ainda que X_1 é sempre observada e X_2 nunca é observada quando $X_1 \neq 0$. As configurações admissíveis passam a ser as do modelo anterior acrescidas de δ_6 , e as restrições naturais envolvendo as probabilidades $\lambda_{ij}^{\delta_6}$ diferem das do modelo inicialmente considerado apenas na medida em que as equações $\lambda_{0j}^{\delta_6} = 1, j=1,2,3$ são substituídas por $\lambda_{0j}^{\delta_6} + \lambda_{0j}^{\delta_6} = 1, j=1,2,3$.

Como o tipo de observação de X_2 depende do valor que X_2 terá para cada unidade amostral, é de esperar que, em geral, a distribuição condicional dado \underline{W} do padrão observável de registo, $\underline{m} = (m'_k, k=1, \dots, N)'$, $m'_k = (m_k^\ell, \ell=0,1, \dots, 6)$ dependa de valores não observados. Na verdade, tal distribuição pode definir-se pela expressão

$$f(\underline{m}|\underline{W}, \lambda) = \left\{ \prod_{\ell=0,4,5} \prod_{k:m_k^\ell=1} \gamma [I_{E_\ell}(\underline{W}_k) - m_k^\ell] \right\} \left\{ \prod_{j=1}^3 (\lambda_{0j}^{\delta_6})^{n_{\delta_6} y_{j\delta_6}} \right\}$$

onde $y_{j\delta_6} = \sum_{k:W_k=e_{0j}} m_k^6$ é a frequência não observada de crianças classificadas em δ_6 e pertencentes à cela $(0,j)$.

Como na amostra observada $n_{\delta_6} \equiv \sum_{j=1}^3 y_{j\delta_6} = 0$, segue-se que a probabilidade condicional do padrão observado de registo para este modelo também não depende dos valores "missing". Assim, este processo de registo é igualmente MAR independentemente dos $\{\lambda_{ij}^{\delta_6}\}$ satisfazerem ou não as condições M_1 . Obviamente que este mecanismo deixará de ser MAR se na amostra observada houver pelo menos um k tal que $m_k^6 \neq 0$, a não ser que seja modelado por M_1 , i.e., quando a probabilidade de X_2 ser parcialmente observada em δ_6 só dependa de X_2 na medida em que X_2 seja diferente de zero.

Note-se que a função de verosimilhança de θ baseada na amostra inicialmente referida é a mesma para os dois mecanismos de censura embora a distribuição das frequências observáveis seja naturalmente diferente (veja-se (3.2)). Como tal, é de esperar que as propriedades de procedimentos inferenciais clássicos sobre θ difiram, em geral, para os dois mecanismos. A

Apesar da frequência com que na literatura estatística se admite, implícita ou explicitamente, o modelo M_1 , ele pode ser violado em várias situações, particularmente quando a ausência de registo para certos indivíduos se deve à sua recusa em responder. A suposição de que a não-resposta é independente da verdadeira categoria a que o indivíduo pertence conduz a que n_0 não exerça qualquer influência na estimação de θ - recorde-se (3.7) - quando uma análise sensata à priori do problema concreto pode indicar que o número de não respondentes contém informação sobre θ .

Em ordem a ilustrar o carácter demasiadamente restritivo que M_1 pode ter, considere-se a seguinte situação:

Exemplo 2: Pretende-se indagar a proporção de toxicómanos numa dada população, de dimensão suficientemente grande, dividida nas seguintes categorias consideradas disjuntas

- 1: consumidores de droga pesada
- 2: consumidores de droga leve
- 3: não consumidores de qualquer droga

Os subconjuntos susceptíveis de ser observados através de um inquérito amostral são

$$\Pi = \{ \{1\}, \{2\}, \{3\} \} ; \quad \delta_0 = \{1, 2, 3\}$$

$$\mathbb{T} = \{ \{1, 2\}, \{2, 3\}, \{1, 3\} \}$$

Atendendo ao significado das categorias envolvidas e ao comportamento do meio social face ao problema do consumo de drogas, é natural esperar-se, nomeadamente, que

$$\lambda_i^{123} > \lambda_3^{123} \quad , \quad i = 1, 2$$

$$\lambda_1^{12} > \lambda_2^{12} ; \quad \lambda_2^{23} > \lambda_3^{23} ; \quad \lambda_1^{13} > \lambda_3^{13}$$

A não ser que se obtenham dados completos, este juízo apriorístico acarreta a impossibilidade de o mecanismo de resposta ser ignorável e, por conseguinte, a verosimilhança correcta deve ser dada por (3.3) e não por (3.7). A fixação de valores para todas as chances $\lambda_i^\delta / \lambda_1^\delta$, $i, i' \in \delta \in \mathbb{T} \cup \{\delta_0\}$, permitirá avaliar o impacto do mecanismo de resposta em inferências sobre θ . A

Noutras situações em que a análise à priori da validade de M_1 não se afigura tão linear, a própria definição das configurações admissíveis, mormente quando Π não é admissível e \mathbb{T} não constitui uma partição das s categorias, pode inviabilizar M_1 pela violação das restrições (3.8) concernentes aos parâmetros perturbadores. Para o ilustrar, considere-se o exemplo simples de uma tabela 2×2 com as celas ordenadas lexicograficamente, para a qual se admite que

$\Delta = \left\{ \{1,3,4\}, \{2,4\} \right\}$. A inconsistência das 3 equações

$$\lambda_{134}^{134} = 1 ; \lambda_{24}^{24} = 1 ; \lambda_{134}^{134} + \lambda_{24}^{24} = 1$$

torna o modelo M_1 sem sentido. Desta forma, particularmente quando $\Delta = T$, torna-se necessário assegurar a compatibilidade de M_1 com Δ antes que se comece cegamente a traçar inferências sobre $\underline{\theta}$ a partir de (3.7).

3.2.2. Processo de registo "classicamente" ignorável

Atrás ficou patente que as inferências correctas sobre $\underline{\theta}$ devem apoiar-se na interpretação de que os dados observados (aí expressos pelo valor observado, digamos \underline{m}_0 , do padrão de registo, dado o modo como este foi definido) são gerados a partir da distribuição conjunta de $(\underline{W}, \underline{m})$.

Referiram-se ainda condições que garantem que tal procedimento é equivalente, numa óptica verosimilancista ou Bayesiana, a ignorar o processo de registo. Isto é, equivalente a fazer as inferências de interesse, fixando o padrão de censura no seu valor observado e supondo que os dados observados, descritos por \underline{W}_0 , são gerados da distribuição marginal de \underline{W} .

Em ordem a compreendermos como este último procedimento pode ser considerado correcto, do ponto de vista da inferência clássica, suponhamos que o padrão de censura é redefinido de modo a que já não seja suficiente, por si só, para a definição dos dados observáveis. Isto pode ser feito, no âmbito de uma estrutura geral de dados categorizados, considerando os conjuntos $\delta \in \Delta$ aglutinados em classes constituindo partições de $\{1, \dots, s\}$ - o que exigirá, em geral, a ampliação de Δ - e definindo o padrão de censura pelos vectores \underline{m}_k , $k=1, \dots, N$, do tipo de \underline{m}_k , mas indicando apenas a partição onde cada indivíduo é classificado. Ou seja, somente a componente t , digamos m_k^t , será não nula e igual a 1, se o indivíduo k é classificado na classe t .

Esta nova forma de definir o padrão de registo, embora seja incapaz de traduzir cabalmente os dados, não altera a definição do processo de registo, i.e., da distribuição condicional do padrão de dados dado \underline{W} . De facto, dado que $\underline{W}_k = \underline{e}_i$, $m_k^\delta = 1$ sse $m_k^t = 1$, $k=1, \dots, N$, $i=1, \dots, s$, para o conjunto δ da classe t que contém i . As considerações anteriores sobre um mecanismo MAR mantêm-se assim para esta reformulação do padrão de censura. O valor observado de \underline{m} , digamos \underline{m}_0 , e \underline{W}_0 (definido como antes), em conjunto, passam agora a definir exhaustivamente todos os dados observados. A distribuição multinomial das frequências observáveis \underline{n} passa a ser calculada exclusivamente a partir de (3.1), tendo em conta que $\underline{n} = Q'y$.

Como o acto de ignorar o processo de registo significa, nomeadamente, fixar \underline{m} em \underline{m}_0 , a distribuição amostral correcta de \underline{W}_0 , nessa base, deve ser dada pela

distribuição condicional $f(W_0 | \bar{m}_0; \theta, \lambda)$ e não pela distribuição marginal $f(W_0 | \theta)$. A ignorabilidade do processo de registo, do ponto de vista da inferência clássica, requer então que $f(W_0 | \bar{m}_0; \theta, \lambda)$ coincida com (3.9), o que é claramente garantido se a probabilidade de \bar{m}_0 (ou mais restritivamente, a distribuição de \bar{m}) condicional em W não depende de W . No quadro de uma estrutura geral de dados categorizados incompletos, a reformulação do padrão de registo em termos de \bar{m} permite assim aplicar as condições de Rubin (1976) para ignorabilidade do mecanismo de registo.

Com o objectivo de verificar que tipo de modelos para o mecanismo de registo garante a sua ignorabilidade para propósitos inferenciais sobre θ , via teoria da amostragem, vamos considerar por agora o caso especial em que T constitui uma partição do conjunto de categorias (em l subconjuntos),

Neste caso, os vectores indicadores das modalidades gerais de classificação (ausência de censura, censura parcial e censura total) são definidos por $\bar{m}_k = (m_k^c, m_k^p, m_k^o)'$, $k=1, \dots, N$, tal que $\bar{m}_k | W_k = e_i \sim M_3(1, \beta_i)$, onde $\beta_i = (\lambda_i^1, \lambda_i^\delta, \lambda_i^0)'$, com δ o subconjunto (único) de T que contém i .

Conhecer $\bar{m} = (\bar{m}_k', k=1, \dots, N)'$ implica conhecer a estatística $(N_c, N_p, n_o)'$ = $\sum_{k=1}^N \bar{m}_k$, a qual se pode exprimir alternativamente em termos das frequências observáveis $\underline{n} = (N_c', N_p', n_o)'$ como $\underline{U}\underline{n}$, onde \underline{U} é a matriz $3 \times (s+l+1)$ diagonal em blocos de blocos diagonais $\underline{1}'_s$, $\underline{1}'_l$ e $\underline{1}$.

Dado que a distribuição $f(W, \bar{m} | \theta, \lambda)$ em (3.1) implica, com a reparametização para $\underline{\mu}$ que

$$\underline{n} | N, \underline{\mu} \sim M_{s+l+1}(N, Q' \underline{\mu}) \quad (3.13)$$

resulta então que $\underline{U}\underline{n} | N, \underline{\mu} \sim M_3(N, UQ' \underline{\mu})$, isto é

$$f(N_c, N_p, n_o | N, \underline{\mu}) = \frac{N!}{N_c! N_p! n_o!} \gamma_c^{N_c} \left(\sum_{\delta \in T} \gamma_\delta \right)^{N_p} \gamma_o^{n_o} \quad (3.14)$$

onde

$$\begin{aligned} \gamma_c &= \underline{1}'_s \underline{\mu}_c ; \gamma_o = \underline{1}'_s \underline{\mu}_o \\ \gamma_\delta &= \underline{1}'_{d_\delta} \underline{\mu}^\delta, \delta \in T \end{aligned} \quad (3.15)$$

Consequentemente, a distribuição de (N_c, N_p) condicional a (N_c, N_p) é um produto de multinomiais independentes expressa por

$$f(N_c, N_p | N_c, N_p, \underline{\mu}) = \frac{N_c!}{\prod_{i=1}^s n_i!} \prod_{i=1}^s \left(\frac{\mu_{ii}}{\gamma_c} \right)^{n_i} \times \frac{N_p!}{\prod_{\delta \in T} n_\delta!} \prod_{\delta \in T} \left(\frac{\gamma_\delta}{\sum_{\delta \in T} \gamma_\delta} \right)^{n_\delta} \quad (3.16)$$

Para esta configuração particular de dados incompletos foi possível obter uma estatística, condicional à qual, porém, a distribuição das frequências observáveis depende de λ .

Sob o modelo M_1 , a distribuição (3.16) continua dependendo de $\underline{\theta}$ e $\underline{\lambda}$. Assim, embora $\underline{\theta}$ se possa estimar (apenas através de (N_c, N_p)) independentemente de $\underline{\lambda}$ sob M_1 , a distribuição condicional (em \bar{m}_0) dos respectivos estimadores e estatísticas adequadas para testar alguma hipótese sobre $\underline{\theta}$ permanece dependente de $\underline{\lambda}$. O mesmo ocorre com a respectiva distribuição não condicional calculada de (3.13). Sob M_1 , os procedimentos sobre $\underline{\theta}$ via teoria da amostragem devem, pois, ser baseados na distribuição não condicional dos dados e acompanhados de procedimentos correspondentes sobre $\underline{\lambda}$ (vide, e.g., a análise para tabelas de contingência bidimensionais de Chen and Fienberg [1974]).

A eliminação dos parâmetros perturbadores do ponto de vista das inferências clássicas sobre $\underline{\theta}$ exige assim a adopção de hipóteses mais restritivas sobre esses parâmetros, com as suas implicações na compatibilidade das configurações observadas.

Na situação de censura por uma partição que temos vindo a considerar, a suposição adicional de que as probabilidades de classificação completa não dependam da categoria (i.e., $\lambda_i^i = \lambda_1^*$, $i=1, \dots, s$), no contexto de um modelo MAR descrito por M_1 , implica que $\lambda_\delta^\delta = 1 - \lambda_1^* - \lambda_0^* \equiv \lambda_2^*$, $\forall \delta \in T$, onde $\lambda_0^* = \lambda_{\delta_0}^\delta$. Ou seja, todas as probabilidades condicionais de classificação completa, de classificação parcial e de não classificação são constantes.

Sob este modelo para o mecanismo de censura, denotado por M_2 , a distribuição do padrão de dados observável, \bar{m} , condicional em \underline{W} depende apenas dos valores da estatística (N_c, N_p, n_o) - reveja-se (3.6) - e por conseguinte, coincide com a probabilidade marginal de \bar{m} . Esta propriedade vai implicar que o núcleo de (3.16) coincida com (3.9) pelo que a análise condicional em (N_c, N_p, n_o) corresponde a ignorar o mecanismo de censura. Note-se ainda que, independentemente do padrão dos dados incompletos, o confinamento aos dados completamente categorizados para propósitos inferenciais implica uma verosimilhança multinomial de parâmetro $\underline{\theta}$ (núcleo da distribuição de N_c dado N_c) se e só se as probabilidades de registo completo não dependem da categoria. Esta observação é tanto mais pertinente quanto é facto que grande parte da literatura parece escamoteá-la quando despreza os dados incompletos.

O facto de (3.16) não depender de $\underline{\lambda}$ sob M_2 significa que (N_c, N_p, n_o) é uma estatística suficiente específica para $\underline{\lambda}$, para empregar a terminologia de Basu (1977). Por outro lado, pode constatar-se ainda que (3.14) se torna independente de $\underline{\theta}$ sob M_2 , o que traduz que a referida estatística é também ancilária parcial para $\underline{\theta}$. Numa palavra, sob o modelo M_2 a estatística (N_c, N_p, n_o) é S-ancilária para $\underline{\theta}$. Em consequência disto, o princípio da condicionalidade generalizado justifica o desprezo do mecanismo de censura, quando modelado por M_2 , nas inferências clássicas sobre $\underline{\theta}$, as quais revestem, pois, uma natureza condicional.

Do ponto de vista de qualquer metodologia que use instrumentos da inferência clássica, tudo se passa como se a amostra fosse constituída por $N-n_0$ elementos, estratificados em duas amostras independentes, de dimensões N_c e N_p , fixadas previamente, extraídas de populações multinomiais relacionadas. Fica assim evidenciado que a prática de encarar um mecanismo aleatório de censura como se fosse determinístico, com o padrão de dados incompletos fixado por delineamento (e portanto de definir a verosimilhança por meio da expressão (3.9), só se afigura correcta para processos particulares e restritivos de censura.

A situação considerada de T constituir uma partição do conjunto de categorias é um caso particular da situação focada na secção seguinte e que será analisada com maior detalhe, dado que é sobre ela que se aplicam os métodos clássicos de inferência descritos no Capítulo 4.

3.3. Caso de um padrão de censura em partições genéricas

Em várias situações, os elementos de Π_{UT} estão estruturados em T conjuntos de categorias, Π_t , $t=1, \dots, T$, cada um dos quais constitui uma partição do conjunto das s categorias. É o que acontece, por exemplo, numa tabela de contingência multidimensional onde para cada unidade populacional

- ou são registadas as categorias de todas as variáveis;
- ou não são registadas as categorias de uma ou mais variáveis.

Os dados podem então ser resumidos numa tabela completamente categorizada e em várias tabelas marginais suplementares.

Não é essencial que o padrão observável de dados implique uma tabela completa. Em vez dos conjuntos singulares de categorias estarem organizados na partição mais fina Π , eles podem estar dispersos unicamente por vários conjuntos Π_t distintos de Π - é o caso em que os dados se podem organizar em diversas tabelas com categorias combinadas. As considerações subsequentes sobre vários mecanismos geradores da perda de dados mantêm-se na sua essência para esta situação.

Admitiremos, contudo, pela sua ocorrência frequente, que as observações possíveis podem ser expostas numa tabela com as categorias de $\Pi_1 \equiv \Pi$ (relativas aos elementos completamente categorizados) e em T tabelas adicionais envolvendo os elementos parcial e totalmente censurados. A tabela suplementar t , $t=2, \dots, T+1$ apresenta as categorias reunidas em $\Pi_t = \{\delta_j^t, j=1, \dots, s_t\}$, em número de s_t , considerando-se $\Pi_{T+1} = \{\delta_0\}$ ($s_{T+1}=1$). Por conveniência, o número de categorias de Π_1 será denotado a partir de agora, neste capítulo, por s_1 , passando s a indicar o número total de celas nas $T+1$ tabelas.

O processo de registo será definido, de uma forma condensada, em termos do vector $\bar{m} = (\bar{m}_k, k=1, \dots, N)'$, com $\bar{m}_k = (m_k^1, \dots, m_k^T, m_k^{T+1})'$ apresentando como valores possíveis os vectores \underline{e}_t da base canónica de \mathbb{R}^{T+1} . Sublinhe-se que, em virtude da definição de Δ , cada unidade k só tem $T+1$ possibilidades de classificação e afirmar que, dado $\underline{w}_k = \underline{e}_i$, $i=1, \dots, s_1$, $m_k^t = 1$ é equivalente a dizer que k é classificado no

único subconjunto δ_j^t da partição Π_t que contém i .

Desta forma, $\Delta_i = T+1, \forall i$, e portanto, é de $s_1(T+1)$ o nº total de probabilidades condicionais de registo, agora denotadas por $\lambda_i^t = P(\bar{m}_k = \varepsilon_t | W_k = e_i)$, $t=1, \dots, T+1$, $i=1, \dots, s_1$. Fazemos

$$\underline{\lambda}^t = (\lambda_i^t, i=1, \dots, s_1), t=1, \dots, T+1; \quad \underline{\beta}_i = (\lambda_i^t, t=1, \dots, T+1), i=1, \dots, s_1$$

onde, em termos da notação da secção 3.2, $\lambda^1 \equiv \lambda_c$, $\lambda^{T+1} \equiv \lambda_o$ e $(\lambda^{2'}, \dots, \lambda^{T'})' = \lambda_p$.

O vector coluna de todos os $\{\lambda_i^t\}$ pode assim ser representado por $\underline{\lambda} = (\lambda_i^t, t=1, \dots, T+1)'$ ou por $\underline{\beta} = (\beta_i^t, i=1, \dots, s_1)'$. As restrições naturais envolvendo estes parâmetros são assim dadas por

$$\mathbf{1}_{T+1}' \underline{\beta}_i = 1, i=1, \dots, s_1 \quad (3.17)$$

A distribuição correspondente a (3.1) converte-se em

$$f(W, \bar{m} | \theta, \underline{\lambda}) = \prod_{i=1}^{s_1} \prod_{t=1}^{T+1} (\theta_i \lambda_i^t)^{y_{it}} \quad (3.18)$$

onde $y_{it} = \sum_{k: \bar{m}_k = \varepsilon_t} W_{ki} \equiv \sum_{k: W_k = e_i} m_k^t$ traduz o número (não observado para $t=2, \dots, T+1$) de indivíduos da amostra que pertencem à categoria i e que foram classificados na cela da tabela t compatível com tal categoria.

Fazendo $\underline{y}_t = (y_{it}, i=1, \dots, s_1)'$, $t=1, \dots, T+1$, a expressão anterior revela que $\underline{y} = (y_1' y_2' \dots y_{T+1}')'$ tem, dado $(\theta, \underline{\lambda})$ uma distribuição multinomial de dimensão $s_1(T+1)-1$.

Seja $Z_t = (z_{t1} \dots z_{ts_t})$, $t=2, \dots, T+1$, a matriz particionada $s_1 \times s_t$ cujos vectores coluna z_{tj} têm coordenadas iguais a 1 nas posições $i \in \delta_j^t$, sendo as restantes nulas. Note-se, desde já, que Z_t é indicadora da partição Π_t uma vez que os seus elementos são 0 e 1, as suas colunas são ortogonais e nenhuma linha é nula. As frequências observáveis podem definir-se então como

$$\underline{N}_1 \equiv (n_{1i}, i=1, \dots, s_1)' = \underline{y}_1 \quad (3.19)$$

$$\underline{N}_t \equiv (n_{tj}, j=1, \dots, s_t)' = Z_t' \underline{y}_t, t = 2, \dots, T+1$$

e portanto, $\underline{n} = (N_t', t=1, \dots, T+1)'$ apresenta, dado $(\theta, \underline{\lambda})$, a distribuição multinomial $(s-1)$ -dimensional

$$f(\underline{n} | \theta, \underline{\lambda}) = \frac{N!}{\prod_{t=1}^{T+1} \prod_{j=1}^{s_t} n_{tj}!} \left\{ \prod_{i=1}^{s_1} (\theta_i \lambda_i^1)^{n_{1i}} \prod_{t=2}^{T+1} \prod_{j=1}^{s_t} \left(\sum_{i \in \delta_j^t} \theta_i \lambda_i^t \right)^{n_{tj}} \right\} \quad (3.20)$$

a qual patenteia a não identificabilidade dos $s_1(T+1)-1$ parâmetros $(\underline{\theta}, \underline{\lambda})$ linearmente independentes. De um ponto de vista não Bayesiano, as inferências sobre $\underline{\theta}$ para um mecanismo geral de registo exigem o conhecimento de $\underline{\lambda}$ mediante, por exemplo, a sua estimação por informação \tilde{a} priori.

Definamos os vectores de \mathbb{R}^{s_t} , $\underline{W}_k^t = (W_{kj}^t, j=1, \dots, s_t)$, $t = 2, \dots, T$ onde

$$W_{kj}^t = \begin{cases} 1 & \text{se } W_k \in \bigcup_{i \in \delta_j^t} \{e_i\} \\ 0 & \text{c.c.} \end{cases}$$

A fixação de Δ e de \underline{m}_0 permite definir as observações não censuradas totalmente, \underline{W}_0 , pelos valores de \underline{W}_k para k tal que $\underline{m}_k = \underline{\varepsilon}_1$ e de \underline{W}_k^t para k tal que $\underline{m}_k = \underline{\varepsilon}_t$, $t=2, \dots, T$. A distribuição marginal de \underline{W}_0 calculada sem condicionamento no padrão observado \underline{m}_0 é então dada por

$$f(\underline{W}_0 | \underline{\theta}) = \prod_{i=1}^{s_1} \theta_i^{n_{1i}} \times \prod_{t=2}^T \prod_{j=1}^{s_t} (z_{tj}' \cdot \underline{\theta})^{n_{tj}} \quad (3.21)$$

Esta função pode ser interpretada dizendo que \underline{n} condicionalmente a $\{N_{t.} = \sum_{s=1}^t N_{st}, t = 1, \dots, T+1\}$ tem uma distribuição multinomial - produto. Vista como uma função de $\underline{\theta}$ dado \underline{n} , ela representará, portanto, a verosimilhança correcta se $N_{T+1} = 0$ e o mecanismo de registo for determinístico. Isto é, se o padrão de dados surge intencionalmente com os $\{N_{t.}, t=1, \dots, T\}$ previamente fixados por delineamento e representando a dimensão de amostras independentes de populações multinomiais relacionadas. Este processo de registo será denotado por M_0 .

A comparação de (3.21) com o núcleo $L(\underline{\theta}, \underline{\lambda} | \underline{n})$ de (3.20) mostra que as inferências sobre $\underline{\theta}$ baseadas nas duas verosimilhanças não são em geral equivalentes, o que traduz a não ignorabilidade do mecanismo geral de registo para esse tipo de abordagem inferencial.

Suponhamos agora o seguinte modelo MAR para o mecanismo de registo

$$M_1 : \lambda_i^t = \lambda_j^t, \forall i \in \delta_j^t, j = 1, \dots, s_t; t=1, \dots, T+1 \quad (3.22)$$

Do facto de $\underline{m}_k | \underline{W}_k = e_i; \beta_i \overset{\sim}{\text{iid}} M_{T+1}(1, \beta_i)$ segue-se que a distribuição condicional de \underline{m} sob M_1

$$f_1(\underline{m} | \underline{W}; \underline{\lambda}) = \prod_{i=1}^{s_1} (\lambda_i^1)^{n_{1i}} \times \prod_{t=2}^{T+1} \prod_{j=1}^{s_t} (\lambda_j^t)^{n_{tj}} \quad (3.23)$$

depende, dado $\underline{\lambda}$, somente dos valores observados $\{N_{t.}\}$. Seja $\underline{\lambda}_{M_1}$ o vector $\underline{\lambda}$ sob

(3.22) rearranjado como $(\underline{\lambda}_{M_1}^t, t=1, \dots, T+1)$ onde $\underline{\lambda}_{M_1}^t = (\lambda_1^t \dots \lambda_{s_t}^t)$, $t=1, \dots, T+1$.

Seja ainda $\underline{z} = (z_2 \dots z_t \dots z_{T+1})$, $\bar{x} = (I_{s_1} \underline{z})'$ e $\underline{\psi}(\underline{\theta}) = \bar{x}\underline{\theta}$. A expressão (3.20) mostra então que a distribuição por amostragem não condicional (em \bar{m}_0) de \underline{n} é tal que $\underline{n} | \underline{\theta}, \underline{\lambda}_{M_1} \rightsquigarrow M_s(N, D_{\underline{\psi}} \underline{\lambda}_{M_1})$ onde $D_{\underline{\psi}}$ é a matriz diagonal com os s elementos do vector $\underline{\psi} = \underline{\psi}(\underline{\theta})$ na diagonal principal, isto é

$$f_1(\underline{n} | \underline{\theta}, \underline{\lambda}_{M_1}) = \frac{N!}{\prod_{t=1}^{T+1} \prod_{j=1}^{s_t} n_{tj}!} \left\{ \prod_{i=1}^{s_1} (\theta_i \lambda_i^1)^{n_{1i}} \times \prod_{t=2}^{T+1} \prod_{j=1}^{s_t} (z'_{tj} \cdot \theta) \lambda_j^t \right\}^{n_{tj}} \quad (3.24)$$

A respectiva função de verosimilhança, digamos $L_1(\underline{\theta}, \underline{\lambda}_{M_1} | \underline{n})$, factoriza-se em

$$L_1(\underline{\theta}, \underline{\lambda}_{M_1} | \underline{n}) = L_1(\underline{\lambda}_{M_1} | \underline{n}) \times L_1(\underline{\theta} | \{N_t, t=1, \dots, T\}) \\ = \left\{ \prod_{t=1}^{T+1} \prod_{j=1}^{s_t} (\lambda_j^t)^{n_{tj}} \right\} \left\{ \prod_{i=1}^{s_1} \theta_i^{n_{1i}} \prod_{t=2}^T \prod_{j=1}^{s_t} (z'_{tj} \theta)^{n_{tj}} \right\} \quad (3.25)$$

onde $L_1(\underline{\theta} | \{N_t, t=1, \dots, T\})$ coincide com a verosimilhança quando se ignora o processo de registo (dada em (3.21)). Como $\underline{\theta}$ e $\underline{\lambda}$ são considerados parâmetros de variação independente, o resultado (3.25) indica que o mecanismo de registo é ignorável para as inferências sobre $\underline{\theta}$ baseadas na verosimilhança, desde que se admita adicionalmente, para a abordagem Bayesiana, que $\underline{\theta}$ e $\underline{\lambda}_{M_1}$ são distribuídos independentemente à priori.

No contexto de tabelas de contingência com observações perdidas em algumas variáveis, a estimação de máxima verosimilhança de $\underline{\theta}$ pode proceder-se através do método de factorização da verosimilhança de Rubin (1974). No caso de não ser observável uma tabela completa, a expressão (3.24) pode ainda reflectir a falta de identificabilidade de $\underline{\theta}$. A sua estimação tornar-se-á possível se cada e_i for um vector coluna de pelo menos uma matriz Z_t , o que assegurará a identificabilidade de cada θ_i .

A função de probabilidade (3.24) admite ainda a factorização, na qual se considera $Z_1 = I_{s_1}$,

$$f_1(\underline{n} | \underline{\theta}, \underline{\lambda}_{M_1}) = \frac{N!}{\prod_{t=1}^{T+1} N_t!} \prod_{t=1}^{T+1} [(\underline{\theta}' Z_t) \underline{\lambda}_{M_1}^t]^{N_t} \times \prod_{t=1}^T \left\{ \frac{N_t!}{\prod_{j=1}^{s_t} n_{tj}!} \prod_{j=1}^{s_t} \left(\frac{(z'_{tj} \cdot \theta) \lambda_j^t}{(\underline{\theta}' Z_t) \underline{\lambda}_{M_1}^t} \right)^{n_{tj}} \right\} \quad (3.26)$$

onde o primeiro factor corresponde à distribuição multinomial para o vector $(N_1, \dots, N_T, N_{T+1})$ dado N e o segundo à distribuição condicional (multinomial-produto) de (N_1, \dots, N_t) dado (N_1, \dots, N_T) .

O núcleo desta distribuição condicional é ainda, sem hipóteses adicionais

sobre $\lambda_{\sim M_1}$, diferente de (3.21), o que implica que a distribuição por amostragem condicional (em \bar{m}_0) de uma estatística é geralmente diferente da sua distribuição marginal, calculada de $f(W_0 | \theta)$.

Suponhamos agora o seguinte modelo MAR mais restritivo

$$M_2: \lambda_i^t = \lambda_t^t, \quad i=1, \dots, s_1; \quad t = 1, \dots, T+1 \quad (3.27)$$

O vector dos parâmetros perturbadores sob (3.27) é expresso por $\lambda_{\sim M_2} = (\lambda_{\sim M_2}^t, t=1, \dots, T+1)'$, com $\lambda_{\sim M_2}^t = \lambda_{t \sim s_1}^t$. De (3.23) resulta que a função de probabilidade condicional de \bar{m} dado \underline{W} para este modelo é dada por

$$f_2(\bar{m} | \underline{W}; \lambda) = \prod_{t=1}^{T+1} (\lambda_t^t)^{N_t}. \quad (3.28)$$

ilustrando, pelo facto de $N_{t.} = \sum_{k=1}^N m_k^t$, $t=1, \dots, T+1$, que ela não depende de \underline{W} , coincidindo, pois, com a respectiva probabilidade marginal.

Este resultado vai implicar que o núcleo da distribuição de \underline{n} dado $\{N_{t.}\}$ coincida com (3.21) - atente-se no 2º factor de (3.26) e no facto de $\lambda_{\sim M_1}^t = \lambda_{t \sim s_t}^t$ sob M_2 e de $(Z_{t.}' \theta)' 1_{\sim s_t} = 1$ pelas características de Z_t , $t=1, \dots, T$.

Deste modo, o mecanismo de registo M_2 é ignorável para as inferências sobre θ baseadas na teoria da amostragem. Na sequência do que foi dito em 3.2, a mesma conclusão pode ser extraída se o vector \bar{m} é encarado como reflectindo os diversos padrões observáveis, no quadro mais lato de Δ englobar, para além de $\bigcup_{t=1}^{T+1} \prod_t$, outros conjuntos com frequências observadas nulas.

Sob M_2 temos então

$$f_2(\{N_{t.}, t=1, \dots, T+1\} | N; \{\lambda_t^t\}) = \frac{N!}{\prod_{t=1}^{T+1} N_{t.}!} \prod_{t=1}^{T+1} (\lambda_t^t)^{N_t}. \quad (3.29)$$

e

$$f_2(N_1, \dots, N_T | N_1, \dots, N_T; \theta) = \prod_{t=1}^T f_2(N_{t.} | N_{t.}; \theta) = \prod_{t=1}^T \left\{ \frac{N_{t.}!}{\prod_{j=1}^{s_t} n_{tj}!} \prod_{j=1}^{s_t} (z_{tj}' \cdot \theta)^{n_{tj}} \right\} \quad (3.30)$$

Assim, quando o processo de registo estipula que cada indivíduo só pode ser classificado, quando o é efectivamente, num conjunto de uma partição em T partições admissíveis, a estatística $(N_1, \dots, N_T, N_{T+1})$ é S -ancilária para θ se o mecanismo de registo for modelado por M_2 . Por conseguinte, a argumentação em torno do princípio da condicionalidade justifica que os dados relevantes possam ser reinterpretados como tendo surgido da experiência condicional ao valor observado de (N_1, \dots, N_T) , ou seja, gerados pelo mecanismo M_0 .

Denotemos agora o vector das probabilidades amostrais de δ_j^t , $j=1, \dots, s_t$ por $\phi_t(\underline{\theta})$, i.e.

$$\phi_t(\underline{\theta}) = (\phi_{tj}(\underline{\theta}), j=1, \dots, s_t)' = \underline{z}'_t \underline{\theta}, \quad t=2, \dots, T \quad (3.31)$$

Seja ainda $\underline{\theta}_{(tj)} = (\theta_i, i \in \delta_j^t)'$ com dimensão d_j^t (cardinal de δ_j^t) e $\rho_{(tj)}(\underline{\theta})$ o vector das probabilidades amostrais de cada uma das d_j^t categorias condicionais a δ_j^t , isto é

$$\rho_{(tj)}(\underline{\theta}) = (\rho_{(tj)}^i(\underline{\theta}), i \in \delta_j^t)' = \frac{\underline{\theta}_{(tj)}}{\phi_{tj}(\underline{\theta})}, \quad t=2, \dots, T \quad (3.32)$$

Recorde-se (vide (3.10)) que, sob qualquer modelo MAR M_1 , $\rho_{(tj)}(\underline{\theta})$ define as probabilidades de qualquer indivíduo pertencer a cada categoria $i \in \delta_j^t$ dado que foi classificado em δ_j^t .

Consideremos ainda $\phi_t(N_1) = \underline{z}'_t N_1$ e $\rho_{(tj)}(N_1) = \frac{N_1(tj)}{\phi_{tj}(N_1)}$, com

$N_1(tj) = (n_{1i}, i \in \delta_j^t)$, as mesmas funções de (3.31) e (3.32) mas com $\underline{\theta}$ substituído por N_1 .

O resultado distribucional expresso em (3.30) permite concluir então que para $t=2, \dots, T$

$$\phi_t(N_1) | N_1; \theta \rightsquigarrow M_{s_t}(N_1, \phi_t(\underline{\theta})) \quad (3.33a)$$

$$N_t + \phi_t(N_1) | N_t, N_1; \underline{\theta} \rightsquigarrow M_{s_t}(N_t + N_1, \phi_t(\underline{\theta})) \quad (3.33b)$$

$$N_1 | \phi_t(N_1); \underline{\theta} \rightsquigarrow \prod_{j=1}^{s_t} M_{d_j^t}(\phi_{tj}(N_1), \rho_{(tj)}(\underline{\theta})) \quad (3.33c)$$

onde o factor j da última relação distribucional se reporta à distribuição de $N_1(tj)$ dado $\phi_{tj}(N_1) \equiv \underline{z}'_{tj} N_1$.

Estes resultados, válidos sob M_2 e sob M_0 , permitem interpretar, em termos de distribuições multinomiais marginais e condicionais (em ambos os casos com $\{N_t\}$ considerados fixos), a função de verosimilhança relevante sob um mecanismo de registo M_1 para certos padrões especiais de censura, como veremos na secção seguinte.

3.4. Caso de um padrão de censura em partições encaixadas

No caso tratado anteriormente, a condição de Π_t representar uma partição do espaço amostral significa que Π_t , $t=2, \dots, T+1$ está directamente encaixada na partição mais fina $\Pi_1 = \{\{i\}, i=1, \dots, s_1\}$ no sentido em que cada parte de Π_t é uma união de partes de Π_1 . Deste modo, o caso de duas partições é um exemplo especial de um padrão de censura em partições encaixadas (em um nível).

Quando $T > 2$, pode acontecer que entre cada par de sucessivas partições

Π_t exista uma relação do tipo daquela entre Π_1 e cada Π_t . Isto é, o conjunto de T partições pode ser ordenado ascendentemente de modo que Π_{t-1} seja mais fina do que Π_t , $t > 2$. A matriz $\underline{Z} = (Z_t, t \geq 2)$ indicadora das partições Π_t , $t \geq 2$ é assim tal que cada coluna de \underline{Z}_t é uma soma de colunas de \underline{Z}_{t-1} pelo que $Z_t = Z_{t-1} L_t$, onde L_t é a matriz $s_{t-1} \times s_t$, do tipo das matrizes $\{Z_t\}$, indicadora da característica de cada parte de Π_t ser uma união de partes de Π_{t-1} . Esta estrutura hierárquica das partições Π_t , $t=1, \dots, T$, que denominamos de partições encaixadas (em $T-1$ níveis), pode ser representada por uma árvore, cujo tronco é a partição mais grossa Π_T , necessariamente $\{\delta_o\}$ quando o não registo é admissível. Uma ilustração do encaixe em 2 níveis é dada pelo caso em que $s_1 = 12$ com $\Pi_2 = \{\{1,2,3\}, \{4,5,6\}, \{7,8,9\}, \{10,11,12\}\}$ e $\Pi_3 = \{\{1,2,3,4,5,6\}, \{7,8,9,10,11,12\}\}$, no qual $L_2 = Z_2$ e $L_3 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}'$.

No contexto de tabelas de contingência com os $\{\Pi_t\}$ correspondentes às partições definidas por tabelas marginais, este padrão em partições encaixadas coincide com o conceito de padrão de dados monótono de Rubin (1974), segundo o qual as unidades k e as variáveis categorizadas X_j podem ser arranjadas de modo que para cada unidade k , X_{kj} observado implica $X_{kj'}$ observado para todo o $j' < j$. Será o caso do exemplo ilustrativo anterior se ele traduzir as $2 \times 2 \times 3$ categorias ordenadas lexicograficamente de uma tabela tridimensional (X_1, X_2, X_3) onde, por conseguinte, Π_2 corresponderá à marginal (X_1, X_2) e Π_3 à marginal X_1 . Fica também claro que a ocorrência de duas ou mais marginais suplementares com a mesma dimensão não é compatível com o padrão de censura encaixada.

A justificativa para a consideração em separado deste padrão de censura reside na particularidade de permitir simplificar os procedimentos inferenciais sobre θ sob qualquer modelo M_1 para o mecanismo de censura. Para constatar tal facto, restringamo-nos por agora ao método da máxima verosimilhança e consideremos em primeiro lugar o caso de censura por uma partição ($T=2$) diferente de $\{\delta_o\}$ - note-se que esta é irrelevante sob M_1 .

Recordando (3.25) e tendo presente a notação introduzida na secção anterior, é fácil verificar que $L_1(\theta | n)$ admite a factorização

$$\begin{aligned} L_1(\theta | n) &= \prod_{j=1}^{s_2} \prod_{i \in \delta_j^2} \theta_i^{n_{1i}} \times \prod_{j=1}^{s_2} \left(\phi_{2j}(\theta) \right)^{n_{2j}} \\ &= \prod_{j=1}^{s_2} \left(\phi_{2j}(\theta) \right)^{n_{2j} + \phi_{2j}(N_1)} \times \prod_{j=1}^{s_2} \prod_{i \in \delta_j^2} \left(\rho_{(2j)}^i(\theta) \right)^{n_{1i}} \quad (3.34) \end{aligned}$$

Dada a definição de $\phi_{2j}(\theta)$ e de $\rho_{(2j)}^i(\theta)$, esta expressão permite calcular explicitamente o estimador de máxima verosimilhança de θ , $\hat{\theta}$, através da relação

$$\hat{\theta}_i = \hat{\phi}_{2j}(\theta) \hat{\rho}_{(2j)}^i(\theta), \quad i \in \delta_j^2, \quad j=1, \dots, s_2 \quad (3.35a)$$

$$\text{onde } \hat{\phi}_{2j}(\underline{\theta}) = \frac{N_{2j} + \phi_{2j}(N_1)}{N_{2j} + 1'_{s_2} \phi_{2j}(N_1)} \quad (3.35b)$$

$$\hat{\rho}_{(2j)}(\underline{\theta}) = N_{1(2j)} / \phi_{2j}(N_1) \equiv \rho_{(2j)}(N_1), \quad j=1, \dots, s_2 \quad (3.35c)$$

Pelos resultados (3.33), a factorização (3.34) revela que $L_1(\underline{\theta} | \underline{n})$ é o produto dos núcleos das distribuições, sob M_2 , de $N_{2j} + \phi_{2j}(N_1)$ dado $\{N_{2j}, N_1\}$ e de $N_{1(2j)}$ dado $\phi_{2j}(N_1)$, $j=1, \dots, s_2$ (condicionais a $\underline{\theta}$). Deste modo, o estimador de máxima verosimilhança de $\underline{\theta}$ sob M_1 coincide com o respectivo estimador do método dos momentos para o modelo probabilístico (3.33 b,c), válido sob M_2 (note-se que $1'_{s_2} \phi_{2j}(N_1) = N_{1(2j)}$).

A determinação em forma fechada do estimador de máxima verosimilhança de $\underline{\theta}$ sob M_1 ocorre igualmente no padrão geral de censura em partições encaixadas devido a uma factorização da verosimilhança do tipo da indicada em (3.34). Para o ilustrarmos consideremos o caso do encaixe em 2 níveis ($T = 3$).

Atendendo a que cada parte de $\Pi_3 = \{\delta_h^3, h=1, \dots, s_3\}$ é uma união de partes de $\Pi_2 = \{\delta_\ell^2, \ell=1, \dots, s_2\}$, passaremos a denotar Π_2 por $\{\delta_{hm}^2, m=1, \dots, j_h; h=1, \dots, s_3\}$, com $\sum_{h=1}^{s_3} j_h = s_2$, de modo a explicitar que $\{\delta_{hm}^2, m=1, \dots, j_h\}$ é a classe dos δ_ℓ^2 directamente encaixada em δ_h^3 , para todo o $h = 1, \dots, s_3$. Em sintonia, representaremos $Z_2 = (Z_{2h}, h=1, \dots, s_3)$, onde a sub-matriz $Z_{2h} = (z_{2hm}, m=1, \dots, j_h)$ indica a partição do conjunto δ_h^3 representado pelo vector coluna z_{3h} de Z_3 . Consequentemente, $\phi_2(\underline{\theta}) \equiv Z_2' \cdot \underline{\theta} = (\phi_{2h}'(\underline{\theta}), h=1, \dots, s_3)'$ onde $\phi_{2h}'(\underline{\theta}) = Z_{2h}' \cdot \underline{\theta}$ com componentes $\phi_{2hm}'(\underline{\theta}) = z_{2hm}' \cdot \underline{\theta}$, $m=1, \dots, j_h$. Como os vectores componentes de Z_{2h} têm as coordenadas referentes a δ_h^3 nulas, os sub-vectores de $\phi_2(\underline{\theta})$ podem ser alternativamente definidos por $\phi_{2h}'(\underline{\theta}) = \bar{Z}_{2h}' \cdot \underline{\theta}_{(3h)}$, onde \bar{Z}_{2h} é a matriz $d_h^3 \times j_h$ (d_h^3 representa o cardinal de δ_h^3) obtida de Z_{2h} por omissão daquelas linhas nulas, e $\underline{\theta}_{(3h)} = (\theta_i, i \in \delta_h^3)'$.

Seja ainda

$$\Gamma_{(2h)}(\underline{\theta}) = (\Gamma_{(2h)}^m, m=1, \dots, j_h)' = \frac{\phi_{2h}'(\underline{\theta})}{\phi_{3h}'(\underline{\theta})} = \bar{Z}_{2h}' \cdot \underline{\theta}_{(3h)} / z_{3h}' \cdot \underline{\theta}$$

o vector das probabilidades amostrais de qualquer indivíduo pertencer a cada conjunto δ_{hm}^2 dado que pertence a δ_h^3 e

$$\rho_{(2hm)}(\underline{\theta}) = (\rho_{(2hm)}^i(\underline{\theta}), i \in \delta_{hm}^2)' = \underline{\theta}_{(2hm)} / \phi_{2hm}'(\underline{\theta}) = \frac{1}{\Gamma_{(2h)}^m} \underline{\theta}_{(2hm)} / \phi_{3h}'(\underline{\theta})$$

onde $\underline{\theta}_{(2hm)}$ é o vector $d_{hm}^2 \times 1$ dos elementos θ_i , $i \in \delta_{hm}^2$, o vector das probabilidades amostrais de cada indivíduo pertencer à categoria i dado que pertence a δ_{hm}^2 .

Note-se que o facto de $Z_3 = Z_2 L_3$, ou seja, de $z_{3h} = Z_2 L_{3h} = \sum_{m=1}^{j_h} z_{2hm}$, para $h = 1, \dots, s_3$, onde L_{3h} é o vector coluna h de L_3 , vai implicar que

$\phi_{3h}(\theta) = L'_{3h} \phi_2(\theta) = \sum_{m=1}^{j_h} \phi_{2hm}(\theta)$, e assim $L'_{3h} \phi_2(N_1) = \sum_{m=1}^{j_h} \sum_{i \in \delta_{hm}^2} n_{1i}$. Em consonância com a partição de Z_2 , seja $N_2 = (n_{2\ell}, \ell=1, \dots, s_2)'$ particionado nos sub-vectores $N_{2(h)} = (n_{2hm}, m=1, \dots, j_h)'$, pelo que $L'_{3h} N_2 = L'_{j_h} N_{2(h)}$.

Tendo presente estes aspectos notacionais e observações, podemos concluir sucessivamente de (3.25) que

$$\begin{aligned}
 L_1(\theta | n) &= \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \prod_{i \in \delta_{hm}^2} \theta_i^{n_{1i}} \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} (z'_{2hm} \cdot \theta)^{n_{2hm}} \times \prod_{h=1}^{s_3} (z'_{3h} \cdot \theta)^{n_{3h}} \\
 &= \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \prod_{i \in \delta_{hm}^2} (\theta_i / z'_{2hm} \cdot \theta)^{n_{1i}} \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} (z'_{2hm} \cdot \theta)^{n_{2hm} + \phi_{2hm}(N_1)} \times \prod_{h=1}^{s_3} (z'_{3h} \cdot \theta)^{n_{3h}} \\
 &= \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \prod_{i \in \delta_{hm}^2} \left(\rho^i(2hm)(\theta) \right)^{n_{1i}} \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \left(\frac{z'_{2hm} \cdot \theta}{z'_{3h} \cdot \theta} \right)^{n_{2hm} + \phi_{2hm}(N_1)} \\
 &\quad \times \prod_{h=1}^{s_3} (z'_{3h} \cdot \theta)^{n_{3h} + L'_{3h} [N_2 + \phi_2(N_1)]} \\
 &= \prod_{h=1}^{s_3} \left(\phi_{3h}(\theta) \right)^{n_{3h} + L'_{3h} [N_2 + \phi_2(N_1)]} \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \left(\Gamma^m(2h)(\theta) \right)^{n_{2hm} + \phi_{2hm}(N_1)} \\
 &\quad \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \prod_{i \in \delta_{hm}^2} \left(\rho^i(2hm)(\theta) \right)^{n_{1i}}
 \end{aligned} \tag{3.36}$$

A expressão (3.36) revela que a verosimilhança envolve primeiramente todos os dados relacionados com as categorias de $\Pi_3 (N_3 + L'_{3h} N_2 + \phi_3(N_1))$ depois a partição de uma parte deles ($L'_{3h} [N_2 + \phi_2(N_1)]$) nos dados relacionados com as categorias de Π_2 , e finalmente, a partição de uma parte destes ($\phi_2(N_1)$) nos dados completamente categorizados.

Observe-se agora que, sob a validade de M_2 (ou de M_0), podemos asseverar, por (3.30), que

$$N_{2+\phi_2}(N_1) | N_{2.}, N_{1.}; \theta \rightsquigarrow M_{s_2}(N_{2.} + N_{1.}, \phi_2(\theta)) \tag{3.37a}$$

$$N_{3+L'_{3h}[N_2+\phi_2(N_1)]} | \{N_{t.}\}; \theta \rightsquigarrow M_{s_3} \left(\sum_{t=1}^3 N_{t.}, \phi_3(\theta) \right) \tag{3.37b}$$

$$N_{2(h)+\phi_{2h}(N_1)} | L'_{3h}[N_2+\phi_2(N_1)]; \theta \rightsquigarrow M_{j_h} (L'_{3h}[N_2+\phi_2(N_1)], \Gamma(2h)(\theta)), \tag{3.37c}$$

independentemente para $h=1, \dots, s_3$

$$N_{1(hm)} | \phi_{2hm}(N_1); \theta \rightsquigarrow M_{d_{hm}^2}(\phi_{2hm}(N_1), \rho_{(2hm)}(\theta)) \quad (3.37d)$$

independentemente para $m=1, \dots, j_h$ e $h=1, \dots, s_3$

onde $N_{1(hm)} = (n_{1i}, i\epsilon\delta_{hm}^2)'$.

Deste modo, $L_1(\theta | \underline{n})$ é o produto dos núcleos das distribuições multinomiais sob M_2 (3.37b,c,d). A expressão em forma fechada de $\hat{\theta}$ é então

$$\hat{\theta}_i = \hat{\phi}_{3h}(\theta) \cdot \hat{\Gamma}_{(2h)}^m(\theta) \cdot \hat{\rho}_{(2hm)}^i(\theta), \quad i\epsilon\delta_{hm}^2; \quad m=1, \dots, j_h; \quad h=1, \dots, s_3 \quad (3.38a)$$

onde

$$\hat{\phi}_{3h}(\theta) = \frac{N_3 + L_3' [N_2 + \phi_2(N_1)]}{N_3 + N_2 + N_1} \quad (3.38b)$$

$$\hat{\Gamma}_{(2h)}(\theta) = \frac{N_2(h) + \phi_{2h}(N_1)}{L_3' [N_2 + \phi_2(N_1)]}, \quad h=1, \dots, s_3 \quad (3.38c)$$

$$\hat{\rho}_{(2hm)}(\theta) = \frac{N_{1(hm)}}{\phi_{2hm}(N_1)}, \quad m=1, \dots, j_h; \quad h=1, \dots, s_3 \quad (3.38d)$$

Note-se que

$$\hat{\rho}_{(2hm)}(\theta) = \rho_{(2hm)}(N_1)$$

e usando a reparametrização $\hat{\Gamma}_{(2h)}(\theta) = \eta(\Psi_{2h})$, onde $\Psi_{2h} = (\phi_{2hm}, m=1, \dots, j_h)'$ = $\phi_{2h}(\theta)$,

$$\hat{\Gamma}_{(2h)}(\theta) \equiv \eta(\hat{\Psi}_{2h}) = \eta[N_2(h) + \phi_{2h}(N_1)]$$

A estimação de máxima verosimilhança de θ envolve assim 3 tabelas de frequências que não são mutuamente exclusivas. As probabilidades $\{\rho_{(2hm)}^i(\theta)\}$ são estimadas apenas da tabela completamente categorizada. As probabilidades $\{\hat{\Gamma}_{(2h)}^m(\theta)\}$ são, por sua vez, estimadas da tabela de s_2 categorias envolvendo as frequências N_2 mais as correspondentes somas dos elementos de N_1 . Por fim, as probabilidades $\{\phi_{3h}(\theta)\}$ são estimadas da tabela de s_3 categorias contendo as frequências N_3 e as somas apropriadas dos elementos de N_2 e N_1 .

A conjugação das expressões (3.38) conduz a que (3.38a) se possa exprimir, considerando $N = \sum_{t=1}^3 N_t$, por

$$\hat{\theta}_i = \frac{1}{N} \left\{ n_{1i} + n_{2hm} \left(\frac{n_{1i}}{\sum_{i\epsilon\delta_{hm}^2} n_{1i}} \right) + n_{3h} \left(\frac{n_{1i}}{\sum_{i\epsilon\delta_{hm}^2} n_{1i}} \right) \left(\frac{n_{2hm} + \sum_{i\epsilon\delta_{hm}^2} n_{1i}}{\sum_{m=1}^{j_h} (n_{2hm} + \sum_{i\epsilon\delta_{hm}^2} n_{1i})} \right) \right\} \quad (3.39)$$

ilustrando que $\hat{\theta}_i$, $16\delta_{hm}^2$, é a proporção envolvendo os indivíduos classificados em i , a fracção dos indivíduos classificados em δ_{hm}^2 determinada pela estimativa de probabilidade amostral de pertencer a i dado δ_{hm}^2 , e a fracção dos indivíduos classificados em δ_h^3 calculada das estimativas da probabilidade anterior e da probabilidade de pertencer a δ_{hm}^2 dado δ_h^3 .

O esquema conducente à factorização (3.36) de $L_1(\theta|\underline{n})$ deixa antever como se factoriza a verosimilhança no caso de partições encaixadas em $T-1$ níveis ($T>3$) e a forma de (3.39) permite prever qual a expressão explícita das estimativas nesse caso mais geral.

Frequentemente, os conjuntos de categorias confundidas que são observados não se conseguem aglutinar exaustivamente em partições, embora configurem uma estrutura encaixada no seguinte sentido: Os conjuntos δ podem agrupar-se em classes, algumas das quais constituem partições do conjunto original $\{1, \dots, s_t\}$, e para qualquer par de δ 's, ou um é subconjunto do outro ou são disjuntos.

Neste caso, as classes podem ser estruturadas de modo que cada classe é uma partição de um elemento de outra classe, classe esta, por sua vez, partição de um elemento de uma outra classe, e assim sucessivamente até se chegar à classe que forma a partição (mais grossa, se houver várias partições) do conjunto original de categorias. Em termos ilustrativos, e retomando o exemplo com $s_1=12$, o padrão de censura com as classes observadas $C_3 = \{\{1,2,3,4,5,6\}, \{7,8,9,10,11,12\}\}$, $C_2 = \{\{1,2,3\}, \{4,5,6\}, \{7,8,9,10\}, \{11,12\}\}$ e $C_1 = \{\{7,8\}, \{9,10\}\}$ representa uma estrutura encaixada em 3 níveis.

O método acima desenvolvido para obter uma expressão em forma fechada para $\hat{\theta}$ é aplicável na sua essência a esta situação. A grande diferença está em que cada sub-matriz da matriz \underline{Z} relativa aos conjuntos observados, indicadora de uma classe que não é partição do conjunto original de categorias, particiona-se apenas nas sub-matrizes associadas aos elementos da classe adjacente que são uniões dos elementos daquela classe. Embora este aspecto não dificulte a obtenção de uma factorização da verosimilhança do tipo de (3.36), a consideração da matriz \underline{Z} observada como coincidente com a matriz admissível acarreta problemas na interpretação distribucional dos estimadores dos parâmetros da verosimilhança factorizada. Este inconveniente será eliminado se considerarmos também como admissíveis os conjuntos δ - cujas frequências observadas serão encaradas como zeros amostrais - que, adicionados aos δ 's observados, fazem com que cada classe constitua uma partição sem violarem a estrutura encaixada. Ilustrando com o exemplo anterior, o padrão observado com censura em C_3 e $\{\{7,8,9,10\}, \{11,12\}\}$, pode ser interpretado como se $T = C_3 \cup C_2$, ou seja, como um padrão em partições encaixadas em 2 níveis, com $n_{123} = n_{456} = 0$, para o qual são válidos, nomeadamente, os resultados (3.36), (3.37) e (3.39).

4. ANÁLISE CLÁSSICA CONDICIONAL DE DADOS COM CATEGORIZAÇÃO COMPLETA E PARCIAL

4.1. Introdução

Vimos no capítulo anterior que a ocorrência de dados incompletos pode ser interpretada como consequência da actuação de um mecanismo estocástico de registo. A introdução formal deste mecanismo gera, contudo, um modelo estatístico não identificável e com um número de parâmetros superior ao número de observações, caso não se adoptem hipóteses simplificadoras sobre o mecanismo de registo.

A consideração de um processo de registo estabelecendo que as probabilidades condicionais de registo só dependem dos valores registados (modelo designado por M_1) já permite sanar alguns dos problemas. O modelo estatístico torna-se identificável para várias configurações dos dados incompletos (nomeadamente, quando um subconjunto da amostra é completamente categorizado). Os seus parâmetros reduzem-se a um número desejável. Procedimentos inferenciais, via teoria da amostragem, sobre as probabilidades das categorias podem então ser desbloqueados. Para este efeito, o modelo M_1 não é, porém, suficiente para a erradicação dos parâmetros perturbadores da análise.

A consideração no âmbito de um padrão de registo em partições, de um modelo mais restritivo estipulando que as probabilidades condicionais de registo são constantes para todos os elementos de cada partição (modelo M_2), possibilita já a realização de inferências correctas sobre as probabilidades das celas ignorando o mecanismo de registo.

Estas inferências revestem uma natureza condicional (ao total de elementos registados em cada partição) e é a descrição da teoria estatística subjacente que constitui o objectivo deste capítulo. Nele serão postas em destaque as metodologias de máxima verosimilhança (MV) e de mínimos quadrados ponderados (MQP), sem deixar de fazer referência às suas ligações com outros métodos inferenciais clássicos. Antes porém, será feita uma revisão — que não se reivindica ser completa — da literatura sobre a análise não Bayesiana de dados categorizados incompletos nas suas diversas configurações.

4.2. Breve revisão da literatura

A maioria das primeiras publicações assume apenas implicitamente que o processo responsável pela falta de classificação precisa de uma observação não depende da verdadeira categoria subjacente ou, pelo menos, das categorias "missing". Rubin (1976) fornece, num contexto geral, condições para que as inferências traçadas sobre os parâmetros de interesse, ignorando o processo que gera os dados perdidos, sejam correctas. Embora, em alguns casos, a natureza incompleta dos dados possa, de um modo mais realista, depender dos valores que seriam observados

sob uma categorização completa, actualmente, tais hipóteses de ignorabilidade, explicitadas mais frequentemente, são mantidas na maioria dos trabalhos.

Esta prática afigura-se compreensível devido aos problemas que a introdução de um mecanismo aleatório de censura acarreta para as análises "clássicas", conforme foi atrás salientado. A consideração de um mecanismo não ignorável exige algum conhecimento prévio sobre os seus parâmetros. É neste quadro que se inserem, designadamente:

- o trabalho de Little (1980,1982), no qual se estende a definição de ignorabilidade em inquéritos amostrais de modo a incorporar, também, o mecanismo de amostragem, e se modela o mecanismo de não-resposta através da especificação das chances de registo em cada subconjunto observado para cada par de categorias;
- a análise por MV de Nordheim (1984), pressupondo como conhecidos ou estimados de dados independentes os parâmetros que representam razões de probabilidades de classificação parcial;
- a aplicação interessante à tomografia de emissão de positrões de um modelo Poissoniano por Vardi et al. (1985) onde as probabilidades condicionais de registo são consideradas conhecidas;
- o estudo da estrutura de tabelas de contingência por Fay (1986), através de modelos causais log-lineares para o padrão de não-resposta, tal como foram formulados por Goodman (vide suas referências).

Todas as referências citadas a seguir assumem condições de ignorabilidade do mecanismo de registo para efeitos de traçar inferências sobre as probabilidades das celas, ainda que algumas delas (e.g., Chen and Fienberg [1974]) lidem também com a estimação dos parâmetros relacionados com a categorização parcial. Como veremos na sequência, os procedimentos mais comuns são baseados no método de MV. Por motivos de clareza da exposição, separá-los-emos dos outros métodos.

4.2.1. Métodos de máxima verosimilhança

As primeiras análises de dados incompletos preocuparam-se com configurações especiais em contextos específicos. Isto é ilustrado nos procedimentos de Craig (1953) e Watson (1956), os quais trataram com alguns padrões simples de frequências aglutinadas em tabelas de contingência bidimensionais geradas por um modelo multinomial sob a hipótese de independência.

Tentando ligar este tipo de problema envolvendo dados discretos agrupados e censurados com a análise para dados completos, Hartley (1958) introduziu um procedimento iterativo geral para calcular estimativas de MV desse modo. Esta ideia básica foi usada mais tarde por Blight (1970) para amostras censuradas e truncadas da família exponencial e por Carter and Myers (1973) para dados agrupados de uma combinação linear de distribuições discretas.

Hocking and Oxspring (1971,1974) consideraram a situação de várias tabelas categorizadas, umas das quais completamente, supondo os dados obtidos por delineamento e modelados por uma distribuição multinomial-produto. Eles derivaram as equações de verosimilhança de uma forma que sugere um procedimento iterativo específico para a determinação dos estimadores de MV e da sua matriz de covariâncias assintótica estimada. Condições que garantem a sua convergência são dadas em Hartley and Hocking (1971), os quais usaram um procedimento semelhante para dados de distribuições normais. Parece que Blumenthal (1968) foi o primeiro a tomar em conta o processo gerador da classificação parcial, ainda que num contexto específico. Usando modelos log-lineares em tabelas de contingência de dupla entrada, geradas de distribuições de Poisson e multinomiais, e detalhando o caso de marginais suplementares, Chen and Fienberg (1974) generalizaram o trabalho de Blumenthal (1968) e de Hocking and Oxspring (1971), modelando o processo pelo qual algumas observações perdem a sua identidade de linha ou de coluna. Admitiram implicitamente que a perda de dados foi "aleatória", ou seja, um mecanismo MAR do tipo do modelo M_1 .

Como outro exemplo de um procedimento particular, Goodman (1968) analisou tabelas de contingência bidimensionais com certos padrões de celas combinadas por aplicação do seu método de particionamento em sub-tabelas separáveis. Posteriormente, na sua análise de tabelas multidimensionais com uma estrutura latente, Goodman (1974) usou um procedimento iterativo para obter as estimativas de MV dos parâmetros de modelos multiplicativos identificáveis. Sob modelagem log-linear de tabelas de frequências com algumas celas indistinguíveis, geradas por esquemas de Poisson e multinomiais, Haberman (1974,1976) derivou as equações de verosimilhança por métodos de projecção ortogonal e propôs algoritmos de escalonamento iterativo para a sua solução. Chen and Fienberg (1976) propuseram um outro método de solução dessas equações relativas a tabelas multidimensionais sob modelos identificáveis, generalizando os resultados mencionados de Craig (1953) e Watson (1956), bem como os de Cohen (1971) respeitantes a tabelas 2^k censuradas.

Várias tentativas foram feitas no sentido de se definir um procedimento geral de MV aplicável a qualquer padrão de dados incompletos independentemente do modelo gerador. Woodbury (1971) e Orchard and Woodbury (1972) tentaram dar uma metodologia geral (o seu "missing information principle") que levasse às estimativas de MV usando quase as mesmas técnicas que são aplicadas aos dados completos. Na mesma direcção, Sundberg (1974,1976) estendeu os tratamentos de Hartley (1958) e Blight (1970) a vários tipos de perda de dados dentro da família exponencial, e analisou as propriedades de convergência do método iterativo definido. Na sequência, Dempster, Laird and Rubin (1977) clarificaram, unificaram e generalizaram essa metodologia, criando a designação EM (Expectação - Maximização) para o correspondente algoritmo e ilustrando-o com uma larga gama de exemplos bem variados. Este trabalho permite-nos ver os métodos de Goodman (1974) e de Chen and Fienberg (1976) como aplicações do algoritmo EM.

Há muitas vezes necessidade de conjugar o algoritmo EM com outros algoritmos para executar o passo M. Para tal fim, Fuchs (1982) empregou o algoritmo de ajustamento proporcional iterativo (IPF) para estimar os parâmetros multinomiais sob modelagem log-linear de tabelas multidimensionais suplementadas por tabelas de dimensão inferior. Little and Schluchter (1985) usaram um procedimento semelhante num contexto mais geral envolvendo conjuntamente dados categorizados e contínuos. Ambos os artigos discutem ainda o caso especial onde a estimação é simplificada pela ocorrência de um padrão de dados monótono, como Rubin (1974) demonstrara. O algoritmo descrito em Haber (1984) combina também o algoritmo EM com o método de escalonamento iterativo de Darroch-Ratcliff (1972) e pode ser usado em ambos os casos de tabelas parcialmente categorizadas e de tabelas truncadas.

Outros métodos podem ser aplicados para estimação de MV em problemas de dados incompletos. Por exemplo, Haberman (1977), propondo uma classe de modelos, não necessariamente log-lineares, para tabelas de frequências com categorias combinadas, estudou o comportamento assintótico de estimadores de MV e comparou os algoritmos EM, Newton-Raphson e "Scoring" para solução das respectivas equações de verosimilhança. Um outro exemplo, aplicável ao caso de tabelas completa e parcialmente categorizadas, é o algoritmo em 2 passos desenvolvido por Haber and Brown (1986). Este algoritmo, empregando o método de escalonamento iterativo generalizado e um método para obtenção de estimativas de MV sob restrições lineares (como o de Gokhale [1973], por exemplo) permite ajustar modelos contendo restrições log-lineares e lineares a dados multinomiais.

4.2.2, Outros métodos

Dada a estrutura do modelo que consideramos ser gerador dos dados arranjados numa tabela completa e em várias tabelas parcialmente categorizadas, a abordagem por MQP a este problema consiste numa extensão relativamente linear da metodologia de Grizzle, Starmer and Koch (1969), doravante designada abreviadamente por metodologia GSK. Devido à sua possível gama de aplicações, devemos ainda citar uma outra extensão, a de Bonnett et al. (1985), a qual toma em consideração uma classe geral de distribuições de Poisson condicionais, funções resposta com matriz de covariância singular e restrições lineares estocásticas para os parâmetros de modelos lineares. No que concerne ao ajuste do modelo log-linear tradicional, a sua metodologia envolvendo inversas generalizadas surge como uma alternativa às formulações de Grizzle and Williams (1972) e Imrey et al. (1981, 1982).

A primeira tentativa publicada de aplicação da metodologia GSK ao problema de dados incompletos supra-citado parece ter sido devida a Koch and Reinfurt (1970), os quais assumiram explicitamente que as observações incompletas ocorrem por delineamento (modelo M_0 do cap. 3) e não aleatoriamente. Considerando os da

dos incompletos estruturados em sub-tabelas marginais, eles estenderam a abordagem GSK e desenvolveram um procedimento em duas fases para determinar estimativas ajustadas que reflectem todos os dados disponíveis e testar várias hipóteses de interesse. Kock, Imrey and Reinfurt (1972) preocuparam-se em descrever alguns aspectos adicionais e em ilustrar este procedimento de MQP através de vários exemplos.

Stanish, Gillings and Koch (1978) aplicaram esta abordagem na análise de dados longitudinais incompletos através de métodos "multivariate ratio". Woolson and Clarke (1984) trataram o mesmo tipo de problema em estudos longitudinais, adicionando à variável resposta mais uma categoria (a categoria "missing"), em ordem a obter o mesmo número de categorias para todas as subpopulações multinomiais definidas pelos níveis dos factores. Tal modificação permite uma aplicação mais padronizada da metodologia GSK básica e dos programas de computador convencionais. Este mesmo procedimento é adoptado por Haber (1986) ao testar a independência por par com dados incompletos de estudos longitudinais e de experiências de captura-recaptura (neste último caso, o tamanho da população é também estimado sob aquela hipótese de independência).

Dado o suporte probabilístico das inferências condicionais citadas em 4.1, a análise por informação de discriminação mínima do problema de restrições externas em várias amostras multinomiais independentes, como é descrita em Gokhale and Kullback (1978, cap. 5), pode ser aplicada de um modo idêntico a esta situação de dados incompletos.

Finalmente, para mencionar ainda outros métodos de análise, devemos referir aqueles descritos em Titterington (1984) e Titterington and Jiang (1983). Numa tentativa de obter estimativas rápidas na presença de grandes conjuntos de dados com possibilidade de registos incompletos, eles propuseram e estudaram vários procedimentos recursivos (alguns dos quais relacionados com o método "Scoring" e outros com o algoritmo EM) conducentes a estimadores que possuem boas propriedades assintóticas sob certas condições de regularidade.

4.3 Métodos de máxima verosimilhança

4.3.1. Solução das equações de verosimilhança para o modelo "saturado"

De acordo com o que foi visto no capítulo 3, o modelo M_2 é um modelo MAR pelo que as equações de verosimilhança correspondem a (3.11). Dado o padrão de censura em partições que trataremos ao longo deste capítulo, tais equações podem ser definidas, em consonância com a notação de 3.3 que usaremos neste capítulo, por

$$\hat{\theta}_i = \left[n_{1i} + \sum_{t=2}^T \sum_{j=1}^{s_t} (n_{tj} \hat{\theta}_i / z'_{tj} \hat{\theta}) I_{\delta_j^t}(i) \right] / N, \quad i=1, \dots, s_1 \quad (4.1)$$

onde $N = \sum_{t=1}^T N_t$, $N_t = \sum_{j=1}^{s_t} N_{tj}$, $t=1, \dots, T$, $Z_t = (z_{tj}, j=1, \dots, s_t)$ é a matriz $s_t \times s_t$ indicadora da partição Π_t , $t=2, \dots, T$, e $I_{\delta_j}^t$ é a função indicadora do conjunto $\delta_j^t \in \Pi_t$. Numa formulação matricial, útil para a sua implementação em programas de computador, (4.1) corresponde a

$$\hat{\theta} = (N_1 + \sum_{t=2}^T D_{\hat{\theta}} Z_t D_{Z_t \hat{\theta}}^{-1} N_t) / N \quad (4.2)$$

onde $D_{\hat{a}}$ indica a matriz diagonal com os elementos do vector \hat{a} na diagonal principal.

É fácil comprovar que (4.2) representa a concretização da expressão de Sundberg (1974) para o sistema das equações de verosimilhança relativas a dados incompletos da família exponencial. Isto é, sendo \underline{x} o vector das frequências que se obteria na eventualidade do processo de amostragem ser plenamente observado, (4.2) corresponde a $E(\underline{x} | N, \hat{\theta}) = E(\underline{x} | \underline{n}, \hat{\theta})$, onde $\underline{n} = (N_1', \dots, N_T)'$. Como referimos no capítulo anterior, o sistema (4.2) pode ser resolvido iterativamente pelo simples mas lento algoritmo EM.

A simplicidade conceptual e analítica do algoritmo EM tem como contrapartida a incapacidade de estimar, por si só, a matriz de covariância assintótica do estimador de MV (veja-se, contudo, a adaptação de Louis [1982], e.g., visando calcular a matriz de informação observada). Em ordem a derivar a matriz de informação esperada, anotemos previamente alguns aspectos notacionais. Seja $\theta_{tj} = Z_{tj}' \theta = (\theta_{tj})$ o vector das probabilidades amostrais das celas da tabela t , $t=2, \dots, T$ e $\bar{\theta}_t$ (resp. $\bar{\theta}$) o vector $\theta_t(\theta)$ truncado da última componente, i.e., $\bar{\theta}_t = A_t \theta_t$ ($\bar{\theta} = A_1 \theta$), onde $A_t = (I_{s_t-1} \ 0_{(s_t-1)})$, $t=1, \dots, T$. Correspondentemente, seja $\bar{p}_t = A_t p_t$, $t=1, \dots, T$, onde $p_t = N_t / N_t$ é o vector das proporções dentro da tabela t .

Em virtude de (3.30), a função de verosimilhança condicional é $L(\bar{\theta} | \underline{n}) =$

$\prod_{t=1}^T L_t(A_t Z_t' \theta | N_t)$, onde $Z_1 = I_{s_1}$, e $L_t(\cdot)$ é numericamente igual à função de probabilidade multinomial de parâmetros N_t , e $Z_t' \theta$. Denotando por $V_{\bar{p}_t}(\bar{\theta}_t) = (D_{\bar{p}_t} - \bar{\theta}_t \bar{\theta}_t') / N_t$, a matriz de covariâncias, condicional a N_t , de \bar{p}_t , é fácil verificar que

$$\frac{\partial \log L_t}{\partial \bar{\theta}_t} = \left[V_{\bar{p}_t}(\bar{\theta}_t) \right]^{-1} (\bar{p}_t - \bar{\theta}_t) \quad (4.3)$$

e que a matriz Hessiana relativa a $\log L_t$ é

$$H_t(\bar{\theta}_t) = \frac{\partial^2 \log L_t}{\partial \theta_t \partial \theta_t'} = - \left[D_{\bar{v}_t} + \frac{(1 - J_t' \bar{p}_t) N_t}{(1 - J_t' \bar{\theta}_t)^2} J_t J_t' \right] \quad (4.4)$$

onde \bar{v}_t é o vector $(s_t-1) \times 1$ de elementos n_{tj} / θ_{tj}^2 , $j=1, \dots, s_t-1$ e $J_t = 1_{(s_t-1)}$. Esta expressão reflecte claramente a concavidade da função $\log L_t$.

Tendo em conta que $\bar{\theta}_t = \bar{z}_t' \bar{\theta}$, onde \bar{z}_t é a matriz Z_t sem a última linha e a última coluna, e usando a regra em cadeia da diferenciação matricial, obtêm-se

$$\frac{\partial \log L_t}{\partial \bar{\theta}} = \frac{\partial \bar{\theta}_t'}{\partial \bar{\theta}} \frac{\partial \log L_t}{\partial \bar{\theta}_t} = \bar{z}_t \left[\underset{\sim}{V}_{\underset{\sim}{P}_t}(\bar{\theta}_t) \right]^{-1} (\bar{p}_t - \bar{\theta}_t) \quad (4.5)$$

$$\frac{\partial^2 \log L_t}{\partial \bar{\theta} \partial \bar{\theta}'} = \frac{\partial}{\partial \bar{\theta}_t'} \left[\bar{z}_t \frac{\partial \log L_t}{\partial \bar{\theta}_t} \right] \frac{\partial \bar{\theta}_t}{\partial \bar{\theta}'} = \bar{z}_t H_t(\bar{\theta}_t) \bar{z}_t' \quad (4.6)$$

Por conseguinte, o vector gradiente e a matriz Hessiana de $\log L$ são respectivamente dados por

$$\frac{\partial \log L}{\partial \bar{\theta}} = \sum_{t=1}^T \bar{z}_t \left[\underset{\sim}{V}_{\underset{\sim}{P}_t}(\bar{z}_t' \bar{\theta}) \right]^{-1} (\bar{p}_t - \bar{z}_t \bar{\theta}) \quad (4.7)$$

$$H(\bar{\theta}) = \sum_{t=1}^T \bar{z}_t H_t(\bar{z}_t' \bar{\theta}) \bar{z}_t' \quad (4.8)$$

Tomando em consideração que $E[-H_t(\bar{\theta}_t)]_{N_t} = \left[\underset{\sim}{V}_{\underset{\sim}{P}_t}(\bar{\theta}_t) \right]^{-1}$, a matriz de informação de Fisher condicional é então dada por

$$F(\bar{\theta}) = \sum_{t=1}^T \bar{z}_t \left[\underset{\sim}{V}_{\underset{\sim}{P}_t}(\bar{z}_t' \bar{\theta}) \right]^{-1} \bar{z}_t' = X' \left[\underset{\sim}{V}_{\underset{\sim}{P}}(\bar{\theta}) \right]^{-1} X \quad (4.9)$$

onde $\bar{p} = (\bar{p}_1' \bar{p}_2' \dots \bar{p}_T)'$, $\underset{\sim}{V}_{\underset{\sim}{P}}(\bar{\theta})$ é a matriz diagonal em blocos, $\text{diag}(\underset{\sim}{V}_{\underset{\sim}{P}_1}(\bar{\theta}),$

$\underset{\sim}{V}_{\underset{\sim}{P}_2}(\bar{z}_2' \bar{\theta}) \dots \underset{\sim}{V}_{\underset{\sim}{P}_T}(\bar{z}_T' \bar{\theta}))$ e $X = (\bar{z}_1 \bar{z}_2 \dots \bar{z}_T)'$. Deste modo, para grandes valores de $N_{t.}$, $t=1, \dots, T$, $\hat{\theta}$ tem aproximadamente uma distribuição normal multivariada, com vector média $\bar{\theta}$ e a matriz de covariância $\hat{V}(\bar{\theta}) = [F(\bar{\theta})]^{-1}$.

As expressões (4.7), (4.8) e (4.9) permitem maximizar o logaritmo da função de verosimilhança através dos algoritmos de Newton-Raphson e "Scoring" de Fisher. Este último proporciona ainda uma estimativa directa de $\hat{V}(\bar{\theta})$.

Esta característica é também partilhada pelo método iterativo específico projectado por Hocking and Oxspring (1971) para o caso em que a ocorrência das várias tabelas é determinada por delineamento. Como tal mecanismo determinístico é equivalente a um mecanismo aleatório de registo modelado por M_2 , do ponto de vista de realização de inferências sobre θ condicionalmente a $\{N_{t.}\}$, esse procedimento iterativo é aplicável à situação que estamos tratando. Ele é sugerido pela forma de (4.7), a qual permite definir as equações de verosimilhança por

$$\hat{\theta} = \hat{V}(\hat{\theta}) \sum_{t=1}^T \bar{z}_t \left[\underset{\sim}{V}_{\underset{\sim}{P}_t}(\hat{\theta}_t) \right]^{-1} \bar{p}_t \quad (4.10)$$

onde $\hat{\theta}_t = \bar{z}_t' \hat{\theta}$, $t=1, \dots, T$. Para a definição do seu algoritmo, Hocking and Oxspring (1971) usam para a matriz $\hat{V}(\bar{\theta})$ a expressão

$$\hat{\bar{\theta}} = \prod_{t=2}^T \left[\bar{Z}'_1 + B_t(\bar{\theta}) \bar{Z}'_t \right] \bar{v}_{\bar{p}_1}(\bar{\theta}) \quad (4.11)$$

onde $B_t(\cdot)$, $t=2, \dots, T$ são matrizes $(s_1-1) \times (s_t-1)$ definidas por

$$B_t(\bar{\theta}) = - \bar{v}_{\bar{p}_1}(\bar{\theta}) \bar{Z}'_t \left[\bar{Z}'_t \bar{v}_{\bar{p}_1}(\bar{\theta}) \bar{Z}'_t + \bar{v}_{\bar{p}_t}(\bar{\theta}) \right]^{-1} \quad (4.12)$$

com $\bar{v}_{\bar{p}_1}(\bar{\theta})$ denotando a matriz de covariâncias assintótica do estimador de MV de $\bar{\theta}$ quando calculado das primeiras $t-1$ verosimilhanças, i.e., de $\prod_{m=1}^{t-1} L_m(\bar{\theta} | N_m)$ para $t=2, \dots, T$. O seu esquema iterativo fica então definido por

$$\bar{\theta}^{(v+1)} = \bar{c}^{(v)} \sum_{t=1}^T \bar{Z}'_t \bar{D}^{(v)} \bar{p}_t, \quad 0, 1, 2, \dots \quad (4.13)$$

onde $\bar{c}^{(v)}$ representa $\hat{\bar{v}}(\bar{\theta}^{(v)})$ na expressão (4.10) e $\bar{D}^{(v)} = \left[\bar{v}_{\bar{p}_t}(\bar{\theta}^{(v)}) \right]^{-1}$.

Definindo as quantidades

$$q_1(\bar{\theta}) = q_1 = \bar{p}_1; \quad q_t(\bar{\theta}) = q_{t-1}(\bar{\theta}) + B_t(\bar{\theta}) \left[\bar{Z}'_t q_{t-1}(\bar{\theta}) - \bar{p}_t \right], \quad t=2, \dots, T$$

Hocking and Oxspring (1971, sec.3) provam que (4.10) pode ser posta na forma

$$\hat{\bar{\theta}} = \hat{q}_{T-1} + \hat{B}_T(\bar{Z}'_T \hat{q}_{T-1} - \bar{p}_T) = \bar{p}_1 + \sum_{t=2}^T \hat{B}_t(\bar{Z}'_t \hat{q}_{t-1} - \bar{p}_t) \quad (4.14)$$

onde $\hat{B}_t = B_t(\hat{\bar{\theta}})$ e $\hat{q}_t = q_t(\hat{\bar{\theta}})$, $t=2, \dots, T$. Como $\bar{\theta}_t$ é uma função paramétrica identificável em L_t , (4.14) mostra, como foi por eles salientado, que o estimador de MV de $\bar{\theta}$ se obtém do correspondente estimador para os dados completamente categorizados, adicionando apropriadas combinações lineares das diferenças entre os estimadores de $\bar{\theta}_t$ calculados de L_t e de $\prod_{m=1}^{t-1} L_m$.

A expressão (4.14) permite-nos redefinir (4.13) como

$$\begin{aligned} q_1^{(v)} &= \bar{p}_1, \quad \forall v \\ q_t^{(v)} &= q_t(\bar{\theta}^{(v)}) = q_{t-1}^{(v)} + B_t(\bar{\theta}^{(v)}) \left[\bar{Z}'_t q_{t-1}^{(v)} - \bar{p}_t \right], \quad t=2, \dots, T \\ \bar{\theta}^{(v+1)} &= q_T^{(v)}, \quad v=1, 2, \dots \end{aligned} \quad (4.15)$$

Se se considerar o seguinte procedimento modificado

$$u_1 = \bar{p}_1 \quad (4.16)$$

$$u_t = q_t(u_{t-1}) = u_{t-1} + B_t(u_{t-1}) \left[\bar{z}'_t u_{t-1} - \bar{p}_t \right], \quad t=2, \dots, T$$

a. sua comparação com (4.14) requer, para $u_{\sim T}$ coincidir com $\hat{\theta}$, que $B_t(u_{\sim t-1}) = B_t(u_{\sim T})$, $t=2, \dots, T$. Isto é assegurado num padrão de censura encaixado (Hocking and Oxspring [1971, sec.4]), devido a uma estrutura especial das matrizes $B_t(\cdot)$. Esta estrutura pode ser definida por

$$B_t(\bar{\theta}) = -\left(N_t / \sum_{m=1}^t N_m \right) V_{\sim t}(\bar{\theta}) \bar{z}'_t \left[\bar{z}'_t V_{\sim 1}(\bar{\theta}) \bar{z}'_t \right]^{-1} \quad (4.17)$$

como se pode provar por indução. Em suma, o estimador de MV de $\bar{\theta}$ no caso de um padrão de censura em partições encaixadas é explicitamente dado por $u_{\sim T}$ definido em (4.16).

No capítulo 3 deduzimos, pelo método de factorização da verosimilhança, a expressão explícita de $\hat{\theta}$ para um padrão de registo em partições encaixadas. Não é difícil provar que $u_{\sim T}$ em (4.16) corresponde efectivamente a tal expressão. Por exemplo, para um padrão encaixado em um nível ($T=2$), a efectivação de cálculos algébricos simples mas morosos conduz a que a componente u_{2i} de $u_{\sim 2}$ em (4.16), $i=1, \dots, s_1-1$ tal que $i \in \delta_j^2$, $j=1, \dots, s_2-1$, se possa definir como

$$\begin{aligned} u_{2i} &= \bar{p}_{1i} - (N_2./N) \left(\bar{p}_{1i} / \sum_{i \in \delta_j^2} \bar{p}_{1i} \right) \left(\sum_{i \in \delta_j^2} \bar{p}_{1i} - \bar{p}_{2j} \right) \\ &= \bar{p}_{1i} - (N_2./N) \left[\bar{p}_{1i} - \bar{p}_{2j} \left(\bar{p}_{1i} / \sum_{i \in \delta_j^2} \bar{p}_{1i} \right) \right] \\ &= \left[n_{1i} + n_{2j} \left(n_{1i} / \sum_{i \in \delta_j^2} n_{1i} \right) \right] / N, \quad N = N_1. + N_2. \end{aligned} \quad (4.18)$$

que é a expressão (3.35a). Identicamente, para uma configuração encaixada em dois níveis ($T=3$), pode provar-se que u_{3i} para $i \in \delta_{hm}^2$, $i=1, \dots, s_1-1$, $m=1, \dots, j_h-1$, $h=1, \dots, s_3-1$, em (4.16) é

$$\begin{aligned} u_{3i} &= u_{2i} - (N_3./N) \left(u_{2i} / \sum_{i \in \delta_h^2} u_{2i} \right) \left(\sum_{i \in \delta_h^2} u_{2i} - \bar{p}_{3h} \right) \\ &= (N_1. + N_2.) u_{2i} / N + (N_3./N) \bar{p}_{3h} u_{2i} / \sum_{i \in \delta_h^2} u_{2i}, \quad N = \sum_{t=1}^3 N_t. \end{aligned} \quad (4.19)$$

Tendo em conta a expressão de u_{2i} para $i \in \delta_{hm}^2$, obtêm-se

$$u_{3i} = \left[n_{1i} + n_{2hm} \frac{n_{1i}}{\sum_{i \in \delta_{hm}^2} n_{1i}} + n_{3h} \frac{n_{1i}}{\sum_{i \in \delta_{hm}^2} n_{1i}} \frac{n_{2hm} + \sum_{i \in \delta_{hm}^2} n_{1i}}{j_h} \right] / N$$

ou seja, a expressão (3.39).

Sublinhe-se ainda que a comparação destas expressões em forma fechada para $\hat{\theta}$ com (4.2) torna patente que o algoritmo EM pode produzir as estimativas de MV ao fim da 1ª iteração, usando como estimativas iniciais, respectivamente, $\bar{\theta}^{(0)} = \bar{p}_1$ e $\bar{\theta}^{(0)} = \underline{u}_2$, nos dois casos acima ilustrados.

4.3.2 Máxima verosimilhança versus outros métodos de estimação

Dado o tipo de modelo probabilístico que estamos considerando neste problema de dados incompletos, o estimador de MV de $\bar{\theta}$ é também uma aproximação para grandes valores de $\{N_t\}$ do estimador do χ^2 de Pearson mínimo (abreviadamente QPM), como se pode demonstrar seguindo o raciocínio de Cramer (1946, p.426). De facto, a função do χ^2 de Pearson pode ser expressa em termos matriciais por

$$X_P^2(\bar{\theta}) = \sum_{t=1}^T X_{P_t}^2(\bar{\theta}), \quad X_{P_t}^2(\bar{\theta}) = N_t \cdot (\underline{p}_t - \bar{\theta}_t)' D_{\bar{\theta}_t}^{-1} (\underline{p}_t - \bar{\theta}_t) \quad (4.20)$$

onde $\sum_{s=1}^{s_t} \theta_{ts} = 1$, $\theta_{ts} = Z' \theta (Z = I_{s_t})$ e $\underline{p}_t = N_t / N_t$, $t=1, \dots, T$. Tomando em consideração as restrições naturais

$$X_{P_t}^2 \equiv X_{P_t}^2(\bar{\theta}) = (\underline{p}_t - \bar{\theta}_t)' \left[\underline{V}_{\bar{p}_t}(\bar{\theta}_t) \right]^{-1} (\underline{p}_t - \bar{\theta}_t)$$

e portanto, fazendo $\bar{\theta}_* = \underline{X} \bar{\theta}$,

$$X_P^2(\bar{\theta}) = (\underline{p} - \bar{\theta}_*)' \left[\underline{V}_{\bar{p}}(\bar{\theta}) \right]^{-1} (\underline{p} - \bar{\theta}_*) \quad (4.21)$$

Diferenciando X_P^2 em ordem a θ_{tj} , $j=1, \dots, s_t-1$, obtém-se

$$\frac{\partial X_P^2}{\partial \theta_{tj}} = -2N_t \cdot \left(\frac{p_{tj}}{\theta_{tj}} - \frac{p_{ts_t}}{\theta_{ts_t}} \right) - \frac{1}{N_t} \left[\left[\frac{n_{tj} - N_t \cdot \theta_{tj}}{\theta_{tj}} \right]^2 - \left[\frac{n_{ts_t} - N_t \cdot \theta_{ts_t}}{\theta_{ts_t}} \right]^2 \right], j=1, \dots, s_t-1$$

Para N_t suficientemente grande, o 2º termo do 2º membro é desprezável, o que nos permite escrever

$$\begin{aligned} \frac{\partial X_P^2}{\partial \theta_t} &= -2N_t \cdot \left(D_{\bar{\theta}_t}^{-1} + \frac{J_t J_t'}{1 - J_t' \bar{\theta}_t} \right) (\underline{p}_t - \bar{\theta}_t) + O_p(N_t^{-1}) \\ &= -2 \frac{\partial \log L_t}{\partial \bar{\theta}_t} + O_p(N_t^{-1}) \end{aligned} \quad (4.22)$$

tendo em vista (4.3). Assim, a afirmação acima sobre a relação assintótica entre os estimadores de MV e do QPM fica demonstrada.

A expressão (4.21) mostra ainda que o estimador do QPM de $\bar{\theta}$ pode ser visto

como o estimador dos MQP obtido do modelo linear generalizado, condicionado em $\{N_{t.}\}$, $\bar{p} = \sum \bar{\theta} + \bar{\varepsilon}$, onde o erro tem média nula e matriz de covariância $V_{\bar{p}}(\bar{\theta})$. De acordo com Clickner (1976), o estimador de MV pode ser visto de um modo análogo. De facto, (4.22) revela que a remoção do termo desprezável é equivalente à minimização de X_P^2 supondo que $V_{\bar{p}}(\bar{\theta}_{\sim t})$ é independente de $\bar{\theta}$. Assim, as equações de verosimilhança podem ser interpretadas como as equações normais resultantes de (4.21) com $V_{\bar{p}}(\bar{\theta})$ suposto independente de $\bar{\theta}$. Deve também anotar-se que os estimadores de MV e do QPM de $\bar{\theta}$ (ou de $\bar{\theta}_*$) são ambos estimadores BAN quando $N \rightarrow \infty$ com $\{N_{t.}/N\}$ mantidos fixos, como se segue do trabalho de Neyman (1949).

É ainda conveniente realçar que (4.21), (4.3) e (4.4) implicam que a função do χ^2 de Pearson corresponde à generalização para T amostras da função "score" eficiente de Rao (Rao [1947]), dada por

$$S(\bar{\theta}) = \sum_{t=1}^T N_t^{-1} \left(\frac{\partial \log L_t}{\partial \bar{\theta}_{\sim t}} \right)' \left[F(\bar{\theta}_{\sim t}) \right]^{-1} \left(\frac{\partial \log L_t}{\partial \bar{\theta}_{\sim t}} \right) \quad (4.23)$$

onde $F(\bar{\theta}_{\sim t})$ é a matriz de informação de Fisher numa simples observação da tabela t. Este resultado pode ser encarado como um caso particular do resultado geral de Bemis and Bhapkar (1982) para a família exponencial multivariada de distribuições.

4.3.3 Testes de ajustamento de vários modelos

A presença de observações incompletas torna possível testar o ajuste do modelo linear básico que liga parametricamente as T distribuições multinomiais no contexto do modelo "saturado". Em termos da especificação em equações livres, esse modelo estritamente linear é formulado por

$$H_0 : \underline{f}(\underline{\theta}_*) = \underline{X} \bar{\theta} \quad (4.24)$$

onde $\underline{\theta}_* = (\theta_1', \theta_2', \dots, \theta_T)'$ e $\underline{f}(\underline{\theta}_*) = \underline{A} \underline{\theta}_* = \underline{\bar{\theta}}_*$, com $\underline{A} = \text{diag}(A_1, A_2, \dots, A_T)$.

Como na situação de dados completos, o ajuste de H_0 pode ser testado pelas estatísticas usuais. Isto é, pela estatística do χ^2 de Pearson

$$X_P^2 = X_P^2(\hat{\theta}) = (\bar{p} - \underline{X} \hat{\theta})' \left[V_{\bar{p}}(\hat{\theta}) \right]^{-1} (\bar{p} - \underline{X} \hat{\theta}) \quad (4.25)$$

que é algebricamente idêntica à estatística $S(\hat{\theta})$ de Rao, ou pela estatística da razão das verosimilhanças de Wilks

$$G^2 = G^2(\hat{\theta}) = -2 \log \frac{L(\underline{X} \hat{\theta} | n)}{L(\bar{p} | n)} = 2 \sum_{t=1}^T N_t \sum_{j=1}^{s_t} p_{tj} \log \frac{p_{tj}}{\hat{\theta}_{tj}} \quad (4.26)$$

A matriz da forma quadrática em (4.25) é um estimador consistente de $\hat{V}_p(\hat{\theta})$ obtido substituindo $\hat{\theta}$ e $\hat{\theta}_t = (\theta_{tj})$, $t=2, \dots, T$, pelos seus estimadores de MV sob H_0 .

Estas duas estatísticas possuem ambas uma distribuição nula assintótica do χ^2 , com o número ν de graus de liberdade definido pela diferença entre o nº total de celas independentes nas T tabelas e o número de parâmetros estimados i.e., $\nu = (s-1) - (s_1-1)$. Na secção 4.4 será justificada esta afirmação sobre a distribuição assintótica, a qual não varia se os critérios X_p^2 e G^2 forem avaliados em qualquer outro estimador BAN de $\hat{\theta}$.

Sob a validade de H_0 , podemos considerar hipóteses H adicionais, as quais podem ser testadas conjuntamente com H_0 por qualquer dos critérios referidos, a avaliados no estimador de MV de $\hat{\theta}$ sob $H_0 \cap H$.

Se tem interesse considerar uma hipótese linear, digamos $H : C \hat{\theta} = Q$, onde C é uma matriz $s_1 \times (s_1-1)$ de posto $a < s_1-1$, o estimador de MV restringido a $H_0 \cap H$, $\hat{\theta}$, pode ser obtido, aproveitando os cálculos de 4.3.1, através do método dos multiplicadores de Lagrange. É fácil constatar que a maximização da função Lagrangiana, $\psi(\hat{\theta}, \alpha) = \log L(\hat{X} | \hat{\theta} | \hat{n}) - \alpha' C \hat{\theta}$, conduz ao sistema

$$\hat{\theta} = \{ I_{s_1-1} - \hat{V}(\hat{\theta}) C' [C \hat{V}(\hat{\theta}) C']^{-1} C \} \hat{V}(\hat{\theta}) \sum_{t=1}^T \bar{z}_t \left[\hat{V}_{p_t}(\hat{\theta}_t) \right]^{-1} \bar{p}_t \quad (4.27)$$

O uso deste método de maximização restringida (usado, e.g, por Hocking and Oxspring [1974]) pode ser evitado se adoptarmos uma formulação em equações livres para H . Neste sentido, H_0 e H podem ser expressos pelo modelo linear reduzido

$$H_0 \cap H : f(\theta_{**}) = X U \gamma \quad (4.28)$$

onde U é uma matriz $(s_1-1) \times (s_1-1-a)$ de posto completo ortogonal a C e γ é um vector de s_1-1-a parâmetros. A substituição de $\hat{\theta}$ por $U \gamma$ na função de verosimilhança e a repetição da técnica usada em 4.3.1, permitem-nos obter o estimador de MV de γ e, consequentemente, os valores preditos para $\hat{\theta}$ e θ_{**} .

Quando os dados se reportam a uma classificação cruzada segundo várias variáveis categorizadas, é de especial importância a consideração de modelos log-lineares para descrever a estrutura da distribuição conjunta dessas variáveis. Adoptando a formulação tradicional de um modelo log-linear, teremos em notação matricial

$$H : \log \hat{\theta} = \begin{pmatrix} 1 \\ \sim_{s_1} \end{pmatrix} X_I \begin{pmatrix} \eta \\ \beta \end{pmatrix}$$

onde η é uma média geral, β o vector $(v \times 1)$ dos parâmetros relevantes e X_I é uma matriz $s_1 \times v$ de especificação do modelo satisfazendo a condição de identifica-

bilidade para β , i.e.

$$R\left(\begin{bmatrix} 1 & X_I \\ \sim s_1 & \sim I \end{bmatrix}\right) = 1 + R(X_I) = 1 + v$$

com $R(\)$ designando o posto. Deste modo, θ_* fica definido pelo modelo não linear, que incorpora já as restrições naturais

$$H_0 \cap H : \theta_* = X_* \theta(\beta) = X_* \exp(X_I \beta) / \left[\begin{matrix} 1' \\ \sim s_1 \end{matrix} \exp(X_I \beta) \right] \quad (4.29)$$

onde $X_* = (Z_1 \ Z_2 \ \dots \ Z_T)'$. Tomando em consideração que

$$\frac{\partial \log L_t}{\partial \theta_t} = D_{\theta_t}^{-1} N_t ; \quad \frac{\partial \theta}{\partial \beta'} = \{ D_{\theta}(\beta) - \theta(\beta) [\theta(\beta)]' \} X_I$$

é fácil mostrar que a função "score" pode ser definida por

$$\frac{\partial \log L}{\partial \beta} = X_I' \left[N_1 + \sum_{t=2}^T D_{\theta}(\beta) Z_t D_{Z_t \theta}^{-1}(\beta) N_t \right] - X_I' \mu(\beta) \quad (4.30)$$

onde $\mu(\beta) = N \theta(\beta)$ representa o vector das frequências esperadas. Esta expressão implica que as equações de verosimilhança são idênticas àquelas que se obteriam se todos os N elementos fossem completamente categorizados, excepto no ponto em que o correspondente vector x das frequências observadas hipotéticas é substituído pela estimativa do seu valor esperado condicional aos dados realmente observados. Ou seja, x é substituído pelo numerador do 2º membro de (4.2). Assim, as v combinações lineares do estimador de MV de $E[x | n, \theta(\beta)]$ igualam as mesmas combinações lineares de $\mu(\hat{\beta})$. Para o caso especial de ocorrência de uma única tabela parcialmente categorizada, Haberman (1974) obteve equações de verosimilhança correspondentes, através de métodos de projecção ortogonal, tendo ainda discutido propriedades das suas soluções.

Quando H traduz um modelo log-linear hierárquico, (4.30) implica a formulação matricial das equações de verosimilhança registadas em Fuchs (1982). Em termos da aplicação do algoritmo EM, (4.30) traduz ainda que o passo E deriva daquele relativo ao modelo saturado (vide (4.2)), somando sobre os índices de modo a obter o valor ajustado para a estatística suficiente do modelo log-linear em causa. O passo M pode requerer solução iterativa, podendo, então, ser executado por meio de procedimentos iterativos delineados para dados completos, como o algoritmo IPF (vide, e.g., Bishop et al. [1975, cap.3]). A própria ocorrência de uma configuração em partições encaixadas não conduz necessariamente a estimativas directas, como sucede com o modelo "saturado". Por outro lado, o facto de um determinado modelo hierárquico reduzido apresentar estimativas em forma fechada no problema de dados completos, pode deixar de se verificar com a presença de dados incompletos. O exemplo 2 da secção 4.5 ilustra claramente este ponto. A existência de estimativas de θ com expressão explícita depende, não

são do modelo log-linear em questão, mas também do padrão de censura observado.

Independentemente do tipo (hierárquico ou não) do modelo log-linear, as equações (4.30) podem ser resolvidas recorrendo a métodos gerais de otimização. Para a aplicação do usual método de Newton-Raphson, torna-se necessário derivar a expressão para a matriz Hessiana de $\log L$. Após alguns cálculos de diferenciação matricial a partir de (4.30), pode constatar-se que tal matriz pode ser definida pela expressão

$$\underline{H}(\underline{\beta}) \equiv \frac{\partial^2 \log L}{\partial \underline{\beta} \partial \underline{\beta}'} = \underline{X}'_{-I} \left\{ -N\underline{I}_{s_1} + \sum_{t=2}^T \left[\underline{D}_{\underline{a}_t(\underline{\beta})} - \underline{D}_{\underline{b}_t(\underline{\beta})} \underline{Z}_t \underline{Z}'_t \right] \right\} \left\{ \underline{D}_{\underline{\theta}(\underline{\beta})}^{-\underline{\theta}(\underline{\beta})} [\underline{\theta}(\underline{\beta})]' \right\}' \underline{X}_{-I} \quad (4.31)$$

onde

$$\underline{a}_t(\underline{\beta}) = \underline{Z}_t \underline{D}_{\underline{Z}_t \underline{Z}'_t}^{-1} \underline{\theta}(\underline{\beta}) N_t \quad ; \quad \underline{b}_t(\underline{\beta}) = \underline{D}_{\underline{\theta}(\underline{\beta})} \underline{Z}_t \underline{D}_{\underline{Z}_t \underline{Z}'_t}^{-2} \underline{\theta}(\underline{\beta}) N_t, \quad t=2, \dots, T$$

Consequentemente, a matriz de informação de Fisher condicional cuja inversa define a matriz de covariância assintótica do estimador de MV de $\underline{\beta}$, $\hat{\underline{\beta}}$, é

$$\underline{F}(\underline{\beta}) = \underline{X}'_{-I} \left\{ N\underline{I}_{s_1} + \sum_{t=2}^T \left[N_t \underline{D}_{\underline{c}_t(\underline{\beta})} \underline{Z}_t \underline{Z}'_t \right] \right\} \left\{ \underline{D}_{\underline{\theta}(\underline{\beta})}^{-\underline{\theta}(\underline{\beta})} [\underline{\theta}(\underline{\beta})]' \right\}' \underline{X}_{-I} \quad (4.32)$$

onde

$$\underline{c}_t(\underline{\beta}) = \underline{D}_{\underline{\theta}(\underline{\beta})} \underline{Z}_t \underline{D}_{\underline{Z}_t \underline{Z}'_t}^{-1} \underline{1}_{s_t}, \quad t=2, \dots, T$$

Aplicando o método- δ baseado numa expansão Tayloriana linear de $\underline{\theta}(\hat{\underline{\beta}})$, obtemos a seguinte expressão para a matriz de covariância assintótica de $\hat{\underline{\theta}} = \underline{\theta}(\hat{\underline{\beta}})$

$$\underline{V}_{\hat{\underline{\theta}}}(\underline{\beta}) = \left(\underline{D}_{\underline{\theta}(\underline{\beta})}^{-\underline{\theta}(\underline{\beta})} [\underline{\theta}(\underline{\beta})]' \right)' \underline{X}'_{-I} \left[\underline{F}(\underline{\beta}) \right]^{-1} \underline{X}_{-I} \left(\underline{D}_{\underline{\theta}(\underline{\beta})}^{-\underline{\theta}(\underline{\beta})} [\underline{\theta}(\underline{\beta})]' \right) \quad (4.33)$$

A análise de (4.31), (4.32) e (4.33) permite também concluir que elas constituem uma generalização das correspondentes expressões para o caso usual de dados completos.

Como referimos anteriormente, os modelos reduzidos $H_0 \cap H$ podem ser ajustados por (4.25) ou (4.26), com $\hat{\underline{\theta}}$ designando o estimador de MV de $\underline{\theta}$ sob os respectivos modelos. O número de graus de liberdade é $v = (s-T) - (s_1 - 1 - a)$ para o caso do modelo (4.28), e $v = s - T - v$ para o modelo (4.29). O teste de ajustamento para o modelo H dado o modelo H_0 pode ser baseado na estatística da razão das verossimilhanças condicional

$$G^2(\hat{\underline{\theta}} | \hat{\underline{\theta}}) = G^2(\hat{\underline{\theta}}) - G^2(\hat{\underline{\theta}}) = 2 \sum_{t=1}^T N_t \sum_{j=1}^{s_t} p_{tj} \log \frac{\hat{\theta}_{tj}}{\hat{\theta}_{tj}} \quad (4.34)$$

onde $\hat{\underline{\theta}}$ (resp. $\hat{\underline{\theta}}$) representa o estimador de MV de $\underline{\theta}$ sob $H_0 \cap H (H_0)$. Para grandes valores de $\{N_t\}$ pode mostrar-se que a distribuição nula desta estatística é apro

ximadamente $\chi^2_{(v)}$, onde $v=a$ para o modelo (4.28), e $v=s_1-1-v$ para o modelo (4.29), um resultado análogo àquele obtido para o problema padrão de uma tabela completa.

4.3.4. Alguns comentários adicionais

O esqueleto conceptual apresentado no capítulo 3 apoiou-se num esquema de amostragem aleatória simples das N unidades experimentais a classificar. Contudo, os dados são muitas vezes colhidos, como em estudos longitudinais, por um processo conceptualmente equivalente à amostragem aleatória estratificada (vide Koch et al. [1972] e Woolson and Clarke [1984] para alguns exemplos). A extensão da análise feita ao caso de um esquema de amostragem multinomial-produto subjacente não oferece grandes dificuldades se exceptuarmos uma notação mais intrincada.

Quando os dados se reportam à classificação cruzada por várias variáveis não é, às vezes, evidente se todas as variáveis são do tipo "resposta", ou se algumas delas, geralmente referidas como factores, apenas servem para definir as subpopulações de cada uma das quais é aleatoriamente escolhida uma amostra. Em consonância, saber se as inferências sobre os parâmetros correspondentes são imutáveis, sob os modelos probabilísticos multinomial/multinomial-produto subjacentes, é uma questão de interesse para o analista de dados que, não raramente, desconhece o processo de recolha dos dados.

Devemos notar, desde já, que a hipótese básica de ignorabilidade do mecanismo de registo com uma distribuição conjunta subjacente de q multinomiais independentes, dá origem a uma verosimilhança relevante, L_0 , que é caracterizada por um produto de qT multinomiais, se assumirmos que o número de tabelas suplementares é o mesmo para todas as subpopulações. Definamos $\underline{\theta}_* = (\underline{\theta}'_2 \dots \underline{\theta}'_T)'$, $\underline{\theta} = (\underline{\theta}^{(1)'} \underline{\theta}^{(2)'} \dots \underline{\theta}^{(q)'})'$ e $\underline{\theta}_t = (\underline{\theta}_t^{(1)'} \underline{\theta}_t^{(2)'} \dots \underline{\theta}_t^{(q)'})'$, $t=2, \dots, T$ onde $\underline{\theta}_t^{(j)}$ (resp. $\underline{\theta}^{(j)}$), $j=1, \dots, q$ é um vector $s_t^{(j)} \times 1$ ($s_1^{(j)} \times 1$) de probabilidades conjuntas subjacentes relativas ao estrato j pertinentes à tabela t (completa). Seja $L(\underline{\theta}_*)$ a função de verosimilhança dada em (3.30) e $L_0(\underline{\theta}_*)$ a função de verosimilhança atrás referida, com as respectivas probabilidades condicionais expressas como função de $\underline{\theta}_*$. Ambas são consideradas sob as restrições naturais, as restrições lineares H_0 e as restrições log-lineares usuais definindo hipóteses de não interacção.

Do ponto de vista das inferências por MV, a questão reside então em comparar as estimativas de MV de $\underline{\theta}_*$ baseadas em $L(\underline{\theta}_*)$ e em $L_0(\underline{\theta}_*)$. Esta formulação permite-nos compreender que a questão é uma extensão do problema originalmente considerado por Birch (1963). Aplicando uma forma generalizada do teorema 13.4.1 em Bishop et al. (1975, p.447), podemos concluir que os dois conjuntos de estima

tivas são idênticos desde que as estimativas calculadas nesta secção satisficam as condições $\sum_{j=1}^q \theta^{(j)} = N^{(j)}/N$, $j=1, \dots, q$, onde $N^{(j)}$ é o tamanho da j -ésima amostra e $N = \sum_{j=1}^q N^{(j)}$.

Sob o modelo estrutural "saturado", a expressão (4.1) implica que essas condições se verificam. Esta mesma conclusão pode ser extraída de análise da expressão de $\hat{\theta}$ obtida de (4.30), desde que o modelo log-linear em questão inclua o termo u (para usar a terminologia de Bishop et al. [1975], por exemplo) relativo à configuração considerada como fixa no esquema multinomial-produto subjacente, como no caso do problema de dados completos.

Em contrapartida, a presença do modelo estritamente linear H_0 torna problemática a satisfação das condições esboçadas em Woolson and Brier (1981) para a equivalência dos resultados inferenciais com dados completos, para os dois modelos probabilísticos, quando é usada a metodologia por MQP.

Finalmente, não queremos terminar esta secção sem realçar que a aplicação dos métodos de MV descritos não exige a ocorrência de uma tabela completamente categorizada, desde que cada elemento de θ seja identificável em pelo menos uma das tabelas parcialmente categorizadas. Este requisito poderá ser preenchido com uma única tabela incompleta, sob algum modelo reduzido.

4.4. Métodos de mínimos quadrados ponderados

4.4.1. Metodologia do χ^2 modificado em uma fase

Frisamos já que a suposição de uma distribuição multinomial subjacente dá origem a uma distribuição para as frequências observadas que, condicionalmente ao padrão de registo observado é caracterizada por um produto de T multinomiais independentes com parâmetros N_t e $\theta_t = \sum_{i=1}^q \theta_i$, $t=1, \dots, T$, quando o mecanismo de registo é modelado por M_2 . Nesta base, a grande diferença na aplicação da metodologia GSK relativamente à situação padrão jaz no facto de que as categorias de classificação, bem como o seu número, variam de uma distribuição para outra, conforme foi assinalado por Koch et al. (1972).

Em virtude de tal modelo probabilístico, o estimador de MV irrestrito de θ , definido por $\hat{p} = (p_1' \dots p_T')'$, onde $p_t = N_t/N$, $t=1, \dots, T$, satisfaz o modelo linear generalizado, $\hat{p} = \theta + \varepsilon$, onde o vector dos erros tem média nula e matriz de covariância $V_{\hat{p}}(\theta) = \text{diag}(V_{p_1}(\theta), \dots, V_{p_T}(\theta))$ com $V_{p_t}(\theta) = N_t^{-1} (D_{\theta_t}^{-1} - \theta_t \theta_t')$.

O argumento usado em 4.3.2 pode ser aplicado à função do χ^2 modificado (ou de Neyman) definida por

$$X_N^2(\theta) = \sum_{t=1}^T X_{N_t}^2(\theta) \quad , \quad X_{N_t}^2(\theta) = N_t \cdot (p_t - \theta_t)' D_{\theta_t}^{-1} (p_t - \theta_t) \quad (4.35)$$

onde cada elemento de cada \underline{p}_t é suposto positivo. Incorporando as restrições naturais, esta função pode ser reexpressa por

$$\begin{aligned} X_N^2(\underline{\theta}) &= \sum_{t=1}^T (\underline{p}_t - \underline{\theta}_t)' \left[\frac{V_{\underline{p}_t}(\underline{\theta}_t)}{\underline{p}_t} \right]^{-1} (\underline{p}_t - \underline{\theta}_t) \\ &= (\underline{p} - \underline{\theta}_*)' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} (\underline{p} - \underline{\theta}_*) \end{aligned} \quad (4.36)$$

onde $\frac{V_{\underline{p}}(\underline{p})}{\underline{p}}$ é o estimador consistente de $\frac{V_{\underline{p}_t}(\underline{\theta}_t)}{\underline{p}_t}$ obtido por substituição de $\underline{\theta}_t$ por \underline{p}_t , $t=1, \dots, T$. A existência da matriz inversa desse estimador é garantida pela suposição feita sobre \underline{p} (quando tal não se verifica, veja-se Koch et al. [1972, sec.3]).

Vimos já que a comparação entre as T multinomiais é traduzida pela modelo estritamente linear $H_0 : E(\underline{p} | \{N_t\}) \equiv \underline{\theta} = \underline{X} \underline{\theta}$. A comparação entre (4.35) e (4.21) sob H_0 torna assim evidente que o estimador do χ^2 modificado mínimo (abreviadamente QNM) de $\underline{\theta}$ pode ser encarado de um modo idêntico ao estimador do QPM, excepto no aspecto em que o critério dos mínimos quadrados generalizados é modificado pela substituição de $\frac{V_{\underline{p}_t}(\underline{\theta}_t)}{\underline{p}_t}$ pelo estimador consistente supra-citado.

Deste modo, o ajustamento de H_0 pode ser testado pela estatística do χ^2 modificado mínimo (ou estatística residual dos mínimos quadrados ponderados)

$$X_N^2 = X_N^2(\underline{\theta}) = (\underline{p} - \underline{X} \underline{\theta})' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} (\underline{p} - \underline{X} \underline{\theta}) \quad (4.37)$$

onde $\underline{\theta}$ é o estimador do QNM (ou dos MQP) de $\underline{\theta}$, dado explicitamente em forma fechada por

$$\underline{\theta} = \left\{ \underline{X}' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} \underline{X} \right\}^{-1} \underline{X}' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} \underline{p} \quad (4.38)$$

A formulação equivalente de H_0 em termos de restrições é $C_0 \underline{\theta}_* = \underline{0}_w$ onde $w = (s-T) - (s_1-1)$ e C_0 é uma matriz $w \times (s-T)$ de posto completo ortogonal a \underline{X} . De acordo com Bhapkar (1966), para testar essas restrições, a estatística X_N^2 é algebricamente equivalente à estatística de Wald

$$Q_{C_0} \equiv Q_{C_0}(\underline{p}) = (C_0 \underline{p})' \left[C_0 \frac{V_{\underline{p}}(\underline{p})}{\underline{p}} C_0' \right]^{-1} C_0 \underline{p} \quad (4.39)$$

Para o evidenciarmos, basta pôr (4.38) em (4.37) de modo a exprimir X_N^2 como a forma quadrática

$$X_N^2 = \underline{p}' \underline{S}_X \underline{p}, \quad \underline{S}_X = \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} \left\{ \underline{I}_{s-T} - \underline{X} \left\{ \underline{X}' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} \underline{X} \right\}^{-1} \underline{X}' \left[\frac{V_{\underline{p}}(\underline{p})}{\underline{p}} \right]^{-1} \right\} \quad (4.40)$$

e aplicar o lema de identidade matricial de Koch (1969) para concluir que

$$S_x \equiv C'_0 \left[\begin{array}{ccc} C & V_{\tilde{p}}(\bar{p}) & C' \\ \hline C_0 & \tilde{p} & C'_0 \end{array} \right]^{-1} C_0.$$

É bem conhecido que $\sqrt{N_{t.}} (\bar{p}_{t.} - \bar{\theta}_{t.}) \xrightarrow{d} N_{S_{t-1}}(0, \Sigma_t)$, com $\Sigma_t = D_{\tilde{p}}^{-1} \bar{\theta}_{t.} \bar{\theta}'_{t.}$, $t=1, \dots, T$,

à medida que $N_{t.} \rightarrow \infty$ pelo teorema do limite central. Isto implica, denotando $\Sigma = \text{diag}(\Sigma_1 \dots \Sigma_T)$ e $\xi = (\sqrt{N_1}, J'_1, \dots, \sqrt{N_T}, J'_T)'$, com $J_{t.} = 1_{S_{t-1}}$, que

$D_{\tilde{p}} (\bar{p} - \bar{\theta}_*) \xrightarrow{d} N_{S-T}(0, \Sigma)$ quando $\{N_{t.}\} \rightarrow \infty$. Assim, para grandes valores de $\{N_{t.}\}$, \bar{p} tem uma distribuição normal multivariada aproximada com vector média $\bar{\theta}$ e matriz de covariância $V_{\tilde{p}}(\bar{\theta})$, sob H_0 . Segue-se então do teorema de Slutsky que Q_{C_0} é assintoticamente distribuída como um $\chi^2_{(w)}$ sob H_0 .

Quando $N \rightarrow \infty$ de tal modo que cada $N_{t.}/N$ converge para r_t , $0 < r_t < 1$, $t=1, \dots, T$, o estimador do QNM de $\bar{\theta}$ comporta-se também como um estimador BAN (Neyman 1949). Para vermos como $\hat{\bar{\theta}}$ é assintoticamente equivalente ao estimador de MV $\hat{\bar{\theta}}$, usemos a expansão de Taylor de 1ª ordem em $\bar{\theta}_*$ para (4.38):

$$\hat{\bar{\theta}} = \bar{\theta} + \left(\frac{\partial \bar{\theta}}{\partial \bar{p}'} \right)_{\bar{p}=\bar{\theta}_*} (\bar{p} - \bar{\theta}_*) + o_p(\|\bar{p} - \bar{\theta}_*\|) \quad \text{quando } \bar{p} \rightarrow \bar{\theta}_* \quad (4.41)$$

Diferenciando ambos os membros das equações normais

$$\tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{p})}{\tilde{p}} \right]^{-1} \tilde{X} \tilde{\theta} = \tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{p})}{\tilde{p}} \right]^{-1} \bar{p}$$

em ordem a cada elemento de \bar{p} , digamos a_j , e avaliando os resultados em $\bar{p} = \bar{\theta}_*$, obtemos

$$\tilde{X}' P_j(\bar{\theta}_*) \tilde{X} \tilde{\theta} + \left(\tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{\theta})}{\tilde{p}} \right]^{-1} \tilde{X} \right) \left(\frac{\partial \bar{\theta}}{\partial a_j} \right)_{\bar{p}=\bar{\theta}_*} = \tilde{X}' P_j(\bar{\theta}_*) \bar{\theta}_* + \tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{\theta})}{\tilde{p}} \right]^{-1} \frac{\partial \bar{p}}{\partial a_j}$$

onde $P_j(\bar{\theta}_*) = \left(\frac{\partial}{\partial a_j} \left[\frac{V_{\tilde{p}}(\bar{p})}{\tilde{p}} \right]^{-1} \right)_{\bar{p}=\bar{\theta}_*}$. Assim, sob H_0

$$\left(\frac{\partial \bar{\theta}}{\partial \bar{p}'} \right)_{\bar{p}=\bar{\theta}_*} = \left\{ \tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{\theta})}{\tilde{p}} \right]^{-1} \tilde{X} \right\}^{-1} \tilde{X}' \left[\frac{V_{\tilde{p}}(\bar{\theta})}{\tilde{p}} \right]^{-1} \quad (4.42)$$

Devido às considerações anteriores sobre o comportamento assintótico de $D_{\tilde{p}} (\bar{p} - \bar{\theta}_*)$, o que implica que o 3º termo do 2º membro de (4.41) é um vector com elementos da ordem $o_p(N_{t.}^{-1/2})$, podemos afirmar que $\hat{\bar{\theta}}$ é assintoticamente equivalente à sua aproximação em série de Taylor linear. Tendo em conta (4.41) e (4.42), podemos portanto concluir que, sob H_0 , $\hat{\bar{\theta}}$ tem aproximadamente uma distribuição normal com vector média $\bar{\theta}$ e matriz de covariância $\hat{V}(\bar{\theta}) = [F(\bar{\theta})]^{-1}$ (recorde

-se (4.9)), como pretendíamos demonstrar.

Consideremos agora \underline{C} uma matriz $a \times (s_1 - 1)$ de posto completo $a < s_1 - 1$. O vector $\underline{C} \bar{\theta}$ é então aproximadamente distribuído como $N_a(\underline{C} \bar{\theta}, \underline{C} \hat{V}(\bar{\theta}) \underline{C}')$ e a sua matriz de covariância consistentemente estimada por $\underline{C} \hat{V}(\bar{p}) \underline{C}'$, onde

$$\hat{V}(\bar{p}) = \left[\underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \underline{X} \right]^{-1} \quad (4.43)$$

Desta forma, admitindo que H_0 fornece uma descrição satisfatória da variação de $\bar{\theta}$, a hipótese linear $H: \underline{C} \bar{\theta} = \underline{0}$ pode ser testada com a estatística de Wald

$$Q_c = (\underline{C} \bar{\theta})' \left[\underline{C} \hat{V}(\bar{p}) \underline{C}' \right]^{-1} \underline{C} \bar{\theta} \quad (4.44)$$

cuja distribuição nula assintótica é $\chi^2_{(a)}$.

O ajuste do modelo linear reduzido $H_0 \Omega H$ definido em (4.28) pode ser testado, por analogia com (4.37), (4.39) e (4.40) pela estatística de Wald

$$Q_{c_0, c} = \bar{p}' S_{x, u} \bar{p} \quad (4.45)$$

onde $S_{x, u} = \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \left\{ \underline{I}_{s-T} - \underline{X} \underline{U}' \left\{ \underline{U}' \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \underline{X} \underline{U} \right\}^{-1} \underline{U}' \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \right\}$.

Os estimadores dos MQP do parâmetro $\underline{\gamma}$ e da sua matriz de covariância são respectivamente dados por

$$\begin{aligned} \tilde{\underline{\gamma}} &= \left[\underline{U}' \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \underline{X} \underline{U} \right]^{-1} \underline{U}' \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \bar{p} \\ \underline{V}_{\tilde{\underline{\gamma}}}(\bar{p}) &= \left[\underline{U}' \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \underline{X} \underline{U} \right]^{-1} \end{aligned} \quad (4.46)$$

A diferença entre as estatísticas de ajustamento do modelo linear reduzido $H_0 \Omega H$ e do modelo linear original H_0 é então expressa por

$$Q_{c_0, c} - Q_c = \bar{p}' (S_{x, u} - S_x) \bar{p} \quad (4.47)$$

Notando que em (4.44), $\underline{C} \bar{\theta} = \underline{C} \left[\underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \underline{X} \right]^{-1} \underline{X}' \left[\underline{V}_{\underline{p}}(\bar{p}) \right]^{-1} \bar{p}$ e que $\underline{C} \underline{U} = \underline{0}_{(a, s_1 - 1 - a)}$, a aplicação de novo do lema de identidade matricial de Koch (1969) permite-nos mostrar que (4.47) é equivalente a (4.44). A estatística Q_c é assim interpretável como estatística de Wald condicional para testar as restrições adicionais que o modelo $H_0 \Omega H$ contempla dado que se verificam aquelas implicadas pelo modelo H . Consequentemente, a estatística do χ^2 modificado mínimo admite a mesma partição que a estatística da razão das verosimilhanças na presença de uma sequência hierárquica de modelos.

4.4.2. Metodologias dos mínimos quadrados ponderados em duas fases e híbridas

Para testar a hipótese H anterior, podemos usar alternativamente, no espírito do método de regressão funcional assintótica, MRFA em abreviatura (vide e.g., Imrey et al. [1981,1982]), uma estatística de Wald baseada no estimador de MV de $\hat{\theta}$, concretamente

$$Q_C^\dagger = (\underline{C} \hat{\theta})' (\underline{C} \hat{V}(\hat{\theta}) \underline{C}')^{-1} \underline{C} \hat{\theta} \quad (4.48)$$

onde $\hat{V}(\hat{\theta}) = (\underline{X}' [\underline{V}_p(\hat{\theta})]^{-1} \underline{X})^{-1}$ é o estimador consistente da matriz de covariância assintótica de $\hat{\theta}$, obtida desta substituindo $\bar{\theta}_*$ pelos valores preditos $\hat{\theta}_* = \underline{X} \hat{\theta}$. Esta estatística é assim assintoticamente equivalente a Q_C .

Após a obtenção do estimador $\hat{\theta}$ sob H_0 , a aplicação dos instrumentos da metodologia dos MQP ao modelo $H: \bar{\theta} = \underline{U} \underline{\gamma}$, onde $\bar{\theta}$ iguala o valor esperado assintótico de $\hat{\theta}$, permite-nos obter as seguintes expressões para o estimador do MRFA de $\underline{\gamma}$ e sua matriz de covariância estimada

$$\underline{\gamma}^\dagger = (\underline{U}' \underline{X}' [\underline{V}_p(\hat{\theta})]^{-1} \underline{X} \underline{U})^{-1} \underline{U}' \underline{X}' [\underline{V}_p(\hat{\theta})]^{-1} \hat{\theta}_* \quad (4.49)$$

$$\underline{V}_{\underline{\gamma}^\dagger}(\hat{\theta}_*) = (\underline{U}' \underline{X}' [\underline{V}_p(\hat{\theta})]^{-1} \underline{X} \underline{U})^{-1}$$

Estas expressões são formalmente idênticas àquelas derivadas pela metodologia dos MQP em um estágio dadas em (4.46). A única diferença está no facto dos valores preditos de $\bar{\theta}_*$ por MV jogarem o papel das proporções observadas \bar{p} . Seguindo Imrey et al. (1981), podemos dizer que os estimadores do MRFA (4.49) são obtidos aplicando num segundo estágio a metodologia padrão dos MQP à "tabela de pseudo-contingência" envolvendo as proporções preditas $\hat{\theta}_*$ sob H_0 . O estimador $\underline{\gamma}^\dagger$ pode pois ser interpretado como o estimador do QNM de $\underline{\gamma}$, encarando o critério do χ^2 de Neyman aplicado a essa "tabela de pseudo-contingência". Como anteriormente, podemos facilmente provar que a estatística do χ^2 modificado correspondente para o ajuste do modelo H , $X_N^{2\dagger} = (\hat{\theta}_* - \underline{X} \underline{U} \underline{\gamma}^\dagger)' [\underline{V}_p(\hat{\theta})]^{-1} (\hat{\theta}_* - \underline{X} \underline{U} \underline{\gamma}^\dagger)$ é idêntica a Q_C^\dagger .

Quando se está interessado numa função não linear de θ , como no caso da formulação log-linear de hipóteses de não-interacção em tabelas de contingência multidimensionais, Koch et al. (1972) propõem a aplicação da abordagem GSK numa segunda fase às estimativas ajustadas que reflectem todos os dados disponíveis

$$\tilde{\theta} = \begin{bmatrix} \underline{0} \\ (s_1 - 1) \\ 1 \end{bmatrix} + \begin{bmatrix} \underline{I} \\ s_1 - 1 \\ -\underline{J}' \\ 1 \end{bmatrix} \tilde{\theta} \quad (4.50)$$

onde $\tilde{\theta}$ é dado em (4.38). Deste modo, podemos visualizar este procedimento como uma aplicação da metodologia de regressão funcional assintótica, diferindo da anterior no sentido em que a metodologia por MQP é usada em ambas as fases. Naturalmente que a segunda fase pode ser aplicada, numa abordagem híbrida, ao estimador de MV de θ .

Com o fim de descrever a teoria estatística subjacente a esta metodologia dos MQP em duas fases, seja $G(\theta)$ o vector função de interesse com dimensão $u \leq s_1 - 1$ para o qual é considerado o modelo linear

$$G(\theta) = X_G \alpha \quad (4.51)$$

onde a matriz X_G de especificação do modelo é $u \times v$ com posto completo $v \leq u$ e α é o correspondente vector paramétrico desconhecido. Designemos por $G_*(\theta)$ o vector ($u \times 1$) de funções resultante de $G(\theta)$ por substituição de θ_{s_1} por $1 - J_1' \tilde{\theta}$ e suponhamos que os elementos de $G_*(\cdot)$ têm derivadas parciais contínuas até ordem 2 num subconjunto aberto do simplex probabilístico de \mathbb{R}^{s_1-1} contendo $\tilde{\theta}$. Admitindo ainda que $G_*(\cdot)$ é tal que $G_* = G(\tilde{\theta})$ pode ser aproximado pela sua expansão de Taylor linear sob as condições que garantem a normalidade assintótica de $\tilde{\theta}$, a aplicação do método- δ assegura que G_* tem aproximadamente uma distribuição normal u -variada com vector média $G_*(\tilde{\theta})$ e matriz de covariância $V_{G_*}(\tilde{\theta}) = L \hat{V}(\tilde{\theta}) L'$, onde $L \equiv L(\tilde{\theta}) = \left(\frac{\partial G_*}{\partial b'} \right)_{b=\tilde{\theta}}$. Assume-se ainda que $V_{G_*}(\tilde{\theta})$ é não singular.

Tomando (4.50) em consideração, o que implica que a matriz de covariância assintótica de $\tilde{\theta}$ seja dada por $\hat{V}_+(\tilde{\theta}) = (I_{s_1-1} - J_1)' \hat{V}(\tilde{\theta}) (I_{s_1-1} - J_1)$, pode facilmente mostrar-se que $V_{G_*}(\tilde{\theta}) = H \hat{V}_+(\tilde{\theta}) H' = V_G(\tilde{\theta})$, onde $G = G(\tilde{\theta})$ e $H \equiv H(\tilde{\theta}) = \left(\frac{\partial G}{\partial a'} \right)_{a=\tilde{\theta}}$. Denotemos por V_G o estimador consistente de $V_G(\tilde{\theta})$ definido por

$$V_G = H(\tilde{\theta}) \hat{V}_+(\tilde{\theta}) [H(\tilde{\theta})]' \quad (4.52)$$

onde $\hat{V}_+(\tilde{\theta}) = (I_{s_1-1} - J_1)' \hat{V}(\tilde{\theta}) (I_{s_1-1} - J_1)$ e $\hat{V}(\tilde{\theta})$ é definido por (4.43).

O ajustamento do modelo funcional linear (4.51) pode então ser testado pela estatística residual dos MQP

$$Q_{X_G} = (G - X_G \tilde{\alpha})' V_G^{-1} (G - X_G \tilde{\alpha}) \quad (4.53)$$

cuja distribuição nula assintótica é $\chi^2_{(u-v)}$. O estimador dos MQP de α

$$\tilde{\alpha} = (X_G' V_G^{-1} X_G)^{-1} X_G' V_G^{-1} G \quad (4.54)$$

é um estimador BAN (Bemis and Bhapkar [1982a, 1982b]) com matriz de covariância

assintótica estimada

$$\tilde{V}_{\tilde{\alpha}} = (\tilde{X}'_G \tilde{V}_G^{-1} \tilde{X}_G)^{-1} \quad (4.55)$$

De novo, sob a validade de (4.51) podemos testar hipóteses do tipo $E_{\tilde{\alpha}} = 0_{(c)}$, onde E é uma matriz $c \times v$ de posto completo $c < v$, por meio da estatística de Wald

$$Q_E = (E \tilde{\alpha})' (E \tilde{V}_{\tilde{\alpha}} E')^{-1} E \tilde{\alpha} \quad (4.56)$$

assintoticamente distribuída como $\chi^2_{(c)}$ sob a hipótese nula.

Em particular, consideremos que o objectivo se prende com a procura de um modelo log-linear reduzido, $\log \tilde{\theta} = (\mathbb{1}_{s_1} \ X'_I) (\eta \ \beta')'$, que descreva satisfatoriamente a estrutura da "tabela de pseudo-contingência" $\tilde{\theta}$. Definamos, à luz de Grizzle and Williams (1972), $G(\tilde{\theta}) = \tilde{X}'_S \log \tilde{\theta}$, onde \tilde{X}_S é a sub-matriz $s_1 \times (s_1 - 1)$ relativa aos parâmetros relevantes da matriz de especificação do modelo log-linear saturado. A fixação do modelo $\tilde{X}_I (s_1 \times v)$ implica então, pelo facto de $\tilde{X}'_S \mathbb{1}_{s_1} = 0_{(s_1 - 1)}$, que $G(\tilde{\theta}) = \tilde{X}'_G \beta$, onde \tilde{X}_G é a matriz $(s_1 - 1) \times v$ $\tilde{X}'_S \tilde{X}_I$. Reciprocamente, o modelo $G(\tilde{\theta}) = \tilde{X}'_G \beta$ corresponde às restrições $\tilde{W}_G G(\tilde{\theta}) = 0_{(s_1 - 1 - v)}$, onde a matriz $(s_1 - 1) \times (s_1 - 1 - v)$ \tilde{W}_G é uma base do espaço nulo de \tilde{X}'_G (vide apêndice A), ou seja, corresponde a $\tilde{U}_G \log \tilde{\theta} = 0_{(s_1 - 1 - v)}$, com $\tilde{U}_G = \tilde{W}_G \tilde{X}'_S$. Designando \tilde{X}_I uma matriz base do espaço linear ortogonal a $M(\begin{bmatrix} \mathbb{1}_{s_1} & \tilde{U}'_G \end{bmatrix})$, então essas restrições equivalem ao modelo $\log \tilde{\theta} = (\mathbb{1}_{s_1} \ \tilde{X}'_I) (\eta \ \beta')'$.

Em suma, a formulação ordinária do modelo log-linear é equivalente ao modelo log-linear generalizado $\tilde{X}'_S \log \tilde{\theta} = \tilde{X}'_G \beta$, o que permite resolver o problema da singularidade da matriz de covariância assintótica de $\log \tilde{\theta}$. A análise assintótica destes modelos pode portanto ser baseada em (4.53) - (4.56) onde $G(\tilde{\theta}) = \tilde{X}'_S \log \tilde{\theta}$ e $\alpha = \beta$.

4.4.3. Relação com outros métodos inferenciais

Vimos anteriormente como os estimadores dos MQP de $\tilde{\theta}$ e de γ , respectivamente sob os modelos H_0 e $H_0 \cap H$, dados em (4.38) e (4.46), coincidem com os correspondentes estimadores do QNM devido à identidade entre os critérios do χ^2 modificado e dos MQP.

Pode também mostrar-se que as estatísticas do χ^2 modificado mínimo usadas para ajustar os modelos lineares relativos à 1ª fase da metodologia dos MQP são assintoticamente equivalentes às estatísticas geralmente usadas na metodologia de MV. Para o efeito, seja

onde, para usar a notação de Gokhale and Kullback (1978), $P = \{\theta_{\tilde{r}*}; B \theta_{\tilde{r}*} = d\}$, com $B = (R' K')'$, $R = \text{diag}(1'_{\tilde{s}_1}, \dots, 1'_{\tilde{s}_T})$, $K = C_0 A$, C_0 é a matriz definidora das restrições H_0 , $A = \text{diag}(A_1, \dots, A_T)$, $A_t = (I_{\tilde{s}_t-1} \ 0_{\tilde{s}_t-1})$, $t=1, \dots, T$ e $d = (1'_T \ 0'_w)$, $w = (s-T) - (s_1-1)$.

Incorporando as restrições naturais em (4.60), a minimização pelo método Lagrangiano produz

$$\hat{\theta}_{\tilde{r}*} = \frac{D_{\tilde{r}} \exp(N_{\tilde{r}*}^{-1} K' \lambda)}{\left[N^{-1} n' \exp(N_{\tilde{r}*}^{-1} K' \lambda) \right]} \quad (4.62)$$

onde $N_{\tilde{r}*} = (N_1 \ 1'_{\tilde{s}_1}, \dots, N_T \ 1'_{\tilde{s}_T})$ e λ é o vector (wx1) dos multiplicadores de Lagrange relativos ao 2º conjunto de restrições, a ser determinado de modo que $K' \hat{\theta}_{\tilde{r}*} = 0_w$. O cálculo de (4.62) pode ser feito de um modo iterativo por meio de um algoritmo que é, essencialmente, um procedimento de tipo Newton-Raphson (vide Gokhale and Kullback [1978, caps. 5 e 6] para uma descrição deste e outros algoritmos).

Segundo Ireland and Kullback (1968) é possível mostrar que $\hat{\theta}_{\tilde{r}*}$ (ou $\hat{\theta}_{\tilde{r}}$, se usarmos em (4.60) as restrições na forma $\bar{\theta}_{\tilde{r}*} = X \bar{\theta}$) é também um estimador BAN de $\theta_{\tilde{r}*}$ (resp. $\bar{\theta}$), e portanto, pode ser usado nas estatísticas de ajustamento acima referidas sem modificar a sua distribuição nula assintótica.

No âmbito desta metodologia, a hipótese H_0 é testada pela estatística de IDM

$$2 N I(\hat{\theta}_{\tilde{r}*}; p) = 2 \sum_{t=1}^T N_t \sum_{j=1}^{s_t} \hat{\theta}_{tj} \log \frac{\hat{\theta}_{tj}}{p_{tj}} \quad (4.63)$$

a qual, segundo Kullback (1959, cap.5), possui uma distribuição nula aproximada $\chi^2_{(w)}$ para grandes valores de $\{N_t\}$, constituindo assim uma estatística adicional assintoticamente equivalente àquelas já referidas.

Apesar do facto do estimador de IDM ser diferente do estimador de QNM, Gokhale and Kullback (1978) provam que a estatística do χ^2 modificado mínimo é uma aproximação quadrática da estatística de IDM. Tal aproximação, aí definida por meio da forma quadrática

$$Q = (N \ \tilde{K} \ p)' S^{22} (N \ \tilde{K} \ p) \quad (4.64)$$

onde

$$S_0 = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix} = \tilde{C} D_{\tilde{r}} \tilde{C}', \quad \tilde{C} = N B D_{\tilde{r}*}^{-1}, \quad S^{22} = (S_{22} \ -S_{21} S_{11}^{-1} S_{12})^{-1}$$

está realmente na forma da estatística de Wald Q_c em (4.39). Na realidade, tendo em conta que $\tilde{p} = A p$, $V_{\tilde{p}}(\tilde{p}) = A V_p(p) A'$ e $\tilde{K} = C_0 A$, esta estatística de Wald pode

$$G^2(\hat{\theta}_{\sim*}) = \sum_{t=1}^T N_t G_t^2(\hat{\theta}_{\sim t}), \text{ com } G_t^2(\hat{\theta}_{\sim t}) = 2 \sum_{j=1}^{s_t} p_{tj} \log \frac{p_{tj}}{\hat{\theta}_{\sim t j}} \quad (4.5)$$

a estatística da razão das verosimilhanças para testar H_0 . Seguindo o argumento de Koch and Bhapkar (1982), podemos expandir $G_t^2(\hat{\theta}_{\sim t})$ numa série de Taylor de 2ª ordem em torno de $p_{\sim t}$, para obtermos

$$G_t^2(\hat{\theta}_{\sim t}) = \sum_{j=1}^{s_t} \frac{(\hat{\theta}_{\sim t j} - p_{tj})^2}{p_{tj}} + o_p(\|\hat{\theta}_{\sim t} - p_{\sim t}\|) \quad (4.5)$$

Recordando os resultados distribucionais assintóticos para $\hat{\theta}_{\sim t}$ e $p_{\sim t}$ quando $N_t \rightarrow \infty$, o 2º termo do 2º membro de (4.58) é de uma ordem estocástica menor que N_t^{-1} e assim, por (4.35), podemos afirmar que

$$G^2(\hat{\theta}_{\sim*}) = X_N^2(\hat{\theta}_{\sim*}) + o_p(1) \quad (4.5)$$

Como $\hat{\theta}_{\sim*}$ e $\tilde{\theta}_{\sim*}$ são ambos estimadores BAN de $\bar{\theta}_{\sim*}$, segue-se que sob H_0 $X_N^2(\hat{\theta}_{\sim*}) \stackrel{a}{\approx} X_N^2(\tilde{\theta}_{\sim*}) \equiv Q_c$, onde $\stackrel{a}{\approx}$ significa "assintoticamente equivalente". Devido à equivalência assintótica entre $G^2(\hat{\theta}_{\sim*})$ e $X_P^2(\hat{\theta}_{\sim*})$ fica assegurado que as estatísticas X_N^2 , G^2 e X_P^2 são assintoticamente equivalentes independentemente dos estimadores BAN envolvidos.

Obviamente que a mesma conclusão se mantém para as estatísticas de ajuste do modelo linear $H_0 \Omega H$, i.e., $G^2(\hat{\theta}_{\sim*}) \stackrel{a}{\approx} Q_{c,c}$, onde $\hat{\theta}_{\sim*}$ é o valor predito para $\theta_{\sim*}$ calculado do estimador de MV de $\bar{\theta}_{\sim*}$ (ou de γ) sob esse modelo. Por conseguinte $G^2(\hat{\theta}_{\sim*} | \hat{\theta}_{\sim*}) \stackrel{a}{\approx} Q_c$. Deste modo, os resultados distribucionais referidos em 4.3 encontram-se plenamente justificados.

A gama de questões que têm sido encaradas no âmbito do modelo probabilístico condicional usado cai no chamado problema das restrições externas, para empregar a terminologia de Gokhale and Kullback (1978, cap.5). Portanto, elas podem ser analisadas pela abordagem da teoria da informação, que é baseada no critério de informação discriminante generalizado

$$I(\theta_{\sim*}; p) = \sum_{t=1}^T (N_t / N) I(\theta_{\sim t}; p_{\sim t}), \text{ com } I(\theta_{\sim t}; p_{\sim t}) = \sum_{j=1}^{s_t} \theta_{\sim t j} \log \frac{\theta_{\sim t j}}{p_{\sim t j}} \quad (4.6)$$

Considerando, por exemplo, a questão de verificar a consistência dos dados observados com H_0 , o 1º passo consiste na determinação do estimador de informação discriminante mínima (IDM em abreviatura) de $\theta_{\sim*}$, o qual é definido como o valor $\hat{\theta}_{\sim*}$ que minimiza (5.50) sujeito às restrições dadas. Isto é

$$I(\hat{\theta}_{\sim*}; p) = \min_{\theta_{\sim*} \in P} I(\theta_{\sim*}; p) \quad (4.6)$$

ser reescrita como

$$Q_{c_0} = (\tilde{K} \tilde{p})' (\tilde{K} \tilde{V}_{\tilde{p}}(\tilde{p}) \tilde{K}')^{-1} \tilde{K} \tilde{p} \quad (4.65)$$

Assim, basta mostrar que $\tilde{S}^{22} = N^{-2} (\tilde{K} \tilde{V}_{\tilde{p}}(\tilde{p}) \tilde{K}')^{-1}$. Das expressões definidoras das sub-matrizes de \tilde{S}_0 obtidas da consideração da partição de \tilde{B} em \tilde{R} e \tilde{K} , mostra-se facilmente que

$$\tilde{S}^{22} = N^{-2} (\tilde{U}_{22} - \tilde{U}_{21} \tilde{U}_{11}^{-1} \tilde{U}_{12})^{-1} \quad (4.66)$$

onde

$$\tilde{U}_0 \equiv \begin{bmatrix} \tilde{U}_{11} & \tilde{U}_{12} \\ \tilde{U}_{21} & \tilde{U}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{B} & \tilde{D}_{\tilde{N}^*}^{-1} \tilde{D}_{\tilde{p}} \tilde{B}' \\ & \tilde{D}_{\tilde{p}} \end{bmatrix}$$

Assim

$$\begin{aligned} \tilde{U}_{11} &= \tilde{R} \tilde{D}_{\tilde{N}^*}^{-1} \tilde{R}' = \text{diag}(1/N_1, \dots, 1/N_T); & \tilde{U}_{12} &= \tilde{U}_{21}' = \tilde{R} \tilde{D}_{\tilde{N}^*}^{-1} \tilde{D}_{\tilde{p}} \tilde{K}' = \left[\text{diag}(p_1'/N_1, \dots, p_T'/N_T) \right] \tilde{K}' \\ \tilde{U}_{22} &= \tilde{K} \tilde{D}_{\tilde{p}}^{-1} \tilde{D}_{\tilde{p}} \tilde{K}' = \tilde{K} \left[\text{diag}(D_{\tilde{p}_1}/N_1, \dots, D_{\tilde{p}_T}/N_T) \right] \tilde{K}' \end{aligned}$$

e portanto

$$\tilde{U}_{22} - \tilde{U}_{21} \tilde{U}_{11}^{-1} \tilde{U}_{12} = \tilde{K} \left[\text{diag} \left(\left[\begin{matrix} D_{\tilde{p}_1} & -p_1 p_1' \\ \tilde{p}_1 & \tilde{p}_1' \end{matrix} \right] / N_1, \dots, \left[\begin{matrix} D_{\tilde{p}_T} & -p_T p_T' \\ \tilde{p}_T & \tilde{p}_T' \end{matrix} \right] / N_T \right) \right] \tilde{K}' \quad (4.67)$$

como pretendíamos,

Apesar da identidade entre (4.65) e (4.37) ter sido já evidenciada em 4.4.1, torna-se ilustrativo descobrir a expressão explícita para o estimador do QNM de θ_* e comprovar, então, que (4.65), expresso na forma

$$Q_{c_0} = \begin{bmatrix} 0 \\ \tilde{K} \tilde{p} \end{bmatrix}' (\tilde{B} \tilde{D}_{\tilde{N}^*}^{-1} \tilde{D}_{\tilde{p}} \tilde{B}')^{-1} \begin{bmatrix} 0 \\ \tilde{K} \tilde{p} \end{bmatrix} \quad (4.68)$$

corresponde a $X_N^2(\tilde{\theta}_*)$. Para o efeito, escrevamos (4.35) como

$$X_N^2(\theta_*) = (\tilde{p} - \theta_*)' \tilde{D}_{\tilde{N}^*} \tilde{D}_{\tilde{p}}^{-1} (\tilde{p} - \theta_*) \quad (4.69)$$

Diferenciando a função Lagrangiana, $\phi(\theta_*, \tau) = X_N^2(\theta_*) - \tau' (\tilde{B} \theta_* - \tilde{d})$, em ordem a θ_* e igualando a zero, obtêm-se

$$\tilde{\theta}_* - p = -(1/2) D_{N_*}^{-1} D_p B' \tilde{\tau} \quad (4.70)$$

Como $B(p - \tilde{\theta}_*) = B p - d = (0' (K p)')'$, segue-se de (4,70) que

$$\tilde{\tau} = 2(B D_{N_*}^{-1} D_p B')^{-1} \begin{bmatrix} 0 \\ K p \end{bmatrix} \quad (4.71)$$

o que implica que o estimador do QNM de θ_* é

$$\tilde{\theta}_* = p - D_{N_*}^{-1} D_p B' (B D_{N_*}^{-1} D_p B')^{-1} \begin{bmatrix} 0 \\ K p \end{bmatrix} \quad (4.72)$$

e que a correspondente expressão de $X_N^2(\tilde{\theta}_*)$ é idêntica ao 2º membro de (4.68).

Como nota final, acrescentamos que a abordagem por MV a este tipo de problema pode ser vista no âmbito da metodologia de IDM. De facto, definamos um estimador de θ_* que minimiza o diferente critério de informação discriminante generalizado $I_0(\theta_*; p) = I(p; \theta_*)$, i.e.

$$I_0(\theta_*; p) = \sum_{t=1}^T (N_t / N) \sum_{j=1}^{s_t} p_{tj} \log \frac{p_{tj}}{\theta_{tj}} \quad (4.73)$$

para $\theta_* \in P$. Como $I_0(\cdot)$ é $2N$ vezes o critério da razão das verosimilhanças, este estimador de IDM modificado coincide com o respectivo estimador MV, e portanto, $2NI_0(\hat{\theta}_*; p) = G^2(\hat{\theta}_*)$. Parafraseando Berkson(1972), a estatística $2NI$ está para a estatística $2NI_0 = G^2$ como X_N^2 está para X_p^2 .

4.5. Aplicações

Esta secção tem por finalidade ilustrar e comparar as metodologias de MV e de MQP descritas nas secções 4,3 e 4,4. Os dados que serão objecto de análise constituem um módulo componente dos dados resultantes de um estudo envolvendo um determinado grupo de crianças asmáticas, realizado entre 1983 e 1985 no ambulatório de Pneumologia do Instituto da Criança da Faculdade de Medicina da USP.

Os dados do 1º exemplo estão relacionados com a avaliação da função pulmonar de 167 crianças asmáticas. Crianças com asma podem ou não ter crises (denominadas broncoespasmos) ocasionadas pela prática de exercícios. Quando a criança tem broncoespasmo induzido por exercício (BIE), diz-se que é BIE positivo (nível 1 da variável BIE). Caso contrário, diz-se que é BIE negativo (nível 2). A

prova de função pulmonar consistiu em submeter cada criança a um exercício cicloergométrico até atingir uma frequência cardíaca superior a 180 batimentos por minuto, momento a partir do qual se mede o tempo de duração do exercício previamente fixado. Findo este, verificou-se se a criança apresentou dilatação (ausência de broncoespasmo) ou constrição dos brônquios (entrada em broncoespasmo).

Este estudo visava a classificação de cada criança segundo a variável BIE determinada após exercícios com duração respectivamente de 5 e 7 minutos. Contudo, por razões que se desconhecem (como a simples perda de alguns registos; suspeição na fiabilidade de certos resultados que conduziram à sua censura, ausência de certos elementos seleccionados a algumas provas, etc), apenas 23 crianças apresentaram registos completos. Das restantes, 81 foram apenas registadas a 5 minutos e 39 tiveram censurados os registos relativos a 5 minutos. As restantes 24 não apresentaram qualquer registo relativo a essas duas variáveis. Designando por X_1 e X_2 a variável BIE medida após um exercício respectivamente de 5 e 7 minutos, as frequências observadas estão expostas na seguinte tabela:

Tabela 4.5.I.1: Dados sobre broncoespasmo induzido por exercícios de 5 (X_1) e 7 minutos (X_2) em 167 crianças asmáticas

	X_2		não-resposta a X_2
	1	2	
X_1	1	4	50
	2	2	31
		23	81
não-resposta a X_1	27	12	39
			24

Os dados relativos ao 2º exemplo estão ligados à classificação cruzada de 167 crianças asmáticas pelas variáveis categorizadas:

- X_1 , indicando o broncoespasmo induzido por um exercício de determinada duração;
- X_2 , representando a frequência da necessidade de medicação anti-asmática, esporádica aquando de crises (nível 1) e contínua (nível 2);
- X_3 , definindo se a presença de fumaça desencadeia (nível 1) ou não (nível 2) crise asmática.

Tal como no exemplo anterior, alguns elementos amostrais não foram registados completamente, devido eventualmente a algumas das razões já apontadas ou a informações deficientes prestadas pelas mães sobre o valor de X_2 e X_3 . O padrão de registo observado consistiu numa tabela completa envolvendo 58 crianças e nas tabelas marginais suplementares (X_1, X_2) e X_1 respeitantes a 108 e 1 criança.

anças, respectivamente. Registam-se a seguir as frequências observadas:

Tabela 4.5.II.1: Dados sobre a categorização de 167 crianças asmáticas em termos de broncoespasmo induzido por exercício (X_1), frequência de medicação (X_2) e afectação por fumaça (X_3).

		$X_3=1$		$X_3=2$	
		X_2		X_2	
		1	2	1	2
X_1	1	8	26	2	1
	2	8	7	2	4
					58

		X_2		não-resposta a X_2
		1	2	
X_1	1	16	48	1
	2	27	17	0
		108		1

Em ambos os exemplos admitiremos que a amostra de crianças seleccionadas foi obtida por um processo conceptualmente equivalente à amostragem aleatória simples de uma população correspondente considerada suficientemente grande, na qual as probabilidades de cada categoria são constantes. O processo de amostragem (não totalmente observado) é assim modelado por uma distribuição multinomial. Assumiremos também que os registos são gerados, condicionalmente ao processo de amostragem, de distribuições multinomiais, de modo que o resultado conjunto dos dois processos é regulado pela distribuição multinomial (3.1).

Em qualquer dos casos, observou-se um padrão de registo em partições. A análise estatística sobre as probabilidades das categorias ignorará o mecanismo de registo, o que significa que estamos adoptando um modelo M_2 para o processo de registo e que as correspondentes inferências revestem uma natureza condicional.

Os cálculos por via iterativa segundo a metodologia de MV foram realizados através de programas executados no módulo CM do pacote SOC (1988). Os resultados relativos à metodologia de MQP foram obtidos pelo uso do programa GENCAT de Landis et al. (1976). Deve, contudo, observar-se que a efectivação de várias análises de MQP em dados categorizados incompletos não requer necessariamente o uso deste programa específico (embora de enorme flexibilidade) ou de suas versões mais actualizadas. Por exemplo, o programa CATMOD do SAS (1985), pacote mais difundido e de grande generalidade, é um meio alternativo para a execução de análises de modelos estritamente lineares.

EXEMPLO I: Os dados deste exemplo serão usados apenas para indagar se a distribuição da variável indicadora de broncoespasmo não varia com o tempo de duração do exercício.

Suporemos que o padrão de registo admissível coincide com o padrão observado, i.e., cada elemento da população em estudo tem uma probabilidade positiva para qualquer das 4 hipóteses exaustivas e mutuamente exclusivas:

- não sofrer qualquer censura (registo na tabela completa (X_1, X_2));
- sofrer censura do valor de X_2 (registo na tabela marginal X_1);
- sofrer censura do valor de X_1 (registo na tabela marginal X_2);
- sofrer censura total.

A suposição de que tais probabilidades de registo não dependem dos valores não observados implica que as 24 crianças, sem registo do valor da variável BIE a 5 e a 7 minutos, não são tomadas em consideração nas inferências sobre o parâmetro de interesse $\theta = (\theta_{11} \theta_{12} \theta_{21} \theta_{22})'$, onde θ_{ij} é a probabilidade positiva de cada criança pertencer às categorias i e j das variáveis X_1 e X_2 , respectivamente. O vector $n = (N_1' N_2' N_3')$ das 8 frequências é distribucionalmente caracterizado pelo produto das distribuições multinomiais condicionais $M_4(N_1, \theta)$, $M_2(N_2, Z_2' \theta)$ e $M_2(N_3, Z_3' \theta)$, onde $N_1 = 23$, $N_2 = 81$, $N_3 = 39$, $Z_2 = I_2 \otimes I_2$ e $Z_3 = I_2 \otimes I_2$ - o símbolo \otimes denota o produto de Kronecker à esquerda (i.e. $A \otimes B = [A b_{ij}]$).

O modelo linear básico é traduzido por $H_0: \bar{\theta}_* = (\bar{\theta}_1' \bar{\theta}_2' \bar{\theta}_3')$ onde $\bar{\theta}_1 = (\theta_{11} \theta_{12} \theta_{21})'$, $\bar{\theta}_2$ (resp. $\bar{\theta}_3$) é a probabilidade de $X_1=1$ na tabela X_1 (de $X_2 = 1$ na tabela X_2), $X = [I_3 (1 \ 1 \ 0)' (1 \ 0 \ 1)']'$. Na tabela seguinte encontram-se os resultados da estimação de $\bar{\theta}$ sob H_0 :

Tabela 4.5.I.2: Estimativas de $\{\theta_{ij}\}$ e respectivos desvios padrões estimados

celas ij	p_{ij}	$\hat{\theta}_{ij}$ (MV)	$\tilde{\theta}_{ij}$ (MQP)	$\hat{\sigma}(p_{ij})$ x 10	$\hat{\sigma}(\hat{\theta}_{ij})$ x 10	$\tilde{\sigma}(\tilde{\theta}_{ij})$ x 10
11	0,522	0,460	0,459	1,042	0,670	0,674
12	0,174	0,174	0,177	0,790	0,597	0,592
21	0,217	0,248	0,251	0,860	0,607	0,577
22	0,087	0,118	0,113	0,588	0,532	0,484

As estimativas de MV $\{\hat{\theta}_{ij}\}$ - solução de (4.2) - foram determinadas pelo algoritmo EM, a partir de vários valores iniciais. Alguns candidatos naturais para a inicialização do algoritmo são o vector $p_1 = N_1 / N_1 = (p_{ij})$ das proporções observadas dentro da tabela completa, dado na 1ª coluna da tabela 4,5,I.2, os estimadores de MV de θ ignorando uma tabela suplementar de cada vez - dados explicitamente pelas expressões (3.35) - e o estimador $\tilde{\theta} = (\tilde{\theta}_{ij})$ dos MQP definido por (4.38) e registado na 3ª coluna da mesma tabela.

Os desvios padrões estimados foram calculados como a raiz quadrada dos elementos diagonais, respectivamente, de $V_{p_1}(p_1) = (D_{p_1} - p_1 p_1') / N_1$, e das matrizes de terminadas de $\hat{V}(\hat{\theta}) = [F(\hat{\theta})]^{-1}$ em (4.9), e de $\hat{V}(\tilde{p})$ em (4.43), pela relação entre θ e $\bar{\theta}$.

Parece-nos haver uma concordância razoável entre os valores obtidos pelas duas abordagens, particularmente quando se tem em conta que algumas das sub-amostras não têm um tamanho grande. A diferença entre \hat{p} e os valores preditos sob H_0 não é significativa já que os respectivos testes de ajustamento de H_0 conduziram aos valores $X_P^2(\hat{\theta})=0,61$, $G^2(\hat{\theta})=0,62$ e $Q_{c_0}=0,65$ (determinados respectivamente por (4,25), (4,26) e (4,37)). Estes valores observados quando confrontados com a distribuição do $X_{(2)}^2$ correspondem a um nível crítico P (nível descritivo) da ordem de 0,70. Os dados globais são assim consistentes com o modelo H_0 , pelo que se pode dizer que as crianças parcialmente classificadas parecem ser tão representativas da população em estudo quanto aquelas que não sofreram qualquer censura.

Condicionalmente à validade de H_0 podemos responder, então, à questão de interesse, testando a hipótese de homogeneidade marginal $H : C \bar{\theta} = 0$, onde $C = (0 \ 1 \ -1)'$. A estimativa de MV de $\bar{\theta}$ restringida a $H_0 \cap H$, dada na 1ª linha da tabela seguinte, foi calculada por solução iterativa de (4.27) — o algoritmo convergiu rapidamente quando inicializado com as estimativas de MV irrestrita (\bar{p}_1) e restringida a H_0 (em um número máximo de 5 iterações):

Tabela 4.5.I.3: Estimativas de MV e de MQP de θ sob os modelos lineares H_0 e de simetria (em parênteses as estimativas para os dados completos).

	cela (i,j)			
	11	12	21	22
$\hat{\theta}_{ij}$	0,450 (0,522)	0,212 (0,196)	0,212 (0,196)	0,126 (0,086)
$\tilde{\theta}_{ij}$	0,450 (0,524)	0,216 (0,194)	0,216 (0,194)	0,118 (0,088)

Com base nessa estimativa, as estatísticas do teste de $H_0 \cap H$ tomaram os valores $X_P^2 = 1,59$ e $G^2 = 1,61$, o que corresponde a um nível crítico P da ordem de 0,66 ($v=3$). O resultado do teste de H condicional a H_0 é igualmente não significativo, já que a estatística G^2 correspondente (vide (4.3.4)) teve o valor observado $G^2 = 0,98$ ($P \approx 0,32$ para $v = 1$). Os correspondentes testes de ajustamento pela metodologia de MQP conduziram às mesmas conclusões, já que $Q_{c_0,c} = 1,66$ ($P \approx 0,65$) e $Q_c = 1,01$ ($P \approx 0,32$), valores estes calculados respectivamente de (4.37) (ou (4.45)), com \tilde{X} substituído por $\tilde{X} U$, onde $U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}'$, e de (4.44). A estimativa do parâmetro do modelo reduzido $H_0 \cap H : \bar{\theta}_* = \tilde{X} U \gamma$ dada em (4.46) permite calcular os valores preditos de $\{\theta_{ij}\}$ por MQP e que se expõem na 3ª linha da tabela 4.5.I.3.

Em suma, os dados analisados são compatíveis com a hipótese de igualdade das probabilidades de um BIE positivo a 5 e a 7 minutos. Deve, contudo, salientar-se que este modelo tem um ajuste menos satisfatório do que no caso de restringirmos a análise à sub-amostra completamente categorizada. A aplicação das

técnicas usuais para dados completos pelos mesmos dois métodos conduzem às estimativas registadas entre parênteses na tabela 4.5.I,3 (a estimativa de MV é dada explicitamente por $\hat{\theta}_{ii} = p_{ii}$ e por $\hat{\theta}_{ij} = (p_{ij} + p_{ji})/2$, para $i \neq j$) e a valores altamente não significativos das estatísticas de teste correspondentes ao nível crítico $P \approx 0,74$ ($v=1$). A inclusão das marginais suplementares aumenta os desvios face à simetria (comparem-se $\hat{\theta}$ e $\bar{\theta}$ da tabela 4,5,I,2 com \bar{p}_1), embora tal afastamento não seja ainda suficiente para pôr abertamente em causa a homogeneidade das distribuições binomiais marginais de X_1 e de X_2 ,

EXEMPLO II: A análise dos dados deste exemplo visará responder à questão de maior relevância que se prende com o estudo da relação entre a variável X_1 indicadora do broncoespasmo induzido por exercício e as variáveis clínicas X_2 e X_3 , indicando de algum modo a intensidade da asma, na população definida,

Pela forma de obtenção dos valores destas variáveis, é razoável admitir que X_1 é sempre observada e que os únicos subconjuntos de categorias combinadas resultam da perda de informação sobre a categorização em X_2 e X_3 . Esta posição sobre o padrão de registo admissível implica que os dados observáveis se podem configurar em 4 partições do conjunto das 8 categorias, denotadas por Π_t , $t=1,2,3,4$, concernentes respectivamente a uma tabela completa, a uma marginal bidimensional, (X_1, X_2) , a uma marginal unidimensional, X_1 , e a outra marginal de dupla entrada, (X_1, X_3) .

No quadro do modelo probabilístico multinomial assumido, a consideração do modelo para o processo de registo que estipula que as probabilidades condicionais de registo em cada tabela são independentes das categorias dessa tabela, acarreta que as frequências observáveis podem ser interpretadas à luz do modelo multinomial-produto condicional N_t/N_t , $\theta \rightsquigarrow M_{s_t}(N_t, Z_t' \theta)$, onde θ é o vector de elementos $\theta_{ijk} = P(X_1=i, X_2=j, X_3=k)$ em ordenação lexicográfica e

$$s_t = \begin{cases} 8 & , t=1 \\ 4 & , t=2,4 \\ 2 & , t=3 \end{cases} ; \quad N_t = \begin{cases} 58 & , t=1 \\ 108 & , t=2 \\ 1 & , t=3 \\ 0 & , t=4 \end{cases} ; \quad Z_t = \begin{cases} I_8 & , t=1 \\ I_2 \otimes I_4 & , t=2 \\ I_4 \otimes I_2 & , t=3 \\ (I_2 \otimes I_2) \otimes I_2 & , t=4 \end{cases}$$

Como na amostra observada, $N_4=0$, a distribuição condicional de N_4 é degenerada, pelo que a distribuição relevante envolve apenas as 3 restantes multinomiais. Por outro lado, $N_3=1$, o que põe em causa a aplicabilidade dos métodos clássicos deste capítulo, devido à sua natureza assintótica. Deste modo, restringir-nos-emos às 166 crianças classificadas na tabela completa e na tabela marginal suplementar (X_1, X_2) . As frequências observadas estão organizadas nos vectores $\tilde{N}_1 = (n_{ijk}^{(1)})$ e $\tilde{N}_2 = (n_{ij}^{(2)})$ ordenados respectivamente como θ e $\theta_2 = Z_2' \theta$.

A relação paramétrica entre as duas multinomiais é descrita por $H_0 : \bar{\theta}_* \equiv (\bar{\theta}' \ \bar{\theta}_2')' = \underline{X} \bar{\theta}$, onde $\bar{\theta}$ (resp. $\bar{\theta}_2$) é o vector θ (θ_2) truncado da sua última componente e $\underline{X} = (\underline{I}_7 \ \underline{Z}_2)'$ com \underline{Z}_2 indicando Z_2 após truncatura das últimas linha e coluna. O processo de estimação de $\bar{\theta}$ quer sob H_0 quer excluindo os dados parcialmente categorizados conduziu aos seguintes resultados:

Tabela 4.5.II.2 : Estimativas de $\{\theta_{ijk}\}$ e respectivos desvios padrões estimados

	celas ijk							
	111	112	121	122	211	212	221	222
\hat{p}_{ijk}	0.138	0.034	0.448	0.017	0.138	0.034	0.121	0.070
$\hat{\theta}_{ijk}$	0.125	0.031	0.435	0.017	0.178	0.045	0.107	0.062
(MV)	(0.126)	(0.031)	(0.437)	(0.018)	(0.177)	(0.044)	(0.107)	(0.060)
$\tilde{\theta}_{ijk}$	0.126	0.031	0.439	0.017	0.174	0.044	0.108	0.061
(MQP)	-----	-----	-----	-----	-----	-----	-----	-----
$10^2 \hat{x}_0(\hat{p}_{ijk})$	4.53	2.40	6.53	1.71	4.53	2.40	4.28	3.33
$10^2 \hat{x}_0(\hat{\theta}_{ijk})$	3.07	2.15	4.08	1.67	3.58	2.36	3.19	2.80
$10^2 \tilde{x}_0(\tilde{\theta}_{ijk})$	3.13	2.25	4.07	1.70	3.36	2.27	3.31	2.95

As estimativas de MV de $\{\theta_{ijk}\}$ dadas na 2ª linha da tabela 4.5.II.2 foram calculadas por (3.35), atendendo a que os dados censurados analisados formam uma partição. Na 3ª linha, entre parênteses, registam-se, por motivos ilustrativos, as estimativas de MV obtidas por inclusão da observação registada na tabela marginal X_1 . A inclusão desta tabela implica um padrão de registo em partições encaixadas em 2 níveis, pelo que as correspondentes estimativas de MV são definidas explicitamente por (3.39). Como era de esperar, esses 2 conjuntos de estimativas de θ concordam estreitamente. Os restantes valores registados na tabela anterior foram calculados pelo modo indicado no exemplo anterior. Na 1ª linha figura o valor da estimativa de MV irrestrita de θ , $\hat{p}_1 = N_1/N_1$, calculada usando apenas a sub-amostra completa.

Da análise dos resultados da tabela 4.5.II.2 sobressai a concordância estreita entre os valores obtidos pelas duas metodologias em confronto e a previsível diminuição dos erros padrões dos estimadores restritos de θ relativamente aos de \hat{p}_1 , como acontece no exemplo I.

O teste de ajustamento do modelo linear H_0 forneceu os resultados $X_P^2 = 1.43$, $G^2 = 1.47$ e $Q_c = 1.54$, todos altamente não significativos (os valores de P relativos a um $\chi^2_{(3)}$ oscilam entre 0.67 e 0.70), pelo que tal modelo não é desmentido pelos dados observados.

O estudo da associação entre as 3 variáveis categorizadas foi feito através de modelagem log-linear. Na sua formulação discriminada (não matricial), o modelo log-linear saturado é denotado por

$$\log \theta_{ijk} = \eta + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk} + u_{123ijk}$$

com as habituais restrições de identificabilidade (os parâmetros linearmente independentes escolhidos reportam-se a $i=j=k=1$). As estimativas de MV, sob H_0 , do parâmetro β de um modelo log-linear reduzido H , com matriz de especificação X_I , foram calculadas pela solução de (4.30) via algoritmo de Newton-Raphson. A partir delas, determinou-se a correspondente estimativa de θ e de θ_* através de (4.29) e, com base nestas, a estatística G^2 do teste de H dado H_0 , por (4.34). Os desvios padrões assintóticos estimados de $\hat{\beta}$, $\hat{\sigma}(\hat{\beta})$, e de $\hat{\theta}$ foram calculados, respectivamente, a partir da inversa de (4.32) e de (4.33).

A análise por MQP, condicional a H_0 , foi aplicada numa 2ª fase à "tabela de pseudo-contingência" formada pela estimativa dos MQP de θ sob H_0 , registada na tabela 4.5.II.2, dotada da sua matriz de covariância assintótica calculada de $\hat{V}(\hat{p})$ em (4.43). Os modelos log-lineares analisados foram expressos na formulação $G(\theta) \equiv X'_S \log \theta = X_G \beta$, como é indicado no final da sub-secção 4.4.2, e ajustados pela estatística Q_{X_G} em (4.53). A estimativa dos MQP de β e os seus desvios padrões assintóticos estimados, $\hat{\beta}$, foram calculados a partir de (4.54) e (4.55).

Na tabela 4.5.II.3 registam-se os valores atrás indicados para uma série de modelos log-lineares. A cada modelo estão associados os valores de $\hat{\beta}$, $\hat{\sigma}(\hat{\beta})$, $\hat{\theta}$ e $\hat{\sigma}(\hat{\theta})$, por esta ordem (i.e., da 1ª à 4ª linha), e de G^2 e Q_{X_G} , bem como dos respectivos níveis críticos relativos à distribuição do χ^2 com ν graus de liberdade.

Antes de procedermos à análise dos resultados da tabela 4.5.II.3, devemos referir que certos modelos log-lineares hierárquicos apresentam estimativas de MV em forma explícita para θ . É o caso, por exemplo, dos modelos de independência completa e de independência entre (X_1, X_2) e X_3 . De facto, as equações de verosimilhança ($\partial \log L / \partial \beta$) $\beta = \hat{\beta} = 0$ em (4.30) podem ser resolvidas directamente, conduzindo às seguintes expressões:

Modelo $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3$

$$X'_1 \equiv Y' = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix} \quad \begin{aligned} \hat{\theta}_{i..} &= (n_{i..}^{(1)} + n_{i..}^{(2)}) / N \\ \hat{\theta}_{.j.} &= (n_{.j.}^{(1)} + n_{.j.}^{(2)}) / N \\ \hat{\theta}_{..k} &= n_{..k}^{(1)} / N_1 \end{aligned} \quad \hat{\theta}_{ijk} = \hat{\theta}_{i..} \hat{\theta}_{.j.} \hat{\theta}_{..k}, \forall i, j, k$$

Tabela 4.5.II.3 ; Resultados do ajuste de vários modelos log-lineares

Modelo H	Estimativas dos parâmetros log-lineares e seus erros padrões em parênteses (x 10)							Estatísticas de teste de H dado H ₀ e seus níveis críticos					
	u ₁	u ₂	u ₃	u ₁₂	u ₁₃	u ₂₃	u ₁₂₃	v	G ²	P	Q _{X_G}	P	
x ₁ x ₂ x ₃	0.220	-0.246	0.847	-	-	-	-	4	23.39	0.00			
	(0.795)	(0.800)	(1.813)										
(x ₁ , x ₂) x ₃	0.180	-0.280	0.949	-	-	-	-	3	7.07	0.07		18.93	0.00
	(0.899)	(0.839)	(1.794)										
(x ₁ , x ₃) x ₂	0.158	-0.195	0.847	-0.335	-	-	-	3	19.28	0.00		15.52	0.00
	(0.846)	(0.846)	(1.813)	(0.846)									
(x ₂ , x ₃) x ₁	0.177	-0.249	0.697	-0.340	-	-	-	3	22.94	0.00		16.89	0.00
	(0.899)	(0.843)	(1.913)	(0.899)									
(x ₂ x ₃) x ₁	-0.032	-0.246	0.836	-	0.378	-	-	2	2.96	0.23		2.68	0.26
	(1.596)	(0.800)	(1.930)		(1.930)								
(x ₁ x ₃) x ₂	-0.064	-0.217	1.061	-	0.373	-	-	2	6.62	0.04		4.61	0.10
	(1.600)	(0.905)	(1.893)		(2.019)								
(x ₁ x ₂) x ₃	0.220	-0.162	0.818	-	-	-0.125	-	1	13.77	0.00		13.57	0.00
	(0.795)	(1.450)	(1.815)			(1.815)							
u ₁₂₃ =0	0.132	-0.122	0.949	-	-	-0.256	-	1	2.89	0.09		2.68	0.10
	(0.960)	(1.388)	(1.794)			(1.794)							
u ₁₂₃ +u ₁₃ =0	-0.094	-0.195	0.836	-0.334	0.378	-	-	1	0.00	1.00		0.00	1.00
	(1.622)	(0.846)	(1.930)	(0.846)	(1.930)								
u ₁₂₃ +u ₁₃ =0 u ₂₃ =0	-0.008	-0.203	0.794	-0.325	0.283	-	-	2	0.41	0.81		0.42	0.81
	(1.608)	(0.906)	(2.034)	(0.906)	(2.034)								
u ₁₂₃ +u ₁₃ =0 u ₂₃ =0	0.158	-0.111	0.818	-0.335	-	-0.125	-	3	1.57	0.67		1.52	0.68
	(0.846)	(1.476)	(1.815)	(0.846)		(1.815)							
u ₁₂₃ +u ₁₃ =0 u ₂₃ =u ₁₂ =0	0.176	-0.246	0.698	-0.339	-	-0.005	-	3	1.57	0.67		1.52	0.68
	(0.968)	(1.433)	(1.932)	(0.968)		(1.932)							
u ₁₂₃ +u ₁₃ =0 u ₂₃ =u ₁₂ =0	-0.119	-0.018	0.749	-	0.565	-0.394	-	3	1.57	0.67		1.52	0.68
	(1.649)	(1.327)	(2.100)		(1.912)	(1.582)							
u ₁₂₃ +u ₁₃ =0 u ₂₃ =u ₁₂ =0	-0.108	-0.064	1.059	-	0.368	-0.250	-	3	1.57	0.67		1.52	0.68
	(1.630)	(1.424)	(1.893)		(2.019)	(1.794)							

Modelo $(X_1, X_2) \perp\!\!\!\perp X_3$

$$X_1' = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \quad \hat{\theta}_{ij} = \begin{cases} \frac{1}{2N} \left[\sum_{i=1}^2 (n_{ii}^{(1)} + n_{ii}^{(2)}) + (n_{i..}^{(1)} + n_{i.}^{(2)}) + (n_{.i}^{(1)} + n_{.i}^{(2)}) - N \right], & i=j \\ \frac{1}{2N} \left[\sum_{i \neq j} (n_{ij}^{(1)} + n_{ij}^{(2)}) + (n_{i..}^{(1)} + n_{i.}^{(2)}) + (n_{.j}^{(1)} + n_{.j}^{(2)}) - N \right], & i \neq j \end{cases}$$

$$\hat{\theta}_{..k} = n_{..k}^{(1)} / N_1.$$

$$\hat{\theta}_{ijk} = \hat{\theta}_{ij} \cdot \hat{\theta}_{..k}, \quad \forall i, j, k$$

Nas expressões acima, $N=N_1 + N_2$, e o símbolo \cdot denota soma nos índices correspondentes, conforme a notação usual. A mesma característica é partilhada por estes modelos se incluirmos a tabela suplementar unidimensional X_1 (embora as expressões difiram naturalmente). No entanto, os modelos $(X_1, X_3) \perp\!\!\!\perp X_2$ e $(X_2, X_3) \perp\!\!\!\perp X_1$, por exemplo, já não apresentam estimativas de MV directas, contrariamente ao que sucede na situação de dados completos.

O ajuste, em primeiro lugar, do modelo de independência completa conduziu a um resultado fortemente significativo. A este respeito, convém referir que a análise condicionada à sub-amostra completa (executada pelo BMDP4F e GENCAT) revelou um resultado não significativo, ao nível de significância de 5%, por qualquer das abordagens ($X_P^2=8.96$, $G^2=9.55$, $Q_{X_G}=7.76$). A sub-amostra completamente categorizada não constitui evidência suficiente para pôr abertamente em causa a hipótese de independência completa. Os dados adicionais reforçam assim a tendência para a dissonância com tal modelo. Este aspecto verifica-se igualmente para o modelo $(X_1, X_3) \perp\!\!\!\perp X_2$, claramente inconsistente com os dados observados, mas incapaz de ser desmentido pelos dados completos aos níveis de significância usuais ($X_P^2=5.54$, $G^2=5.44$, $Q_{X_G}=5.18$). Este acontecimento exemplifica como as análises que desprezam os dados incompletos podem redundar em conclusões bem diferentes daquelas que, de outro modo, se obteriam.

Nenhum modelo de independência entre cada par de variáveis e a restante tem um ajustamento satisfatório. O modelo $(X_1, X_2) \perp\!\!\!\perp X_3$ tem um pior ajustamento por MV (abordagem menos sensível ao afastamento das condições assintóticas em cada tabela), como sucede igualmente com os dados completos. Aliás, o pobre ajustamento do modelo sem interacção de 2ª ordem aponta para a dificuldade de se conseguir um modelo hierárquico mais reduzido que seja dotado de uma boa capacidade explicativa da variabilidade dos dados.

A análise do vector $G(\hat{\theta})$ permitiu antever o óptimo ajustamento por MQP do modelo não hierárquico $u_{123} = u_{13}$, com matriz de especificação $X_G = (B_1 \ B_2)'$, $B_1 = 8 I_6$, $B_2' = (0 \ 0 \ 0 \ 0 \ -8 \ 0)$, o qual traduz, designadamente, que a presença de

uma maior sensibilidade ao fumo e de uma maior necessidade de medicação anti-asmática aumenta a plausibilidade de um broncoespasmo positivo.

Dado o resultado não significativo dos testes individuais de anulamento de u_{23} e u_{12} , sob esse modelo (aplicação de (4.56)), analisou-se primeiramente o modelo sem u_{23} , o qual continuou a desfrutar de um bom ajuste. A manutenção de um resultado não significativo para o teste de $u_{12}=0$ sob esse modelo levou-nos ao último modelo indicado na tabela, definido por $\tilde{X}_G = \begin{bmatrix} 8 & 1 & 0 \\ \tilde{3} & \tilde{3} & \\ 0 & (4.3) & a \end{bmatrix}$, $\tilde{a}' = (0 \ 8 \ 0 \ -8)$, ou equivalentemente, por $\tilde{X}_I = (\tilde{Y} \ \tilde{b})$, $\tilde{b}' = (0 \ 0 \ 2 \ -2 \ 0 \ 0 \ -2 \ 2)$, onde \tilde{Y} é a matriz de especificação para o modelo de independência completa. Este modelo proporcionou ainda um ajustamento satisfatório pelo que constitui um modelo reduzido capaz de descrever a variação essencial presente nos dados. Deve sublinhar-se ainda que a análise por MV destes modelos não hierárquicos propiciou resultados estreitamente análogos aos da abordagem por MQP.

Atendendo à interpretação dos parâmetros log-lineares, o último modelo hierárquico citado traduz particularmente que a interação entre X_1 e X_2 (resp. entre X_2 e X_3) tem a mesma magnitude em cada nível de X_3 (de X_1) mas sinais opostos. Os dados parecem assim em consonância com as seguintes conclusões:

- Quanto maior for a necessidade de medicação anti-asmática, tanto mais provável será a ocorrência de um broncoespasmo positivo entre as crianças sensíveis ao fumo;
- Quanto mais forte for a sensibilidade ao fumo, tanto mais frequente se tornará a necessidade de medicação anti-asmática entre as crianças com broncoespasmo positivo;
- As variáveis indicadoras de broncoespasmo e de sensibilidade ao fumo estão positivamente associadas entre as crianças com necessidade contínua de medicação, sendo independentes para as restantes.

Na tabela 4.5.II.4 apresentam-se os valores preditos por este modelo para θ calculados a partir da relação $\theta = D^{-1} \exp(\tilde{X}_I \tilde{\beta})$, com $\tilde{\alpha} = 1 \ 1' \exp(\tilde{X}_I \tilde{\beta})$ e $\tilde{\beta} = (u_1 \ u_2 \ u_3 \ u_{13})'$, bem como os respectivos desvios padrões assintóticos estimados (baseados numa expansão de Taylor linear).

Tabela 4.5.II.4 : Valores preditos pelo modelo $u_{123} + u_{13} = u_{23} = u_{13} = 0$ e respectivos desvios padrões estimados

	111	112	121	122	211	212	221	222
$\hat{\theta}_{ijk}$ (MV)	0.147	0.022	0.427	0.012	0.183	0.027	0.102	0.080
$\hat{\theta}_{ijk}$ (MQP)	0.145	0.025	0.426	0.013	0.176	0.030	0.095	0.090
$10^2 \hat{\sigma}(\hat{\theta}_{ijk})$	2.24	0.88	3.90	0.62	2.95	0.86	3.10	2.64
$10^2 \tilde{\sigma}(\hat{\theta}_{ijk})$	2.46	0.89	3.72	0.65	2.88	0.89	2.77	3.19

Em jeito de conclusão, devemos observar que as metodologias aplicadas aos exemplos analisados proporcionam resultados razoavelmente concordantes, particularmente se tivermos em mente que algumas das sub-amostras são apenas de tamanho moderado. Podemos certamente concluir que, independentemente das diferenças nos requisitos computacionais, cada uma delas constitui uma alternativa para uma análise em grandes amostras de dados categorizados incompletos, nas condições de ignorabilidade do mecanismo de registo.

oOo

5. ANÁLISE BAYESIANA DE DADOS COM CATEGORIZAÇÃO COMPLETA E PARCIAL

5.1. Introdução com breve revisão da literatura

O início da aplicação de métodos Bayesianos a problemas de dados incompletos parece datar aproximadamente do início da década de 70 com o estudo de Karson and Wroblewski (1970) do caso especial de uma amostragem binomial. O modelo trinomial decorrente para as observações registadas foi definido com base numa reparametrização envolvendo as probabilidades marginais e condicionais (ao tipo de classificação) das categorias e as probabilidades marginais de resposta/não resposta. A sua análise Bayesiana deste modelo baseou-se numa distribuição envolvendo dependência *a priori* das probabilidades das categorias para com os restantes parâmetros.

Motivado pela necessidade de uma análise Bayesiana de dois processos de Bernoulli univariados sob dependência *a priori* entre os seus parâmetros, Antelman (1972) procedeu ao primeiro estudo sobre uma classe de distribuições que generaliza a família Dirichlet (as chamadas distribuições Dirichlet-Beta) e destacou a sua importância na amostragem binomial com não-respostas.

Kaufman and King (1973), considerando ainda o mesmo esquema de amostragem, trataram de problemas inerentes à determinação do tamanho óptimo de amostras. As suas análises *a posteriori* e *pré a posteriori* assentaram em distribuições beta independentes para as probabilidades amostrais e probabilidades condicionais de resposta.

A conhecida generalização multivariada deste tipo de distribuição *a priori* foi usada por Basu and Pereira (1982) na sua extensão das análises anteriores do problema de não-respostas (censura total) a populações multinomiais. O mesmo problema foi ainda analisado com base numa distribuição *a priori* Dirichlet-Multinomial quando enquadrado em populações politómicas finitas.

Shefrin (1981,1983) explicitou, num contexto geral, resultados *a posteriori* com base em expansões relativamente aos valores possíveis dos dados incompletamente observados. Esta prática foi também utilizada por Gunel (1984) no quadro do modelo binomial acoplado a um membro da classe de distribuições *a priori* de Antelman (1972). A sua análise dirigiu-se a testar algumas hipóteses sobre as probabilidades do resultado conjunto dos processos de amostragem e de resposta, com base em factores de Bayes para a hipótese nula.

Métodos semelhantes de testar hipóteses sobre as probabilidades amostrais foram aplicados por Smith and Gunel (1984) no problema de uma tabela 2x2 com sub-tabelas marginais suplementares, sob um mecanismo de registo ignorável, a partir de distribuições *a priori* Dirichlet. Este mesmo problema foi analisado, do ponto de vista de estimação, por Albert (1985), com base em misturas de distribuições

Dirichlet que incorporam crenças vagas à priori sobre a estrutura de associação entre as duas variáveis categorizadas.

Kadane (1985) procedeu ao estudo da associação numa tabela de contingência 2×2 com base numa amostra que inclui observações censuradas parcialmente em cada uma das variáveis e completamente, sob vários prismas: desprezo dos dados incompletos, inclusão dos dados parcialmente censurados no quadro de um mecanismo de resposta ignorável e inclusão de todos os dados observados sob condicionamento nas probabilidades condicionais de resposta.

Dickey, Jiang and Kadane (1987) descreveram finalmente o problema de dados categorizados incompletos numa formulação geral e procederam à estimação das probabilidades amostrais através da família de distribuições Dirichlet generalizadas de Dickey (1983) - extensão da classe de distribuições de Antelman (1972). Contudo, as respectivas inferências basearam-se ou num mecanismo de registo ignorável ou no condicionamento aos parâmetros perturbadores do mecanismo de registo.

Face a estes resultados, o objectivo fundamental deste capítulo reside no desenvolvimento da teoria para a realização de inferências não condicionais sobre as probabilidades amostrais sob um mecanismo de registo não ignorável.

Nesse sentido, derivam-se nas sub-secções 5.2.1-5.2.4 os resultados necessários, usando distribuições à priori Dirichlet independentes para os vários tipos de parâmetros. Na sub-secção 5.2.5 generaliza-se esta análise pela adopção, para o vector das probabilidades conjuntas das categorias e dos tipos de classificação, de uma distribuição à priori Dirichlet generalizada que traduz dependência à priori entre as probabilidades amostrais e as várias probabilidades condicionais de registo.

Um outro aspecto diz respeito à realização de inferências condicionais aos parâmetros perturbadores. O conteúdo da sub-secção 5.2.6 generaliza os correspondentes resultados de Dickey et al. (1987) pelo facto de se basear numa distribuição à priori Dirichlet generalizada que reflecte uma dependência à priori do parâmetro de interesse para com os parâmetros perturbadores.

A secção 5.3 trata da teoria inferencial, sob um mecanismo de registo ignorável, desenvolvida por Dickey et al (1987). A nossa descrição visa esclarecer alguns aspectos. Entre eles, a apresentação em forma fechada dos resultados à posteriori por aplicação de expansões multinomiais, e a derivação de expressões explícitas para parâmetros à posteriori no caso de um padrão de censura encaixada em dois níveis.

Finalmente, na última secção ilustram-se algumas destas análises através de dois exemplos originais com dados estruturados em tabelas de contingência.

5.2. Solução Bayesiana para um mecanismo de censura não ignorável

5.2.1. Distribuições à priori

No capítulo 3 vimos que a função de verosimilhança se pode exprimir como

$$L(\underline{\theta}, \underline{\beta} | \underline{n}) = \prod_{i=1}^s (\theta_i \lambda_i^1)^{n_i} \prod_{\delta \in \Gamma} (\sum_{i \in \delta} \theta_i \lambda_i^\delta)^{n_\delta} \times (\sum_{i=1}^s \theta_i \lambda_i^0)^{n_0} \quad (5.1)$$

onde $\underline{\theta} = (\theta_1, \dots, \theta_s)'$, com $\mathbf{1}'_s \underline{\theta} = 1$, é o vector das probabilidades amostrais (positivas) de cada categoria e $\underline{\beta} = (\underline{\beta}'_1 \dots \underline{\beta}'_s)'$, sendo $\underline{\beta}'_i = (\lambda_i^\delta, \delta \in \Delta_i)'$, $\mathbf{1}'_{t_i} \underline{\beta}'_i = 1$, o vector ($t_i \times 1$) das probabilidades não nulas de registo condicionais à categoria i , $i=1, \dots, s$.

Para a especificação da distribuição à priori podemos raciocinar directamente em termos do vector $\underline{\mu}^* = (\theta_1 \underline{\beta}'_1 \dots \theta_s \underline{\beta}'_s)'$, $\mathbf{1}'_d \underline{\mu}^* = 1$, das d probabilidades conjuntas e adoptar para ele uma distribuição à priori Dirichlet, designadamente

$$\underline{\mu}^* | \underline{b}^* \rightsquigarrow D_d(\underline{b}^*) \quad (5.2)$$

onde $\underline{b}^* = (\underline{b}^{*1} \dots \underline{b}^{*s})'$ com \underline{b}^{*i} de dimensão $t_i \times 1$ e coordenadas positivas, $i=1, \dots, s$.

Esta distribuição à priori pode ser definida em termos dos parâmetros θ e $\underline{\beta}'_i$, $i=1, \dots, s$, os quais se exprimem em termos de $\underline{\mu}^*$ particionado nos vectores $\underline{\mu}^{*i} = \theta_i \underline{\beta}'_i$, $i=1, \dots, s$, por

$$\underline{\theta} = \underline{S}^{*'} \underline{\mu}^* = (\mathbf{1}'_{t_i} \underline{\mu}^{*i}, i=1, \dots, s)' \quad (5.3)$$

$$\underline{\beta}'_i = \underline{\mu}^{*i} / \mathbf{1}'_{t_i} \underline{\mu}^{*i}, i=1, \dots, s$$

onde, como foi visto no início da secção 3.2, \underline{S}^{*} é a matriz diagonal em blocos de blocos diagonais $\mathbf{1}'_{t_1}, \dots, \mathbf{1}'_{t_s}$. A matriz $d \times s$ \underline{S}^{*} indica assim uma partição das d celas, formadas a partir das categorias populacionais e de Δ , em s subconjuntos. Por conseguinte, $\underline{\theta}$ e $\underline{\beta}'_i$, $i=1, \dots, s$ reflectem, respectivamente, as probabilidades conjuntas agrupadas de acordo com tal partição e as probabilidades condicionais a cada parte dessa partição.

A caracterização da distribuição de Dirichlet enunciada e.g. em Basu and Pereira (1982, lema 1) pode ser estendida de modo a cobrir uma situação do tipo de (5.3), o que está consubstanciado no teorema C.1.1 do apêndice C. De acordo com tal resultado, a distribuição (5.2) é equivalente por (5.3) a admitir conjuntamente que

$$\begin{cases} \underline{\theta} | \underline{b}^* \rightsquigarrow D_s(\underline{S}^{*'} \underline{b}^*) \\ \underline{\beta}'_i | \underline{b}^* \rightsquigarrow D_{t_i}(\underline{b}^{*i}), i=1, \dots, s \\ \underline{\theta} \parallel \underline{\beta}'_1 \parallel \dots \parallel \underline{\beta}'_s | \underline{b}^* \end{cases} \quad (5.4)$$

Os parâmetros conhecidos das distribuições paramétricas à priori são designados de acordo com a notação usada para as respectivas variáveis aleatórias.

Ou seja, o vector especificador da distribuição de $\underline{\beta}_i = (\lambda_i^\delta, \delta \Theta \Delta_i)'$ será definido por $\underline{b}_i^* = (b_i^\delta, \delta \Theta \Delta_i)'$, $i=1, \dots, s$.

A adopção da distribuição à priori (5.2) para $\underline{\mu}^*$ conduz, portanto, a que somente seja necessário quantificar a opinião à priori sobre $\underline{\beta}$, o que pode ser feito, por exemplo, através dos métodos de avaliação de Dickey et al. (1983) ou de Chaloner and Duncan (1987).

Um esquema alternativo para a definição da distribuição à priori é o de funcionar directamente em termos de $\underline{\theta}$ e $\underline{\beta}$. A própria quantificação da opinião à priori é certamente menos problemática quando dirigida a estes parâmetros. Devido ao significado de $\underline{\theta}$ e $\underline{\beta}_i$, $i=1, \dots, s$, poderemos assumir à priori distribuições Dirichlet independentes para os $s+1$ vectores, ou seja

$$\begin{cases} \underline{\theta} | \underline{a}^* \rightsquigarrow D_s(\underline{a}^*) \\ \underline{\beta}_i | \underline{b}_i^* \rightsquigarrow D_{t_i}(\underline{b}_i^*) \quad , \text{ com } \underline{b}_i^* = (b_i^\delta, \delta \Theta \Delta_i)' \quad , i=1, \dots, s \\ \underline{\theta} || \underline{\beta}_1 || \dots || \underline{\beta}_s | \underline{a}^*, \underline{b}^* \end{cases} \quad (5.5)$$

onde $\underline{b}^* = (\underline{b}_1^{*'} \dots \underline{b}_s^{*'})' \in \mathbb{R}_+^d$ e $\underline{a}^* \in \mathbb{R}_+^s$.

Se o parâmetro $\underline{a}^* = (a_1^* \dots a_s^*)'$ for especificado de modo que $a_i^* \neq 1' \underline{t}_i \underline{b}_i^*$ para algum $i=1, \dots, s$, isto é, se $\underline{a}^* = \underline{S}^* \underline{b}^* + \underline{c}$ com $\underline{c} = (c_1 \dots c_s)' \in \mathbb{R}^s$ garantindo $\underline{a}^* \in \mathbb{R}_+^s$, então (5.5) é equivalente pelo resultado enunciado em Dickey et al. (1987, teorema 4.3) - veja-se demonstração do teorema C.3.1 no apêndice - a admitir que a distribuição à priori para $\underline{\mu}^*$ pertence à família de distribuições Dirichlet generalizadas construída por Dickey (1983) com base nas funções hipergeométricas bi dimensionais de Carlson (1977) (vide o resumo contido nos apêndices C.2 e C.3). Mais concretamente, (5.5) com $\underline{a}^* = \underline{S}^* \underline{b}^* + \underline{c} \in \mathbb{R}_+^s$ e $\underline{c} \in \mathbb{R}^s$ tal que $c_j = 1' \underline{c} \notin \{0, 1, 2, \dots\}$ (ou apenas \underline{c} inteiro não negativo) é equivalente a afirmar que

$$\underline{\mu}^* | \underline{b}^*, \underline{S}^*, \underline{c} \rightsquigarrow \mathcal{D}_d^S(\underline{b}^*, \underline{S}^*, \underline{c}) \quad (5.6)$$

o que significa que a função densidade de probabilidade à priori definida no simplex probabilístico $(d-1)$ -dimensional é

$$g(\underline{\mu}^* | \underline{b}^*, \underline{S}^*, \underline{c}) = \frac{\prod_{i=1}^s \prod_{\delta \Theta \Delta_i} (\mu_i^\delta)^{b_i^\delta - 1} \prod_{j=1}^s (S_j^{*'} \cdot \underline{\mu}^*)^{c_j}}{B(\underline{b}^*) \mathcal{R}_c(\underline{b}^*, \underline{S}^*, -\underline{c})} \quad (5.7)$$

onde $\{S_j^*\}$ representam os vectores coluna de \underline{S}^* , $c_j = 1' \underline{c}$ e $\mathcal{R}_c(\underline{b}^*, \underline{S}^*, -\underline{c})$ é a função hipergeométrica múltipla de Carlson definida por

$$\mathcal{R}_c(\underline{b}^*, \underline{S}^*, -\underline{c}) = E_{\underline{\mu}^* | \underline{b}^*} \left[\prod_{j=1}^s (S_j^{*'} \cdot \underline{\mu}^*)^{c_j} \right] \quad (5.8)$$

cujo símbolo $E_{\underline{\mu}^* | \underline{b}^*}$ denota o valor esperado segundo a distribuição $\underline{\mu}^* | \underline{b}^* \rightsquigarrow D_d(\underline{b}^*)$. Observe-se que pelo facto de \underline{S}^* indicar uma partição (veja-se (58) do apêndice C),

$$R_{c.}(\underline{b}^*, \underline{S}^*, -\underline{c}) = \frac{B(\underline{S}^* \underline{b}^* + \underline{c})}{B(\underline{S}^* \underline{b}^*)} \quad (5.9)$$

No caso dos $\{c_j\}$ serem inteiros não negativos, a forma de (5.7) e a relação (5.8) indicam que a distribuição (5.6) pode ser visualizada como a distribuição à posteriori de $\underline{\mu}^*$, resultante da actualização da distribuição Dirichlet (5.2) pela verosimilhança $\prod_{j=1}^s \theta_j^{c_j}$ relativa a uma amostragem multinomial não censurada (em relação às s categorias populacionais). Se a dimensão c . dessa amostra for suficientemente grande, o estimador de máxima verosimilhança de θ_j , c_j/c ., está próximo de θ_j , $j=1, \dots, s$. Assim, neste contexto, a escolha da distribuição à priori (5.6) - ou equivalentemente, (5.5) com $\underline{a}^* = \underline{S}^* \underline{b}^* + \underline{c}$ - exprime que as probabilidades agrupadas $\underline{S}_j^* \underline{\mu}^*$ são à priori melhor conhecidas que as suas componentes $\mu_{j\delta}$, $\delta \in \Delta_j$ (opinião certamente realista) e, além disso, próximas do valor c_j/c ., $j=1, \dots, s$.

Note-se que se apenas q ($q < s$) elementos de \underline{c} são diferentes de zero, (5.6) converte-se em

$$\underline{\mu}^* | \underline{b}^*, \underline{S}^*, \underline{c} \rightsquigarrow \mathcal{D}_d^q(\underline{b}^*, \underline{S}^*, \underline{c})$$

onde \underline{c} é o vector $q \times 1$ obtido de \underline{c} por remoção das coordenadas nulas e \underline{S}^* é a matriz $d \times q$ obtida de \underline{S}^* por eliminação das colunas \underline{S}_j^* para as quais $c_j = 0$. No caso extremo em que $\underline{c} = 0$, (5.6) corresponde a (5.2).

O procedimento inferencial Bayesiano que adoptaremos por agora basear-se-á na distribuição à priori (5.4), a qual é mais conveniente de avaliação subjectiva, embora mais rígida nas opiniões que materializa, do que a distribuição (5.6).

Dada a simetria, face a permutações, da distribuição Dirichlet, e sendo $\underline{\mu} = (\mu'_{c-p-o})'$ a permutação de $\underline{\mu}^*$ definida na secção 3.2, a distribuição (5.2) é equivalente a tomar

$$\underline{\mu} | \underline{b} \rightsquigarrow \mathcal{D}_d(\underline{b}) \quad (5.10)$$

onde $\underline{b} = (\underline{b}'_{c-p-o})'$ é a permutação correspondente de \underline{b}^* , i.e., $\underline{b}'_c = (b_i^i, i=1, \dots, s)$, $\underline{b}'_p = (\underline{b}^{\delta}, \delta \in \Gamma)$, $\underline{b}^{\delta} = (b_i^{\delta}, i \in \delta)'$ e $\underline{b}'_o = (b_i^o, i=1, \dots, s)'$. A função densidade à priori de $\underline{\mu}$ no simplex probabilístico $(d-1)$ -dimensional é dada então por

$$g(\underline{\mu} | \underline{b}) = \frac{1}{B(\underline{b})} \prod_{i=1}^s \mu_{ii}^{b_i^i-1} \times \prod_{\delta \in \Gamma} \prod_{i \in \delta} \mu_{i\delta}^{b_i^{\delta}-1} \times \prod_{i=1}^s \mu_{i\delta_o}^{b_i^o-1} \quad (5.11)$$

5.2.2. Distribuição preditiva à priori

Na secção 3.2. foi visto que o vector $\underline{n} = (N'_{c-p-o})'$ das frequências observadas apresenta a distribuição

$$\underline{n} | N; \underline{\mu} \rightsquigarrow M_{s+l+1}(N, Q' \underline{\mu}) \quad (5.12)$$

onde $\underline{Q}' = \begin{pmatrix} \underline{I}_s & \underline{0} \\ \underline{0} & \underline{Q}'_I \end{pmatrix}$, com \underline{Q}'_I a matriz $(\ell+1) \times (d-s)$ diagonal em blocos com blocos diagonais $\underline{1}'_{d_\delta}$, $\delta \in \mathcal{TU}\{\delta_o\}$. Dada a configuração desta matriz, o teorema C.1.1 aplicado a (5.10) assegura que

$$\underline{Q}' \underline{\mu} | \underline{b} \rightsquigarrow D_{s+\ell+1}(\underline{Q}' \underline{b}) \quad (5.13)$$

A distribuição preditiva *à priori* das observações é assim a mistura de (5.12) por (5.13), ou seja, a distribuição Dirichlet-Multinomial

$$\underline{n} | \underline{N}, \underline{b} \rightsquigarrow DM_{s+\ell+1}(\underline{N}, \underline{Q}' \underline{b})$$

com função de probabilidade

$$f(\underline{n} | \underline{N}, \underline{b}) = \frac{N!}{\prod_{\delta \in \Delta} n_\delta!} \frac{B(\underline{Q}' \underline{b} + \underline{n})}{B(\underline{Q}' \underline{b})} \quad (5.14)$$

Em ordem a interpretar certos resultados *à posteriori* que deduziremos adiante, vamos caracterizar agora a distribuição marginal das frequências que desdobram as contagens observadas n_δ pelas categorias $i \in \delta$. Na secção 3.2 denotaremos essas frequências por $\underline{y} = (\underline{y}'_c \underline{y}'_p \underline{y}'_o)'$, onde $\underline{y}_c = (y_{11} \dots y_{ss})'$, $\underline{y}_o = (y_{1\delta_o} \dots y_{s\delta_o})'$, $\underline{y}_p = (\underline{y}^{\delta'}, \delta \in \mathcal{T})'$ e $\underline{y}^\delta = (y_{i\delta}, i \in \delta)'$, e constatámos que

$$\underline{y} | \underline{N}, \underline{\mu} \rightsquigarrow M_d(\underline{N}, \underline{\mu}) \quad (5.15)$$

com $\underline{\mu}$ particionado em correspondência com o particionamento de \underline{y} .

Por (5.15) e (5.10) conclui-se que a distribuição marginal desse vector parcialmente não observado (apenas \underline{y}_c é observado) é

$$\underline{y} | \underline{N}, \underline{b} \rightsquigarrow DM_d(\underline{N}, \underline{b}) \quad (5.16)$$

Seja $\underline{\omega} = (\underline{N}_c, \underline{N}'_p \underline{n}_o)'$ = $\underline{P}' \underline{y}$ com \underline{P}' a matriz $(\ell+2) \times d$ obtida de \underline{Q}' substituindo as primeiras s linhas pela linha $(\underline{1}'_s \underline{0}'_{(d-s)})$, i.e., \underline{P}' é a matriz diagonal em blocos de blocos diagonais $\underline{1}'_s$ e $\underline{1}'_{d_\delta}$, $\delta \in \mathcal{TU}\{\delta_o\}$. Como \underline{P} indica uma partição das d celas em $\ell+2$ conjuntos, o teorema C.1.3 (generalização do teorema 1 de Basu and Pereira [1982]) garante que (5.16) é equivalente ao conjunto de condições

$$\left\{ \begin{array}{l} \underline{\omega} | \underline{N}, \underline{b} \rightsquigarrow DM_{\ell+2}(\underline{N}, \underline{P}' \underline{b}) \\ \underline{y}_c | \underline{\omega}, \underline{b} \rightsquigarrow DM_s(\underline{N}_c, \underline{b}_c) \\ \underline{y}^\delta | \underline{\omega}, \underline{b} \rightsquigarrow DM_{d_\delta}(n_\delta, \underline{b}^\delta), \delta \in \mathcal{T} \\ \underline{y}_o | \underline{\omega}, \underline{b} \rightsquigarrow DM_s(n_o, \underline{b}_o) \\ \underline{y}_c \perp \perp \underline{y}^{\delta_1} \perp \perp \underline{y} \dots \perp \perp \underline{y}^{\delta_\ell} \perp \perp \underline{y}_o | \underline{\omega}, \underline{b} \end{array} \right. \quad (5.17)$$

Assim, em particular (5.16) implica que

$$E(\underline{y}_o | \underline{\omega}, \underline{b}) = \underline{n}_o \underline{b}_o / \underline{1}' \underline{b}_o$$

$$E(\underline{y}^\delta | \underline{\omega}, \underline{b}) = \underline{n}_\delta \underline{b}^\delta / \underline{1}'_\delta \underline{b}^\delta, \quad \delta \in T \quad (5.18)$$

5.2.3. Distribuições à posteriori

A conjugação pelo teorema de Bayes de (5.1) quando reparametrizada em termos de $\underline{\mu}$, (5.11) e do núcleo de (5.14) origina a função densidade à posteriori

$$g(\underline{\mu} | \underline{n}; \underline{b}) = \frac{\prod_{i=1}^s \mu_{ii}^{b_i^{i+n_i}-1} \prod_{\delta \in T} \prod_{i \in \delta} \mu_{i\delta}^{b_i^\delta-1} \prod_{i=1}^s \mu_{i\delta_o}^{b_i^{\delta_o}-1}}{[B(\underline{b})/B(Q'\underline{b})]B(Q'\underline{b}+\underline{n})} \prod_{\delta \in T} (\underline{1}'_\delta \underline{\mu}^\delta)^{n_\delta} (\underline{1}'_{s\delta_o} \underline{\mu}_o)^{n_o} \quad (5.19)$$

Desdobremos agora o vector \underline{n} na soma dos vectores $\bar{x}_c = (N'_c O'_{(\ell+1)})'$ e $\bar{x}_I = (O'_s N'_I)'$, onde $N'_I = (N'_p n'_o)'$, e seja $\underline{x}_c = (N'_c O'_{(d-s)})'$. Multiplicando e dividindo (5.19) por $B(Q'\underline{b}+\underline{x}_c)/B(\underline{b}+\underline{x}_c)$, e atendendo a que

$$\frac{B(\underline{b})}{B(Q'\underline{b})} = \prod_{\delta \in T \cup \{\delta_o\}} B(\underline{b}^\delta) = \frac{B(\underline{b}+\underline{x}_c)}{B(Q'\underline{b}+\underline{x}_c)}$$

obtemos

$$g(\underline{\mu} | \underline{n}; \underline{b}) = \frac{\prod_{i=1}^s \mu_{ii}^{b_i^{i+n_i}-1} \prod_{\delta \in T} \prod_{i \in \delta} \mu_{i\delta}^{b_i^\delta-1} \prod_{i=1}^s \mu_{i\delta_o}^{b_i^{\delta_o}-1}}{B(\underline{b} + \underline{x}_c)} \frac{\prod_{\delta \in T} (\underline{1}'_\delta \underline{\mu}^\delta)^{n_\delta} (\underline{1}'_{s\delta_o} \underline{\mu}_o)^{n_o}}{B(Q'\underline{b}+\underline{n})/B(Q'\underline{b}+\underline{x}_c)} \quad (5.20)$$

O 1º factor de (5.20) está na forma de uma função densidade Dirichlet com parâmetro $\underline{b}+\underline{x}_c$. Por outro lado, designando $Q'_o = (O'_{(\ell+1,s)} Q'_I)$ a sub-matriz $(\ell+1) \times d$ de Q' e \underline{q}'_δ , $\delta \in T \cup \{\delta_o\}$, os seus vectores linha, é fácil verificar (note-se que $Q'_I \underline{x}_c = \bar{x}_I$) que

$$\frac{B(Q'\underline{b}+\underline{n})}{B(Q'\underline{b}+\underline{x}_c)} = \frac{B[Q'(\underline{b}+\underline{x}_c)+\bar{x}_I]}{B[Q'(\underline{b}+\underline{x}_c)]} = E_{\underline{\mu} | \underline{b}+\underline{x}_c} \left[\prod_{\delta \in T} (\underline{q}'_\delta \underline{\mu})^{n_\delta} (\underline{q}'_{\delta_o} \underline{\mu})^{n_o} \right] =$$

$$= R_{N'_I}(\underline{b}+\underline{x}_c, Q', -\bar{x}_I) = R_{N'_I}(\underline{b}+\underline{x}_c, Q'_o, -N'_I) \quad (5.21)$$

onde $N'_I = \underline{1}'_{\ell+1} N'_I = \sum_{\delta \in T} n_\delta + n_o$.

Assim, (5.20) exprime que a distribuição à posteriori de $\underline{\mu}$ é a distribuição Dirichlet generalizada

$$\underline{\mu} | N'_c, N'_I, \underline{b} \rightsquigarrow \mathcal{D}_d^{s+\ell+1}(\underline{b}+\underline{x}_c, Q', \bar{x}_I) \rightsquigarrow \mathcal{D}_d^{\ell+1}(\underline{b}+\underline{x}_c, Q'_o, N'_I) \quad (5.22)$$

Como Q é indicadora de uma partição, o recurso de novo ao teorema C.3.1 aplicado a (5.22) permite afirmar que

$$Q'_{\underline{\mu}|\underline{n},\underline{b}} \rightsquigarrow D_{s+l+1}(Q'_{\underline{b}+\underline{n}}) \quad (5.23)$$

o que era de esperar, dado (5.13) e a forma da função de verosimilhança $L(\underline{\mu}|\underline{n})$.

A aplicação de expansões multinomiais às potências de algumas formas lineares presentes no 2º factor de (5.20) permite definir a distribuição à posteriori de $\underline{\mu}$, (5.22), como uma mistura de distribuições Dirichlet generalizadas de multiplicidade (segundo a aceção de Dickey [1983, sec. 8]) inferior.

Por exemplo, expandindo a potência relativa aos indivíduos não classificados temos

$$(1'_{\underline{s}} \underline{\mu}_{\underline{o}})^{n_o} = \sum_{\underline{y}_o} \binom{n_o}{\underline{y}_o} \prod_{i=1}^s \mu_{i\delta_o}^{y_{i\delta_o}}$$

onde se subentende que $\sum_{\underline{y}_o}$ é o somatório para todas as contagens $\underline{y}_o = (y_{i\delta_o}, i=1, \dots, s)'$ tais que $1'_{\underline{s}} \underline{y}_o = n_o$. A função de verosimilhança converte-se assim em

$$L(\underline{\mu}|\underline{n}) = \sum_{\underline{y}_o} \binom{n_o}{\underline{y}_o} \prod_{i=1}^s \mu_{ii}^{n_i} \prod_{\delta \in T} (q'_{\delta} \underline{\mu})^{n_{\delta}} \prod_{i=1}^s \mu_{i\delta_o}^{y_{i\delta_o}} \quad (5.24)$$

e a correspondente função de probabilidade preditiva à priori em

$$E_{\underline{\mu}|\underline{b}} [L(\underline{\mu}|\underline{n})] = \sum_{\underline{y}_o} \binom{n_o}{\underline{y}_o} \frac{B(\underline{b}+\bar{\underline{y}})}{B(\underline{b})} \mathcal{R}_{N_p}(\underline{b}+\bar{\underline{y}}, \bar{Q}_o, -N_p) \quad (5.25)$$

onde $\bar{\underline{y}}$ designa o vector obtido de $\underline{y} = (N'_{c-p} \underline{y}'_o)'$ por substituição de \underline{y}_p pelo vector $\underline{0}$ da mesma dimensão, \bar{Q}_o é a matriz Q_o sem o vector coluna q_{δ_o} e $N_p = \sum_{\delta \in T} n_{\delta}$.

Aumentando a matriz Q_o com o vector coluna $q_{\pi} = (1'_{s} \underline{0}'_{(d-s)})'$ obtemos a matriz $P = (q_{\pi} Q_o)$ introduzida na sub-seccção anterior. Então, pelas propriedades da função \mathcal{R} descritas pelas expressões (57) e (58) do apêndice C, podemos escrever

$$\begin{aligned} \mathcal{R}_{N_p}(\underline{b}+\bar{\underline{y}}, \bar{Q}_o, -N_p) &= \mathcal{R}_{N_p}(\underline{b}+\bar{\underline{y}}, P, [0 \ -N'_p \ 0]') \\ &= \frac{B[P'(\underline{b}+\bar{\underline{y}}) + (0 \ -N'_p \ 0)']}{B[P'(\underline{b}+\bar{\underline{y}})]} = \frac{B(P'\underline{b}+\underline{\omega})}{B(P'\underline{b}+\bar{\underline{\omega}})} \end{aligned}$$

onde $\underline{\omega} = (N'_{c-p} N'_o n_o)'$ é o vector de frequências introduzido na sub-seccção anterior e $\bar{\underline{\omega}}$ é o vector obtido dele substituindo N_p por $\underline{0}(\lambda)$. Esta expressão revela que a função \mathcal{R} em (5.25) não depende afinal de \underline{y}_o .

Por outro lado, prova-se facilmente que

$$\frac{B(\underline{b}+\bar{\underline{y}})}{B(\underline{b})} = \frac{B(\underline{b}+\underline{y}_o)}{B(\underline{b}_o)} \frac{B(V'\underline{b}+\underline{n})}{B(V'\underline{b})} \quad (5.26)$$

onde $\underline{v}' = \left(\begin{array}{c|c} \underline{I}_{(d-s)} & \underline{0}_{(d-s,s)} \\ \hline \underline{0}'_{(d-s)} & \underline{1}'_s \end{array} \right)$ e $\underline{\bar{n}} = \left(\begin{array}{c} N_c \\ \underline{n}_o \\ \underline{0}_{(l)} \\ n_o \end{array} \right)$.

Tomando agora em consideração que a função de probabilidade de \underline{y}_o condicional em ω é (recorde-se (5.17))

$$p(\underline{y}_o | n_o, \underline{b}_o) = \binom{n_o}{\underline{y}_o} \frac{B(\underline{b}_o + \underline{y}_o)}{B(\underline{b}_o)}$$

a expressão (5.25) converte-se, por todos estes resultados, em

$$E_{\underline{\mu} | \underline{b}} [L(\underline{\mu} | \underline{n})] = \frac{B(\underline{v}' \underline{b} + \underline{\bar{n}})}{B(\underline{v}' \underline{b})} R_{N_p}(\underline{b} + \underline{\bar{y}}, \underline{\bar{Q}}_o, -N_p) \tag{5.27}$$

Consequentemente, combinando (5.11), (5.24), (5.27) e (5.26), a função densidade à posteriori de $\underline{\mu}$, $g(\underline{\mu} | \underline{b} + \underline{x}_c, \underline{Q}_o, N_I)$, em (5.20) pode escrever-se como

$$g(\underline{\mu} | \underline{b} + \underline{x}_c, \underline{Q}_o, N_I) = \sum_{\underline{y}_o} p(\underline{y}_o | n_o; \underline{b}_o) g(\underline{\mu} | \underline{b} + \underline{\bar{y}}, \underline{\bar{Q}}_o, N_p) \tag{5.28}$$

onde $g(\cdot | \underline{b} + \underline{\bar{y}}, \underline{\bar{Q}}_o, N_p)$ denota a função densidade de uma distribuição $\mathcal{D}_d^{\ell}(\underline{b} + \underline{\bar{y}}, \underline{\bar{Q}}_o, N_p)$. Ou seja, a distribuição à posteriori

$$\mathcal{D}_d^{\ell+1} \left(\begin{array}{c} \underline{b}_c + N_c \\ \underline{b}_p \\ \underline{b}_o \end{array} \right), [\underline{\bar{Q}}_o, \underline{q}_{\delta_o}], \left(\begin{array}{c} N_p \\ n_o \end{array} \right)$$

é a mistura das distribuições Dirichlet generalizadas para $\underline{\mu}$

$$\mathcal{D}_d^{\ell} \left(\begin{array}{c} \underline{b}_c + N_c \\ \underline{b}_p \\ \underline{b}_o + \underline{y}_o \end{array} \right), \underline{\bar{Q}}_o, N_p$$

pela distribuição $DM_s(n_o, \underline{b}_o)$ para \underline{y}_o dado n_o .

Este resultado deixa antever a possibilidade de exprimir a distribuição à posteriori como uma mistura de distribuições Dirichlet através da aplicação do teorema multinomial a todas as potências do 2º factor de (5.20). Com efeito,

denotando $\underline{y}_I = \left(\begin{array}{c} \underline{y}_p \\ \underline{y}_o \end{array} \right)$, temos

$$\prod_{\delta \in \mathbb{T}} (1'_{\underline{d}} \underline{\mu}_{\delta})^{\delta} (1'_{\underline{s}} \underline{\mu}_{\delta})^{n_o} = \sum_{\underline{y}_I} h(\underline{y}_I) \prod_{\delta \in \mathbb{T}} \prod_{i \in \delta} \mu_{i\delta}^{y_{i\delta}} \cdot \prod_{i=1}^s \mu_{i\delta_o}^{y_{i\delta_o}}$$

onde $\sum_{\underline{y}_I}$ designa o somatório múltiplo para todas as frequências \underline{y}_I tais que

$$Q'_I y_I = N_I (Q'_I \text{ é a sub-matriz de } Q' \text{ atrás citada}), \text{ e } h(y_I) = \prod_{\delta \in T} \binom{n_\delta}{y_\delta} \cdot \binom{n_o}{y_o} .$$

Por aplicação do mesmo método que conduziu a (5.28) obtemos

$$g(\underline{\mu} | \underline{b+x}_c, Q_o, N_I) = \sum_{y_I} \frac{h(y_I) B(\underline{b+y})}{\sum_{y_I} h(y_I) B(\underline{b+y})} g(\underline{\mu} | \underline{b+y}) \tag{5.29}$$

onde $g(\underline{\mu} | \underline{b+y})$ é a função densidade da distribuição $D_d(\underline{b+y})$. Ora, é fácil demonstrar que

$$B(\underline{b+y}) = B(Q' \underline{b+n}) \cdot \prod_{\delta \in T} B(\underline{b}^\delta + \underline{y}^\delta) \cdot B(\underline{b}_o + \underline{y}_o)$$

pelo que os pesos em (5.29), tendo em consideração a distribuição de y_I condicional em ω dada em (5.17), correspondem aos valores da função de probabilidade conjunta

$$p(y_I | N_I, \underline{b}_p, \underline{b}_o) = \prod_{\delta \in T} p(y^\delta | n_\delta, \underline{b}^\delta) p(y_o | n_o, \underline{b}_o)$$

Assim

$$g(\underline{\mu} | \underline{b+x}_c, Q_o, N_I) = \sum_{y_I} p(y_I | N_I, \underline{b}_p, \underline{b}_o) g(\underline{\mu} | \underline{b+y}) \tag{5.30}$$

traduzindo que a distribuição à posteriori é a mistura das distribuições Dirichlet para $\underline{\mu}$, $D_d([\underline{b}'_c + N'_c \underline{b}'_p + \sum_{\delta \in T} \underline{b}'_\delta + \underline{b}'_o])$, pelo produto das $l+1$ distribuições independentes

$$\begin{aligned} \underline{y}^\delta | \omega, \underline{b} &\rightsquigarrow DM_{d_\delta}(n_\delta, \underline{b}^\delta), \delta \in T \\ \underline{y}_o | \omega, \underline{b} &\rightsquigarrow DM_s(n_o, \underline{b}_o) \end{aligned}$$

A distribuição à posteriori pode ainda ser definida de forma conveniente em termos de novos parâmetros. Seja $\underline{\gamma} = (\underline{\gamma}_c, \underline{\gamma}'_p, \underline{\gamma}_o)'$ com $\underline{\gamma}_p = (\underline{\gamma}_\delta, \delta \in T)'$ o vector paramétrico definido em (3.11), i.e.

$$\underline{\gamma}_c = \frac{1}{s} \underline{1}' \underline{\mu}; \quad \underline{\gamma}_o = \frac{1}{s} \underline{1}' \underline{\mu}_o; \quad \underline{\gamma}_\delta = \frac{1}{d_\delta} \underline{1}' \underline{\mu}^\delta, \quad \delta \in T \tag{5.31}$$

ou numa forma condensada, $\underline{\gamma} = \underline{P}' \underline{\mu}$. O vector $\underline{\gamma}$ traduz assim as probabilidades marginais de classificação completa, de classificação parcial em cada $\delta \in T$ e de não classificação. Definam-se ainda as probabilidades condicionais de cada categoria dado o tipo de classificação

$$\begin{aligned} \underline{\alpha}_c &= \underline{\mu}_c / \underline{\gamma}_c = (\alpha^i_1, i=1, \dots, s)' \\ \underline{\alpha}_\delta &= \underline{\mu}^\delta / \underline{\gamma}_\delta = (\alpha^i_\delta, i \in \delta)' \\ \underline{\alpha}_o &= \underline{\mu}_o / \underline{\gamma}_o = (\alpha^i_\delta, i=1, \dots, s)' \end{aligned} \tag{5.32}$$

Atendendo a que

$$\mathbf{R}_{N_I}(\underline{b+x}_c, Q_o, -N_I) = \mathbf{R}_{N_I}(\underline{b+x}_c, \underline{P}, \begin{pmatrix} 0 \\ -N_I \end{pmatrix})$$

o resultado (5.22) pode ainda exprimir-se como

$$\mu | \underline{N}_c, \underline{N}_I; \underline{b} \rightsquigarrow \mathcal{D}_d^{\ell+2}(\underline{b} + \underline{x}_c, \underline{P}, \begin{pmatrix} 0 \\ \underline{N}_I \end{pmatrix}) \quad (5.33)$$

Recorrendo de novo ao teorema C.3.1, a distribuição à posteriori (5.33) é equivalente a considerar conjuntamente que

$$\left\{ \begin{array}{l} \gamma | \underline{n}, \underline{b} \rightsquigarrow D_{\ell+2}(\underline{P}'(\underline{b} + \underline{x}_c) + \begin{pmatrix} 0 \\ \underline{N}_I \end{pmatrix}) \\ \alpha_c | \underline{n}, \underline{b} \rightsquigarrow D_s(\underline{b} + \underline{N}_c) \\ \alpha_\delta | \underline{n}, \underline{b} \rightsquigarrow D_{d_\delta}(\underline{b}^\delta), \delta \in T \\ \alpha_o | \underline{n}, \underline{b} \rightsquigarrow D_s(\underline{b}_o) \\ \gamma \parallel \alpha_c \parallel \alpha_{\delta_1} \parallel \dots \parallel \alpha_{\delta_\ell} \parallel \alpha_o | \underline{n}, \underline{b} \end{array} \right. \quad (5.34)$$

Note-se que o parâmetro da distribuição à posteriori de γ pode ainda exprimir-

-se por $\underline{P}'\underline{b} + \underline{\omega}$, onde $\underline{\omega} = \begin{pmatrix} \underline{N}_c \\ \underline{N}_I \end{pmatrix}$ foi introduzido na sub-secção 5.2.2.

Naturalmente que a distribuição à posteriori (5.34) se pode obter pela aplicação do teorema de Bayes, usando a verosimilhança reparametrizada para $(\gamma, \alpha_c, \alpha_o, \alpha_\delta, \delta \in T)$ e a distribuição à priori para estes parâmetros, distribuição esta facilmente obtida de (5.10) se recorrermos ao teorema C.1.1.

O problema das não-respostas considerado por Basu and Pereira (1982) corresponde ao caso particular do problema geral de dados categorizados incompletos em que $\Delta = \Pi \cup \{ \delta_o \}$. Os resultados da sua análise Bayesiana gerada pela distribuição à priori (5.10) saem como particularização dos nossos resultados gerais. Por exemplo, a especialização para esse caso de (5.14), (5.17), (5.23), (5.28) e (5.34) corresponde aos resultados explicitamente citados no referido artigo.

5.2.4. Momentos à posteriori

Atendendo à fórmula (65c) do apêndice C e ao facto de Q ser indicadora de uma partição das d celas, o momento misto da distribuição (5.22) de parâmetros $\underline{a} = \underline{b} + \underline{x}_c$, Q e \bar{x}_I é

$$\begin{aligned} E \mu | \underline{n} \left(\prod_{i=1}^s \mu_{ii}^{r_{ii}} \prod_{\delta \in T} \mu_{i\delta}^{r_{i\delta}} \prod_{i \in \delta} \mu_{i\delta_o}^{r_{i\delta_o}} \right) &= \frac{B(\underline{a} + \underline{r})}{B(\underline{a})} \frac{R_{N_I}(\underline{a} + \underline{r}, Q, -\bar{x}_I)}{R_{N_I}(\underline{a}, Q, -\bar{x}_I)} \\ &= \frac{B(\underline{a} + \underline{r})}{B(\underline{a})} \frac{B[Q'(\underline{a} + \underline{r}) + \bar{x}_I]}{B[Q'(\underline{a} + \underline{r})]} \frac{B(Q'\underline{a})}{B(Q'\bar{x}_I)} \end{aligned} \quad (5.35)$$

onde \underline{r} é o vector dos $r_{i\delta}$, $\delta \in \Delta$, dispostos na mesma ordem que os parâmetros $\mu_{i\delta}$, $\delta \in \Delta$, em μ .

Em particular, fazendo $\underline{r} = \underline{e}_{i\delta}$, onde $\underline{e}_{i\delta}$ é o vector (d×1) com a coordenada correspondente a $\mu_{i\delta}$ igual a 1 e as restantes nulas, obtemos o valor da média a posteriori de $\mu_{i\delta}$, $\delta \in \Delta$, $i \in \delta$. Após alguns cálculos algébricos chega-se a

$$E_{\underline{\mu}|\underline{n}}[\mu_{i\delta}] = \begin{cases} (b_i^i + n_i)/(b.+N) , & \delta = \{i\} \\ (b_i^{\delta} + n_{i\delta} \frac{b_i^{\delta}}{\sum_{j \in \delta} b_j^{\delta}}) / (b.+N) , & \delta \in \Gamma \text{ ou } \delta = \delta_0 \text{ (} n_{\delta_0} \equiv n_0 \text{)} \end{cases} \quad (5.36)$$

onde $b. = \sum_{d=1}^d b.$

Analogamente, tomando $\underline{r} = \underline{e}_{i\delta} + \underline{e}_{j\delta'}$, (5.35) define os momentos mistos à posteriori de 2ª ordem de $\underline{\mu}$. As respectivas expressões são para cada $i, j = 1, \dots, s$ tais que $i \in \delta$ e $j \in \delta'$

$$E_{\underline{\mu}|\underline{n}}[\mu_{i\delta} \mu_{j\delta'}] = \begin{cases} (b_i^i + n_i)(b_j^j + n_j) / u(b., N) , & \delta = \{i\} ; \delta' = \{j\} , j \neq i \\ (b_i^i + n_i)(b_i^i + n_i + 1) / u(b., N) , & \delta = \delta' = \{i\} \\ (b_i^i + n_i)(b_j^{\delta'} + n_{j\delta'} \frac{b_j^{\delta'}}{\sum_{m \in \delta'} b_m^{\delta'}}) / u(b., N) , & \delta = \{i\}, \delta' \in \Gamma U \{\delta_0\} \\ (b_i^{\delta} + n_{i\delta} \frac{b_i^{\delta}}{\sum_{m \in \delta} b_m^{\delta}})(b_j^{\delta'} + n_{j\delta'} \frac{b_j^{\delta'}}{\sum_{m \in \delta'} b_m^{\delta'}}) / u(b., N) , & \delta, \delta' \in \Gamma U \{\delta_0\}, \delta \neq \delta' \\ (b_i^{\delta} + n_{i\delta} \frac{b_i^{\delta}}{\sum_{m \in \delta} b_m^{\delta}})(b_j^{\delta} + n_{j\delta} \frac{b_j^{\delta}}{\sum_{m \in \delta} b_m^{\delta} + 1}) / u(b., N) , & \delta, \delta' \in \Gamma U \{\delta_0\}, \delta = \delta' \\ (b_i^{\delta} + n_{i\delta} \frac{b_i^{\delta}}{\sum_{m \in \delta} b_m^{\delta}})(b_i^{\delta} + 1 + n_{i\delta} \frac{b_i^{\delta}}{\sum_{m \in \delta} b_m^{\delta} + 1}) / u(b., N) , & \delta, \delta' \in \Gamma U \{\delta_0\}, \\ & \delta = \delta', j = i \end{cases} \quad (5.37)$$

onde $u(b., N) = (b.+N)(b.+N+1)$.

Como os nossos objectivos centrais residem nas inferências sobre as probabilidades de cada categoria, interessa-nos calcular os momentos à posteriori de $\underline{\theta}$. Atendendo a (5.3) e designando \underline{S} a matriz obtida de \underline{S}^* , permutando os elementos de cada uma das suas colunas em sintonia com a transformação de $\underline{\mu}^*$ para $\underline{\mu}$, podemos escrever

$$\underline{\theta} = \underline{S}' \underline{\mu} = (\underline{S}'_i \underline{\mu}, i=1, \dots, s)' \quad (5.38)$$

Então, pela expressão (65c) do apêndice C, o momento misto de $\underline{\theta}$ pode definir-se através do seguinte momento generalizado de $\underline{\mu}$

$$E_{\underline{\mu}|\underline{n}}[\prod_{i=1}^s (\underline{S}'_i \underline{\mu})^{v_i}] = \frac{R_{N_I. + v.} (a, [Q' \underline{S}], \begin{pmatrix} -\bar{x}_I \\ -v \end{pmatrix})}{R_{N_I.} (a, Q, -\bar{x}_I)} \quad (5.39)$$

onde $\underline{v} = (v_1, \dots, v_s)'$ e $\underline{v} = \underline{1}'\underline{v}$.

O uso desta expressão coloca um problema computacional já que a função R do numerador não tem uma forma explícita similar à da função R do denominador (dada em (5.21)). Porém, quando todos os v_i são inteiros não negativos, a aplicação do teorema multinomial às potências do 1º membro de (5.39) permite obter uma expansão dessa função R numa soma finita envolvendo funções R de multiplicidade inferior calculáveis em forma fechada,

Na verdade, designando por $\underline{r}_i = (r_{i\delta}, \delta \in \Delta_i)'$ qualquer vector $t_i \times 1$ que dobre o inteiro v_i pelas celas (i, δ) com $\delta \in \Delta_i$, para todo o $i=1, \dots, s$ e denotando $\sum_{\underline{r}_i}$ o somatório para todos os \underline{r}_i tais que $\underline{1}'\underline{r}_i = v_i$, temos

$$\begin{aligned} \prod_{i=1}^s (S_i' \underline{\mu})^{v_i} &= \prod_{i=1}^s \left(\sum_{\underline{r}_i} \binom{v_i}{\underline{r}_i} \prod_{\delta \in \Delta_i} \mu_{i\delta}^{r_{i\delta}} \right) \\ &= \sum_{\underline{r}_1} \dots \sum_{\underline{r}_s} \left(\prod_{i=1}^s \binom{v_i}{\underline{r}_i} \right) \prod_{i=1}^s \mu_{ii}^{r_{ii}} \prod_{\delta \in \Gamma_i} \mu_{i\delta}^{r_{i\delta}} \prod_{i=1}^s \mu_{i\delta}^{r_{i\delta}} \end{aligned}$$

Seja $\underline{r}^* = (\underline{r}_1' \dots \underline{r}_s')$ e represente-se por \underline{r} o rearranjo dos elementos de \underline{r}^* correspondente à ordenação dos $\{\mu_{i\delta}\}$ em $\underline{\mu}$. O último somatório múltiplo refere-se assim a todos os vectores \underline{r}^* (ou \underline{r}) tais que $\underline{S}'\underline{r}^* = \underline{v}$ ($\underline{S}'\underline{r} = \underline{v}$). Deste modo, (5.39) transforma-se em

$$E_{\underline{\mu}|\underline{n}} \left(\prod_{i=1}^s (S_i' \underline{\mu})^{v_i} \right) = \sum_{\underline{r}^*} \left(\prod_{i=1}^s \binom{v_i}{\underline{r}_i} \right) \frac{B(\underline{a}+\underline{r})}{B(\underline{a})} \frac{R_{N_I}(\underline{a}+\underline{r}, Q, -\bar{\underline{x}}_I)}{R_{N_I}(\underline{a}, Q, -\bar{\underline{x}}_I)} \quad (5.40)$$

cuja comparação com (5.39) permite definir a representação para a função R acima referida. Atendendo às expressões em forma fechada das funções R de (5.40), o momento misto à posteriori de θ fica então definido por

$$E_{\theta|\underline{n}} \left(\prod_{i=1}^s \theta_i^{v_i} \right) = \left\{ \sum_{\underline{r}^*} \prod_{i=1}^s \binom{v_i}{\underline{r}_i} \frac{B(\underline{a}+\underline{r})}{B(\underline{a})} \frac{B[Q'(\underline{a}+\underline{r})+\bar{\underline{x}}_I]}{B[Q'(\underline{a}+\underline{r})]} \right\} \frac{B(Q'\underline{a})}{B(Q'\underline{a}+\bar{\underline{x}}_I)} \quad (5.41)$$

A média à posteriori de θ_i , denotada por $E(\theta_i|\underline{n})$, $i=1, \dots, s$, é então

$$E(\theta_i|\underline{n}) = \sum_{\delta \in \Delta_i} \frac{B(\underline{a}+e_{i\delta})}{B(\underline{a})} \frac{B[Q'(\underline{a}+e_{i\delta})+\bar{\underline{x}}_I]}{B[Q'(\underline{a}+e_{i\delta})]} \frac{B(Q'\underline{a})}{B(Q'\underline{a}+\bar{\underline{x}}_I)}$$

onde $e_{i\delta}$ é o vector $d \times 1$ cujas componentes, ordenadas como os $\{\mu_{i\delta}\}$ em $\underline{\mu}$, são nulas à excepção da correspondente à posição de $\mu_{i\delta}$ e que é 1. Atendendo a (5.36) podemos escrever para cada $i=1, \dots, s$

$$\begin{aligned}
 E(\theta_i | \underline{n}) &= \frac{1}{b_i + N} \left\{ (b_i^{i_1} + n_i) + \sum_{\delta \in \Gamma \cap \Delta_i} (b_i^{\delta} + n_{\delta} \frac{b_i^{\delta}}{\sum_{j \in \delta} b_j^{\delta}}) + (b_i^{\delta_0} + n_0 \frac{b_i^{\delta_0}}{\sum_{j=1}^s b_j^{\delta_0}}) \right\} \\
 &= \frac{1}{b_i + N} \left\{ [b_i^{i_1} + \sum_{\delta \in \Gamma \cap \Delta_i} b_i^{\delta} + b_i^{\delta_0}] + \left(n_i + \sum_{\delta \in \Gamma \cap \Delta_i} n_{\delta} \frac{b_i^{\delta}}{\sum_{j \in \delta} b_j^{\delta}} + n_0 \frac{b_i^{\delta_0}}{\sum_{j=1}^s b_j^{\delta_0}} \right) \right\} \quad (5.42)
 \end{aligned}$$

É interessante notar que na expressão entre chavetas, o 1º termo entre colchetes é $\frac{1}{b_i} b_i^*$, ou seja, por (5.4), a componente i do parâmetro θ a priori de θ . O 2º termo é a soma da frequência da categoria i com uma fração de todas as frequências respeitantes aos conjuntos de categorias confundidas que incluem a categoria i . A contribuição de cada n_{δ} com $\delta \in \Gamma$ compatível com i (respectivamente, n_0) para a categoria i é, recordando (5.34), definida pela média α_{δ}^i (resp. $\alpha_{\delta_0}^i$), dessa categoria dada a classificação em δ (resp. δ_0). Deste modo, a média α_{δ}^i (resp. $\alpha_{\delta_0}^i$) da posteriori de θ_i corresponde a uma média ponderada da média α_{δ}^i a priori de θ_i e da soma das proporções amostrais atribuídas à categoria i como atrás indicámos. Isto é

$$E(\theta_i | \underline{n}) = \frac{b_i}{b_i + N} E(\theta_i | \underline{b}) + \frac{N}{b_i + N} \left\{ \frac{n_i}{N} + \sum_{\delta \in \Gamma \cap \Delta_i} \frac{n_{\delta}}{N} E(\alpha_{\delta}^i | \underline{b}) + \frac{n_0}{N} E(\alpha_{\delta_0}^i | \underline{b}) \right\} \quad (5.43)$$

Invocando (5.18), a expressão (5.42) pode ser alternativamente interpretada em termos da média α_{δ}^i a priori de θ_i e das médias das frequências não observadas $y_{i\delta}$ condicionais aos dados, ou seja

$$E(\theta_i | \underline{n}) = \frac{b_i}{b_i + N} E(\theta_i | \underline{b}) + \frac{N}{b_i + N} \left\{ \frac{n_i}{N} + \sum_{\delta \in \Gamma \cap \Delta_i} E\left[\frac{y_{i\delta}}{N} | n_{\delta}, \underline{b}\right] + E\left[\frac{y_{i\delta_0}}{N} | n_0, \underline{b}\right] \right\} \quad (5.44)$$

Note-se que as expressões (5.42) ou (5.44) deixam transparecer o valor da média α_{δ}^i a posteriori de θ_i que se obtém na situação de dados completos, i.e. quando $\Delta = \Pi$. Omitindo os dados parcialmente categorizados, obtemos o resultado de Basu and Pereira (1982).

O vector $s \times 1$ da média α_{δ}^i a posteriori de θ pode ser posto em forma matricial. Para o efeito, designemos por \underline{b}_+^{δ} , $\delta \in \Gamma$, o vector $s \times 1$ com elementos iguais a b_i^{δ} para $i \in \delta$ e iguais a zero para $i \notin \delta$, pelo que $\frac{1}{s} \underline{b}_+^{\delta} = \frac{1}{d_{\delta}} \underline{b}_+^{\delta} = \sum_{j \in \delta} b_j^{\delta}$. Observando a forma de (5.42) para todo o $i=1, \dots, s$, é fácil concluir (note-se que $\underline{S}^* \underline{b}^* = \underline{S}' \underline{b}$) que

$$E(\theta | \underline{n}) = \frac{1}{b_i + N} (\underline{S}' \underline{b} + \underline{A}' \underline{n}) \quad (5.45)$$

onde $\underline{A}' = (\underline{A}_c' \underline{A}_p' \underline{A}_o')$ com $\underline{A}_c' = \underline{I}_s$, $\underline{A}_o' = \underline{b}_o' / \underline{1}' \underline{b}_o'$ e \underline{A}_p' é a matriz $s \times l$ com vectores coluna $\underline{b}_+^{\delta} / \underline{1}' \underline{b}_+^{\delta}$, $\delta \in \Gamma$. A expressão (5.45) afigura-se conveniente para a avaliação do estimador de Bayes $E(\theta | \underline{n})$ no espírito de uma análise pré - α a posteriori, para empregar a terminologia de Raiffa and Schlaifer (1961). Por exemplo, aten-

dendo a que \underline{n} tem uma distribuição preditiva à priori Dirichlet-Multinomial e recorrendo à expressão do vector média e da matriz de covariâncias desta distribuição (vide (24) do apêndice C), é fácil obter as expressões definidoras do vector média e da matriz de covariâncias de $E(\underline{\theta}|\underline{n})$.

Aplicando (5.41) com $v_i=2$, $v_j=0$, $j \neq i$, e tendo em conta (5.37), obtemos o momento simples de 2ª ordem à posteriori de θ_i $i=1, \dots, s$

$$E(\theta_i^2|\underline{n}) = \frac{1}{(b.+N)(b.+N+1)} \left\{ \left((b_i^{i+n_i})(b_i^{i+n_i+1}) \right) + \sum_{\substack{\delta \in TU\{\delta_o\} \\ \delta \in \Delta_i}} (b_i^{\delta+n_i} \frac{b_i^{\delta}}{1'_d b_{\delta}}) (b_i^{\delta+1+n_i} \frac{b_i^{\delta+1}}{1'_d b_{\delta+1}}) + \right. \\ \left. + 2 \sum_{\substack{\delta \in TU\{\delta_o\} \\ \delta \in \Delta_i}} (b_i^{i+n_i})(b_i^{\delta+n_i} \frac{b_i^{\delta}}{1'_d b_{\delta}}) + 2 \sum_{\substack{\delta, \delta' \in TU\{\delta_o\} \\ \delta, \delta' \in \Delta_i \\ \delta \neq \delta'}} (b_i^{\delta+n_i} \frac{b_i^{\delta}}{1'_d b_{\delta}}) (b_i^{\delta'+n_i} \frac{b_i^{\delta'}}{1'_d b_{\delta'}}) \right\} \quad (5.46)$$

A variância à posteriori de θ_i , $i=1, \dots, s$, fica portanto definida com base em (5.46) e (5.42). Igualmente se pode calcular a covariância à posteriori de qualquer par (θ_i, θ_j) .

O método acima empregue para chegar às expressões da média e variância à posteriori de cada θ_i baseou-se na distribuição Dirichlet generalizada (5.22) para $\underline{\mu}$. Em virtude de (5.34), os cálculos, designadamente, desses parâmetros podem vantajosamente ser feitos com base nas distribuições Dirichlet independentes de γ , α_c , α_o e α_δ , $\delta \in T$. Por exemplo, no tocante à média à posteriori de

$$\theta_i = \gamma_c \alpha_i^i + \sum_{\delta \in T \cap \Delta_i} \gamma_\delta \alpha_\delta^i + \gamma_o \alpha_o^i, \text{ temos}$$

$$E(\theta_i|\underline{n}) = E(\gamma_c|\underline{n})E(\alpha_i^i|\underline{n}) + \sum_{\delta \in T \cap \Delta_i} E(\gamma_\delta|\underline{n})E(\alpha_\delta^i|\underline{n}) + E(\gamma_o)E(\alpha_o^i) \\ = \frac{1'_s(b.+N)}{b.+N} \frac{b_i^{i+n_i}}{1'_s(b.+N)} + \sum_{\delta \in T \cap \Delta_i} \frac{1'_d b^{\delta+n_i}}{b.+N} \frac{b_i^{\delta}}{1'_d b_{\delta}} \frac{1'_s b_o+n_o}{b.+N} \frac{b_i^{\delta_o}}{1'_s b_o}$$

que corresponde a (5.42).

Quando as s categorias resultam do cruzamento dos níveis de 2 ou mais variáveis categorizadas, uma das questões centrais é, frequentemente, a averiguação da associação entre elas. Para tal, a atenção pode dirigir-se à(s) razão(ões) dos produtos cruzados com base na(s) qual(ais) se podem interpretar os parâmetros do usual modelo log-linear que medem interações.

Do ponto de vista Bayesiano tem assim interesse calcular momentos mistos à

posteriori, $E\left(\prod_{i=1}^s \theta_i^{v_i} \mid \underline{n}\right)$, onde os v_i , $i=1, \dots, s$ não são todos inteiros não negativos. Por exemplo, no caso ilustrativo da secção 3.1, onde as 4 categorias supostas em ordenação lexicográfica, definem uma tabela de contingência 2×2 , importa calcular os momentos à posteriori de 1ª e 2ª ordem da razão dos produtos cruzados

$$\psi = \psi(\theta) = \frac{\theta_1 \theta_4}{\theta_2 \theta_3}$$

A média à posteriori de ψ corresponde ao momento misto de $\underline{\theta}$ com $v_1=v_4=-v_2=-v_3=1$ enquanto que a variância à posteriori de ψ exige o cálculo do momento de $\underline{\theta}$ relativo a $v_1=v_4=-v_2=-v_3=2$.

Qualquer destes momentos mistos à posteriori de $\underline{\theta}$ pode ser definido por (5.39), mas o método de expansão utilizado para chegar a (5.41) fica inviabilizado pelo facto de nem todos os $\{v_i\}$ serem inteiros não negativos. O desbloqueamento computacional da função R do numerador de (5.39) pode processar-se através das expansões multinomiais que levaram à expressão da distribuição à posteriori de $\underline{\mu}$ como uma mistura de distribuições Dirichlet. Este método é equivalente a calcular o 1º membro de (5.39) através da distribuição (5.30), resultando

$$E_{\underline{\mu} \mid \underline{n}} \left(\prod_{i=1}^s (S'_i \underline{\mu})^{v_i} \right) = \sum_{\underline{y}_I} p(\underline{y}_I \mid \underline{N}_I; \underline{b}) E_{\underline{\mu} \mid \underline{b} + \underline{y}} \left(\prod_{i=1}^s (S'_i \underline{\mu})^{v_i} \right)$$

Note-se de passagem que esta expressão revela, pelo facto de $\underline{\mu} \mid \underline{b} + \underline{y} \rightsquigarrow D_d(\underline{b} + \underline{y})$ implicar $S'_i \underline{\mu} \mid \underline{b} + \underline{y} \rightsquigarrow D_s(S'(\underline{b} + \underline{y}))$, que a distribuição à posteriori de $\underline{\theta}$ é a mesma mistura que $\underline{\mu}$ das correspondentes distribuições Dirichlet para $\underline{\theta}$. Isto é

$$g(\underline{\theta} \mid \underline{n}) = \sum_{\underline{y}_I} p(\underline{y}_I \mid \underline{N}_I; \underline{b}) g(\underline{\theta} \mid S'(\underline{b} + \underline{y})) \quad (5.47)$$

onde $\underline{\theta} \mid S'(\underline{b} + \underline{y}) \rightsquigarrow D_s(S'(\underline{b} + \underline{y}))$.

Tendo em conta a expressão da função de probabilidade condicional de \underline{y}_I e algumas relações entre integrais Dirichlet referidas em 5.2.3., obtemos então

$$E_{\underline{\theta} \mid \underline{n}} \left(\prod_{i=1}^s \theta_i^{v_i} \right) = \frac{\sum_{\underline{y}_I} h(\underline{y}_I) [B(\underline{b} + \underline{y}) / B(\underline{a})] \{B[S'(\underline{b} + \underline{y}) + \underline{v}] / B[S'(\underline{b} + \underline{y})]\}}{B(Q' \underline{a} + \underline{x}_I) / B(Q' \underline{a})} \quad (5.48)$$

onde $h(\underline{y}_I) = \prod_{\delta \in T} \binom{n_\delta}{y_\delta} \cdot \binom{n_0}{y_0}$, pressupondo que \underline{y} é tal que $S'(\underline{b} + \underline{y}) + \underline{v} \in \mathbb{R}_+^s$ para

todo o \underline{y}_I tal que $Q' \underline{y}_I = \underline{N}_I$. A comparação desta expressão com (5.39) define simultaneamente a expansão da função R do numerador de (5.39).

Embora (5.48) revele a possibilidade de cálculo dos momentos mistos à pos-

teriori de θ no problema de dados "missing" sob um mecanismo de registo não ignorável, não é menos verdade que em situações com inúmeros dados incompletos, dispersos em várias configurações, a computação na prática do 2º membro de (5.47) torna-se fastidiosa. Talvez o recurso a métodos aproximados para avaliação de (5.39) em grandes amostras constitua uma alternativa preferível ao método exacto acima definido (vide, e.g., Kadane [1985] e Dickey et al. [1987]). A definição das estratégias computacionais mais adequadas para os problemas deste tipo ainda é um problema em fase de estudos mais aprofundados.

5.2.5. Inferências sob dependência à priori dos parâmetros perturbadores

Os resultados derivados em 5.2.1. - 5.2.4 proporcionam um meio de traçar inferências sobre os parâmetros de interesse no difícil problema de dados categorizados incompletos sob um mecanismo de censura não ignorável. Não se pense, porém, que a solução Bayesiana desenvolvida apresenta um grau de generalidade tal ao ponto de ser aplicável a todo o mecanismo que não é modelado por M_1 .

A análise feita apoiou-se, entre outros aspectos distribucionais, na suposição de que à priori o parâmetro β do mecanismo de registo é independente de θ , aliás, uma condição sine qua non para que um mecanismo de tipo M_1 seja Bayesianamente ignorável. Por outro lado, e acima de tudo, admitiu-se igualmente que as componentes β_i , $i=1, \dots, s$, de β são independentes à priori, hipótese que pode estar em contradição com juízos apriorísticos realistas em várias situações.

Na sub-secção 5.2.1 aludiu-se à possibilidade de usar uma distribuição à priori Dirichlet generalizada para $\underline{\mu}$ - dada em (5.6) com $\underline{c} \in \mathbb{R}_+^s$ - em ordem a dar conta de uma menor incerteza à priori sobre as funções $\theta_i = \sum_{\delta \in \Delta_i} \mu_{i\delta}$, $i=1, \dots, s$ relativamente à existente sobre os $\{\mu_{i\delta}\}$ componentes. De qualquer modo, ficou patente que o uso desta distribuição não colide com as relações da independência acima referidas. Contudo, se, no mesmo espírito, incluímos na densidade Dirichlet novos factores traduzindo um conhecimento à priori sobre as probabilidades de cada tipo de classificação maior do que em relação aos parâmetros $\{\mu_{i\delta}\}$ associados, obtemos uma distribuição Dirichlet generalizada para $\underline{\mu}$ que já reflecte uma estrutura de dependência entre os vectores $\theta, \beta_1, \dots, \beta_s$.

Mais especificamente, consideremos em vez de (5.6) a distribuição à priori

$$\underline{\mu} | \underline{b}, \underline{Z}_o, \underline{d}_o \rightsquigarrow \mathcal{D}_d^{s+l+2}(\underline{b}, \underline{Z}_o, \underline{d}_o) \quad (5.49)$$

onde $\underline{Z}_o = [\underline{S} \ \underline{P}]$, com as sub-matrizes \underline{S} , de colunas \underline{S}_i , $i=1, \dots, s$ e \underline{P} , de colunas \underline{q}_π e \underline{q}_δ , $\delta \in \text{TU}\{\delta_o\}$, definidas como anteriormente, e $\underline{d}_o = (\underline{c}' \underline{e}')'$, com $\underline{c} = (c_1, \dots, c_s)' \in \mathbb{R}_+^s$ e $\underline{e} \in \mathbb{R}_+^{l+2}$ o vector de elementos e_π e e_δ , $\delta \in \text{TU}\{\delta_o\}$.

A sua função densidade

$$g(\underline{\mu} | \underline{b}, \underline{Z}_o, \underline{d}_o) = \frac{\prod_{i=1}^s \prod_{\delta \in \Delta_i} \mu_i^{\delta-1}}{B(\underline{b})} \frac{\prod_{i=1}^s (S'_i \underline{\mu})^{c_i} (q_{\pi}^i \underline{\mu})^{e_{\pi}} \prod_{\delta \in TU\{\delta_o\}} (q_{\delta}^i \underline{\mu})^{e_{\delta}}}{\mathcal{R}_{d_o}(\underline{b}, \underline{Z}_o, -\underline{d}_o)} \quad (5.50)$$

revela que ela pode ser interpretada como a distribuição a posteriori resultante da actualização da distribuição a priori (5.6) $\mathcal{D}_d^s(\underline{b}, \underline{S}, \underline{c})$ pela "verosimilhança" dada pelas potências mencionadas dos elementos de $P^i \underline{\mu}$.

Com a finalidade de ilustrar as implicações de (5.49) na distribuição de $(\underline{\theta}, \underline{\beta})$, consideremos para simplificar o caso em que $\Delta = \prod_{i=1}^s U\{\delta_o\}$. Atendendo a que o jacobiano da transformação $\underline{\mu} \rightarrow (\underline{\theta}, \underline{\beta}_1, \dots, \underline{\beta}_s)$ é $\prod_{i=1}^s \theta_i$ e a que

$$B(\underline{b}) = \prod_{i=1}^s B(\underline{b}_i^*) B(S'_i \underline{b})$$

resulta de (5.50) que a densidade conjunta de $\underline{\theta}, \underline{\beta}_1, \dots, \underline{\beta}_s$ é

$$g(\underline{\theta}, \underline{\beta}_1, \dots, \underline{\beta}_s | \underline{b}, \underline{Z}_o, \underline{d}_o) = \left(\prod_{i=1}^s g(\underline{\beta}_i | \underline{b}_i^*) \right) \frac{\prod_{i=1}^s \theta_i^{a_i^*-1} \left(\sum_{i=1}^s \lambda_i^i \theta_i \right)^{e_{\pi}} \left(\sum_{i=1}^s \lambda_i^o \theta_i \right)^{e_{\delta_o}}}{B(S'_i \underline{b}) \mathcal{R}_{d_o}(\underline{b}, \underline{Z}_o, -\underline{d}_o)}$$

onde $g(\underline{\beta}_i | \underline{b}_i^*)$ representa a função densidade de uma distribuição beta, $D_2(\underline{b}_i^*)$ e $a_i^* = \sum_{t_i} \underline{b}_i^* + c_i$, $i=1, \dots, s$. Integrando-se em ordem a $\underline{\theta}$ obtém-se

$$g(\underline{\beta}_1, \dots, \underline{\beta}_s | \underline{b}, \underline{Z}_o, \underline{d}_o) = \left(\prod_{i=1}^s g(\underline{\beta}_i | \underline{b}_i^*) \right) \frac{B(\underline{a}^*)}{B(S'_i \underline{b})} \frac{\mathcal{R}_{e.}(\underline{a}^*, \underline{\Lambda}_o, -\underline{e})}{\mathcal{R}_{d_o}(\underline{b}, \underline{Z}_o, -\underline{d}_o)} \quad (5.51)$$

onde $\underline{\Lambda}_o$ é a matriz $s \times 2$ de linhas $\underline{\beta}_i$, $i=1, \dots, s$, $\underline{a}^* = (a_1^*, \dots, a_s^*)'$ e $\underline{e} = \underline{e}_{\pi} + \underline{e}_{\delta_o}$.

A densidade conjunta dos $\underline{\beta}_i$, $i=1, \dots, s$ é assim proporcional a um produto de s densidades Dirichlet multiplicado por uma função hipergeométrica \mathcal{R} cuja matriz argumento envolve os $\underline{\beta}_i$, o que traduz uma dependência a priori entre os $\{\underline{\beta}_i\}$. Observe-se que (5.51) se reduz ao correspondente resultado (5.5) quando $\underline{e} = \underline{0}$, como era de esperar, já que (5.49) é (5.6) nesse caso.

Por outro lado, a densidade condicional de $\underline{\theta}$ dado $\underline{\beta}_1, \dots, \underline{\beta}_s$ é

$$g(\underline{\theta} | \{\underline{\beta}_i\}, \underline{b}, \underline{Z}_o, \underline{d}_o) = \frac{\prod_{i=1}^s \theta_i^{a_i^*-1} (\lambda_c^i \theta)^{e_{\pi}} (\lambda_o^i \theta)^{e_{\delta_o}}}{B(\underline{a}^*) \mathcal{R}_{e.}(\underline{a}^*, \underline{\Lambda}_o, -\underline{e})} \quad (5.52)$$

traduzindo assim uma distribuição Dirichlet generalizada dependente dos $\{\underline{\beta}_i\}$.

Naturalmente que a adopção de uma família de distribuições representando uma gama mais lata de crenças a priori obriga-nos a pagar um preço bem caro, designadamente, pelos necessários meios de especificação de um membro dessa família. Em problemas com uma estrutura de dados incompletos complicada, a selec-

ção de uma priori dessa família que dê essencialmente conta de todos os juízos subjectivos sobre θ e β - se é que existe - constitui em geral tarefa gigantesca. Pode-se mesmo dizer que o grande busílis da análise, sob um mecanismo de registo não ignorável, reside precisamente na procura de uma distribuição à priori realista.

Os resultados inferenciais sob (5.49) são obviamente generalização daqueles derivados nas sub-seções anteriores. É fácil constatar que a distribuição preditiva à priori é definida por

$$f(\underline{n}|\underline{b}, \underline{Z}_o, \underline{d}_o) = \binom{N}{\underline{n}} \frac{B(\underline{a})}{B(\underline{b})} \frac{\mathcal{R}_{\underline{d}_o + N \underline{I}}(\underline{a}, \underline{Z}_o, -\underline{d}_o - \underline{x}_I^*)}{\mathcal{R}_{\underline{d}_o}(\underline{b}, \underline{Z}_o, -\underline{d}_o)} \quad (5.53)$$

onde $\underline{a} = \underline{b} + \underline{x}_c$, como foi definido em 5.2.4, e $\underline{x}_I^* = (\underline{0}'_{(s)} \underline{0} \cdot N \underline{I}')'$. Segue-se então que a distribuição à posteriori de $\underline{\mu}$ é

$$\underline{\mu} | \underline{n} \rightsquigarrow \mathcal{D}_d^{s+\ell+2}(\underline{a}, \underline{Z}_o, \underline{d}_o + \underline{x}_I^*) \quad (5.54)$$

resultado de esperar dado o fecho da família de distribuições \mathcal{D} sob amostragem multinomial censurada com um padrão incluído no argumento matricial de \mathcal{D} (Dickey et al. [1987, teor. 3.5]).

Se \underline{c} é um vector de inteiros não negativos, (5.54) pode exprimir-se como uma mistura de distribuições à posteriori do tipo da que se obtém com uma distribuição à priori Dirichlet. Com efeito, denotando agora $\underline{r}_i = (r_{i\delta}, \delta \in \Delta_i)'$ qualquer vector $t_i \times 1$ satisfazendo $\sum_{\delta \in \Delta_i} r_{i\delta} = c_i$ que distribua c_i pelas celas (i, δ) tais que $\delta \in \Delta_i$, $i=1, \dots, s$, podemos escrever

$$\prod_{i=1}^s (S_i' \underline{\mu})^{c_i} = \sum_{\underline{r}} h(\underline{r}) \prod_{i=1}^s \mu_{ii}^{r_{ii}} \prod_{\delta \in \Delta_i} \mu_{i\delta}^{r_{i\delta}}$$

onde \underline{r} é qualquer vector (r_1', \dots, r_s') rearranjado em sintonia com a ordenação

dos $\{\mu_{i\delta}\}$ em $\underline{\mu}$ satisfazendo $\underline{S}' \underline{r} = \underline{c}$ e $h(\underline{r}) = \prod_{i=1}^s \binom{c_i}{r_i}$. Com base nesta expansão, a distribuição à priori (5.49) pode exprimir-se pela mistura

$$g(\underline{\mu} | \underline{b}, \underline{Z}_o, \underline{d}_o) = \sum_{\underline{r}} \left\{ \frac{h(\underline{r}) B(\underline{b} + \underline{r}) \mathcal{R}_e(\underline{b} + \underline{r}, \underline{P}, -\underline{e})}{\sum_{\underline{r}} h(\underline{r}) B(\underline{b} + \underline{r}) \mathcal{R}_e(\underline{b} + \underline{r}, \underline{P}, -\underline{e})} \right\} g(\underline{\mu} | \underline{b} + \underline{r}, \underline{P}, \underline{e}) \quad (5.55)$$

onde $\underline{\mu} | \underline{b} + \underline{r}, \underline{P}, \underline{e} \rightsquigarrow \mathcal{D}_d^{\ell+2}(\underline{b} + \underline{r}, \underline{P}, \underline{e})$.

Do mesmo modo, a densidade à posteriori (5.54) corresponde à mistura

$$g(\underline{\mu}|\underline{n}) = \sum_{\underline{r}} \left\{ \frac{h(\underline{r})B(\underline{a}+\underline{r})\mathcal{R}_{e,+N_I}(\underline{a}+\underline{r},\underline{p},-\underline{e}-\binom{0}{N_I})}{\sum_{\underline{r}} h(\underline{r})B(\underline{a}+\underline{r})\mathcal{R}_{e,+N_I}(\underline{a}+\underline{r},\underline{p},-\underline{e}-\binom{0}{N_I})} \right\} g(\underline{\mu}|\underline{a}+\underline{r},\underline{p},\underline{e}+\binom{0}{N_I}) \quad (5.56)$$

onde $\underline{\mu}|\underline{a}+\underline{r},\underline{p},\underline{e}+\binom{0}{N_I} \rightsquigarrow \mathcal{D}_d^{\ell+2}(\underline{a}+\underline{r},\underline{p},\underline{e}+\binom{0}{N_I})$.

Assim, quer a distribuição misturadora quer as distribuições misturadas, estas do tipo de (5.33), são actualizadas pelos dados. A distribuição (5.33) corresponde a (5.56) ou (5.54) no caso em que $\underline{c} = \underline{0}_{(s)}$ e $\underline{e} = \underline{0}_{(\ell+2)}$.

Desnecessário será dizer que esta distribuição à posteriori já não implica para os parâmetros $\underline{\gamma}$, $\underline{\alpha}_c$, $\underline{\alpha}_\delta$, $\delta \in \text{TU}\{\delta_o\}$ uma caracterização distribucional tão simpática como a de (5.34). Em particular, a estrutura de independência nela registada deixa de existir.

Também o cálculo dos momentos à posteriori se afigura mais complicado. No tocante aos momentos de $\underline{\theta}$ temos para $\underline{v} = (v_1, \dots, v_s)'$ e $\underline{v} = \underline{1}'_s \underline{v}$

$$E_{\underline{\mu}|\underline{n}} \left(\prod_{i=1}^s (S'_{i\sim} \underline{\mu})^{v_i} \right) = \frac{\mathcal{R}_{d_o,+N_I,+v}(\underline{a},\underline{z}_o,-\underline{d}-\binom{\underline{v}}{N_I})}{\mathcal{R}_{d_o,+N_I}(\underline{a},\underline{z}_o,-\underline{d}-\underline{x}^*_I)} \quad (5.57)$$

Esta generalização de (5.39) tem apenas a agravante computacional de ambas as funções \mathcal{R} não admitirem uma forma explícita simples. De qualquer modo, expandindo os factores $\prod_{i=1}^s (S'_{i\sim} \underline{\mu})^{c_i+v_i}$ ou $(q'_{\pi\sim} \underline{\mu})^e \prod_{\delta \in \text{TU}\{\delta_o\}} (q'_{\delta\sim} \underline{\mu})^{e_\delta+n_\delta}$ no caso de $\underline{c}+\underline{v}$ ou \underline{e} , respectivamente, serem vectores inteiros não negativos, exprimimos essas funções \mathcal{R} em somas envolvendo funções \mathcal{R} de multiplicidade inferior e calculáveis em forma fechada, à semelhança do que foi feito em 5.2.4. para a avaliação de (5.39). A questão da praticabilidade computacional deste método permanece, naturalmente.

5.2.6. Inferências condicionadas nos parâmetros perturbadores

Nas sub-seções anteriores ficou definida uma solução Bayesiana para as inferências sobre $\underline{\theta}$ na presença de vários modelos para um mecanismo de censura não ignorável. Este nosso procedimento global complementa de uma forma cabal todas as análises Bayesianas até agora conhecidas. Os procedimentos descritos na literatura Bayesiana (e não só) respeitam quase sempre a mecanismos de censura modelados por M_1 . As únicas excepções conhecidas que abordam um mecanismo informativo no contexto de um padrão geral de dados incompletos (Kadane [1985]; Dickey et al. [1987]) limitam-se a inferências condicionadas nas probabilidades (condicionais) de registo e sob a independência à priori entre estas e $\underline{\theta}$.

Dada a relevância de um estudo deste gênero para averiguar a robustez das inferências sobre θ face a variações nos parâmetros β , o nosso objectivo aqui está em proceder a uma análise condicional através do uso de uma distribuição à priori para θ que seja dependente de β .

Na sub-secção anterior ficou implícito que a adopção da distribuição à priori (5.49) para μ implica que

$$\theta | \beta, b, Z, d, \underline{z}_0 \rightsquigarrow \mathcal{D}_s^{\ell+2}(a^*, \Lambda_0, e) \quad (5.58)$$

onde Λ_0 é a matriz $s \times (\ell+2)$ cuja linha i é o vector β_i preenchido com zeros nas posições correspondentes a $\delta \phi \Delta_i$, $i=1, \dots, s$. Alternativamente, sejam $v_\pi = (\lambda_1^1, \dots, \lambda_s^s)'$ e v_δ , $\delta \phi T U \{ \delta_0 \}$ os vectores $s \times 1$ traduzindo as probabilidades condicionais de registo em δ cuja coordenada i é nula se $i \notin \delta$ (note-se que $v_\pi = \lambda_c$ e $v_\delta = \lambda_0$). Então, a matriz Λ_0 tem como colunas precisamente os vectores v_π e v_δ , $\delta \phi T U \{ \delta_0 \}$.

Como $L(\theta | \beta) = \prod_{i=1}^s \theta_i^{n_i} \prod_{\delta \phi T U \{ \delta_0 \}} (v_\delta' \theta)^{n_\delta}$, demonstra-se facilmente que a distribuição à posteriori de θ condicional em β é

$$\theta | n, \beta \rightsquigarrow \mathcal{D}_s^{\ell+2}(a^* + N_c, \Lambda_0, e + \begin{pmatrix} 0 \\ N_I \end{pmatrix}) \quad (5.59)$$

Fazendo $e = 0_{(\ell+2)}$, o que corresponde a usar como priori a distribuição $D_s(a^*)$, obtemos

$$\theta | n, \beta \rightsquigarrow \mathcal{D}_s^{\ell+1}(a^* + N_c, [v_\delta, \delta \phi T U \{ \delta_0 \}], N_I) \quad (5.60)$$

que é o resultado obtido por Dickey et al (1987, teor. 3.2). Basta atentar que a distribuição (5.60) se pode exprimir como $\mathcal{D}_s^{s+\ell+1}(a^*, \Lambda, n)$, onde Λ é a matriz obtida de Λ_0 substituindo a coluna v_π pela sub-matriz diagonal de ordem s com os elementos de v_π na diagonal principal, devido a que

$$R_N(a^*, \Lambda, -n) = \frac{B(a^* + N_c)}{B(a^*)} R_{N_I}(a^* + N_c, \Lambda_0, \begin{pmatrix} 0 \\ N_I \end{pmatrix})$$

O momento à posteriori de ordem $v = (v_1, \dots, v_s)'$ de (5.59) é então

$$E_\theta | n, \beta \left(\prod_{i=1}^s \theta_i^{v_i} \right) = \frac{B(a^* + N_c + v)}{B(a^* + N_c)} \frac{R_{e + N_I}(a^* + N_c + v, \Lambda_0, -e - \begin{pmatrix} 0 \\ N_I \end{pmatrix})}{R_{e + N_I}(a^* + N_c, \Lambda_0, -e - \begin{pmatrix} 0 \\ N_I \end{pmatrix})} \quad (5.61)$$

A análise de sensibilidade de qualquer momento à posteriori de θ (ou de funções multiplicativas dos $\{\theta_i\}$) a variações nos parâmetros perturbadores pode colocar, de novo, problemas computacionais no cálculo da razão das funções hipergeométricas em situações com uma configuração complexa de dados incompletos.

Em casos mais simples, o método das expansões multinomiais da função integranda dos integrais $R_{e.+N_I}$ pode conduzir a cálculos em computador pouco dispendiosos. Com algumas hipóteses sobre β de modo a reduzir o número de parâmetros distintos - embora mantendo o carácter informativo do mecanismo de registo - e para certos valores especiais destes parâmetros, o cálculo das duas funções R em (5.61) pode simplificar-se consideravelmente ao ponto de não oferecer qualquer dificuldade. É o que se passa com a análise de Kadane (1985, sec. 3) de uma tabela 2×2 com um padrão de dados incompletos envolvendo as duas sub-tabelas marginais e as não-respostas.

Devemos salientar, no entanto, que é possível simplificar algumas análises de sensibilidade se recorrermos a outros estimadores de Bayes. Concretamente, em virtude da forma da densidade à posteriori condicional em β , a moda desta distribuição satisfaz equações do tipo de (3.4) com as seguintes modificações: no 2º membro, n_o e cada n_{δ} e n_i são substituídos, respectivamente, por $e_{\delta} + n_o$, $e_{\delta} + n_{\delta}$ e $a_i^* - 1 + n_i$, nele figurando ainda o termo adicional $e_{\pi} (\hat{\theta}_i \lambda_i^i) / \sum_{j=1}^s \hat{\theta}_j \lambda_j^j$ se $e_{\pi} \neq 0$; no 1º membro, N é correspondentemente substituído por $N + 1 - a^* - s + 1 - \sum_{i=1}^s (\ell + 2) e_i$. Como se evidenciou em 3.2, o uso da moda à posteriori condicional quando $e_i = 0 - \sum_{i=1}^s (\ell + 2)$, em vez da média à posteriori condicional, por exemplo, tem a vantagem de exigir apenas o conhecimento das chances $\lambda_i^{\delta} / \lambda_{i'}^{\delta}$, relativas a todos os pares (i, i') de $\delta \in TU\{\delta_o\}$, cuja especificação é menos problemática do que a que concerne aos $\{\lambda_i^{\delta}\}$ - veja-se, e.g., o exemplo 2 do capítulo 3. A aplicação do algoritmo EM ao conjunto de "frequências" $N_c + a^* - 1 - \sum_{i=1}^s e_i$ e $\begin{pmatrix} 0 \\ N_I \end{pmatrix} + e_i$ permite pois, indagar o grau de variabilidade do comportamento modal de (5.59) face a mudanças nos parâmetros do mecanismo de registo.

5.3. Solução Bayesiana para um mecanismo de registo ignorável

5.3.1. Caso de um padrão genérico de censura

Vimos no capítulo 3 que se for adoptado o modelo M_1 para o mecanismo de registo, a função de verosimilhança relevante é dada por

$$L_1(\theta | N_c, N_p) = \prod_{i=1}^s \theta_i^{n_i} \times \prod_{\delta \in T} (z_{\delta}^{\delta} \cdot \theta)^{n_{\delta}} \quad (5.62)$$

não envolvendo os indivíduos que sofreram censura completa.

Em ordem a assegurar a ignorabilidade do mecanismo de registo consideraremos que θ é distribuído à priori independentemente de λ_{M_1} (o vector λ sob M_1) segundo

$$\theta | u \rightsquigarrow D_s(u), \quad u \in \mathbb{R}_+^s \quad (5.63)$$

Por (5.62) e (5.63) conclui-se então que a função densidade à posteriori é

$$g(\theta | \underline{N}_{\underline{c}}, \underline{N}_{\underline{p}}) = \frac{\prod_{i=1}^s \theta_i^{u_i + n_i - 1}}{B(\underline{u} + \underline{N}_{\underline{c}})} \frac{\prod_{\delta \in T} (z'_{\delta} \theta)^{n_{\delta}}}{R_{N_{\underline{p}}}(\underline{u} + \underline{N}_{\underline{c}}, \underline{Z}, -\underline{N}_{\underline{p}})} \quad (5.64)$$

revelando que

$$\theta | \underline{N}_{\underline{c}}, \underline{N}_{\underline{p}} \rightsquigarrow \mathcal{D}_S^{\ell}(\underline{u} + \underline{N}_{\underline{c}}, \underline{Z}, \underline{N}_{\underline{p}})$$

onde \underline{Z} é a matriz $s \times \ell$ de colunas \underline{z}_{δ} , $\delta \in T$.

Como notam Dickey et al (1987), o resultado (5.64) permite-nos dizer que a família conjugada de (5.62) é a família de distribuições Dirichlet generalizadas com a matriz argumento incluindo \underline{Z} .

Por um raciocínio idêntico ao que foi utilizado em 5.2.3., (5.64) pode exprimir-se como uma mistura de densidades Dirichlet. Com efeito, aplicando o teorema multinomial a todas as potências de formas lineares de θ , obtemos

$$\prod_{\delta \in T} (z'_{\delta} \theta)^{n_{\delta}} = \prod_{\delta \in T} \left\{ \sum_{\underline{M}_{\delta}} \binom{n_{\delta}}{\underline{M}_{\delta}} \prod_{i=1}^s \theta_i^{M_{i\delta}} \right\}$$

onde $\underline{M}_{\delta} = (M_{i\delta}, i=1, \dots, s)'$ é um vector de frequências fictícias desdobrando as n_{δ} observações pelas categorias $i\delta$ (i.e. tal que $M_{i\delta} = 0$, $i\delta$ e $\sum_{s=1}^s M_{s\delta} = n_{\delta}$) e o somatório é entendido como concernente a todos os \underline{M}_{δ} tais que $\sum_{s=1}^s M_{s\delta} = n_{\delta}$. Se dispusermos cada conjunto de vectores \underline{M}_{δ} , $\delta \in T$ numa matriz, obtemos vários conjuntos de matrizes \underline{M} , digamos, satisfazendo cada uma delas a condição $\sum_{s=1}^s \underline{M} = \underline{N}'$. Cada uma destas matrizes é do tipo de $\underline{Z} = (z_{i\delta})$, sendo $M_{i\delta} = 0$ sempre que $z_{i\delta} = 0$ e $M_{i\delta}$ um inteiro não negativo sempre que $z_{i\delta} = 1$. A cada matriz \underline{M} corresponde um vector $\underline{\bar{M}} = \underline{M} \underline{1}_{\ell} = (M_1, \dots, M_s)'$, onde $M_i = \sum_{\delta \in T} M_{i\delta}$, de frequências completas compatível com as observações parcialmente categorizadas $\underline{N}_{\underline{p}}$. Deste modo, podemos escrever

$$\prod_{\delta \in T} (z'_{\delta} \theta)^{n_{\delta}} = \sum_{\underline{M}} h(\underline{M}) \prod_{i=1}^s \theta_i^{M_i} \quad (5.65)$$

onde $h(\underline{M}) = \prod_{\delta \in T} \binom{n_{\delta}}{\underline{M}_{\delta}}$ e o somatório se refere a todas as matrizes \underline{M} tais que

$\sum_{s=1}^s \underline{M} = \underline{N}'$. A expressão (5.65) traduz que a probabilidade amostral dos registos parcialmente censurados é uma média ponderada das probabilidades dos desdobramentos possíveis das frequências incompletas.

Atendendo a (5.65), o integral R em (5.64) converte-se na soma

$$R_{N_{\underline{p}}}(\underline{u} + \underline{N}_{\underline{c}}, \underline{Z}, -\underline{N}_{\underline{p}}) = \sum_{\underline{M}} h(\underline{M}) \frac{B(\underline{u} + \underline{N}_{\underline{c}} + \underline{\bar{M}})}{B(\underline{u} + \underline{N}_{\underline{c}})} \quad (5.66)$$

e, por conseguinte, (5.64) fica expressa pela mistura

$$g(\theta | \underline{N}_{\underline{c}}, \underline{N}_{\underline{p}}) = \sum_{\underline{M}} \left\{ \frac{h(\underline{M}) B(\underline{u} + \underline{N}_{\underline{c}} + \underline{\bar{M}})}{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_{\underline{c}} + \underline{\bar{M}})} \right\} g(\theta | \underline{u} + \underline{N}_{\underline{c}} + \underline{\bar{M}}) \quad (5.67)$$

onde $\theta | \underline{u} + \underline{N}_{\underline{c}} + \bar{M} \rightsquigarrow D_s(\underline{u} + \underline{N}_{\underline{c}} + \bar{M})$. A expressão entre chavetas representa uma função de probabilidade, condicional aos dados observados, para o conjunto de matrizes de frequências possíveis.

Em virtude de (5.64) - vide expressão (65c) do apêndice C -, o momento à posteriori de ordem $\underline{r} = (r_1, \dots, r_s)'$ de $\underline{\theta}$ é

$$E_{\underline{\theta} | \underline{N}_{\underline{c}}, \underline{N}_{\underline{p}}} \left(\prod_{i=1}^s \theta_i^{r_i} \right) = \frac{B(\underline{u} + \underline{N}_{\underline{c}} + \underline{r})}{B(\underline{u} + \underline{N}_{\underline{c}})} \frac{R_{N_{\underline{p}}}(\underline{u} + \underline{N}_{\underline{c}} + \underline{r}, \underline{Z}, -N_{\underline{p}})}{R_{N_{\underline{p}}}(\underline{u} + \underline{N}_{\underline{c}}, \underline{Z}, -N_{\underline{p}})} \quad (5.68)$$

Por (5.66) estes momentos podem ser sempre (embora possivelmente à custa de contas gigantescas) calculados por

$$E_{\underline{\theta} | \underline{N}_{\underline{c}}, \underline{N}_{\underline{p}}} \left(\prod_{i=1}^s \theta_i^{r_i} \right) = \frac{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_{\underline{c}} + \underline{r} + \underline{M})}{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_{\underline{c}} + \underline{M})} \quad (5.69)$$

Por escolha apropriada de \underline{r} , as expressões (5.68) ou (5.69) permitem calcular, designadamente, a média a posteriori de cada θ_i e a covariância à posteriori de cada (θ_i, θ_j) , e ainda, no contexto de uma tabela de contingência, a média e variância à posteriori de razões dos produtos cruzados.

Por outro lado, como o núcleo da densidade à posteriori está na forma de uma verosimilhança relativa às "frequências" $(\underline{N}_{\underline{c}} + \underline{u} - 1, \underline{N}_{\underline{p}})$, a moda à posteriori $\underline{\theta}^* = (\theta_i^*)$ satisfaz as equações do tipo de (3.11), i.e.

$$\theta_i^* = \frac{1}{N_{\underline{c}} + \underline{u} - s + N_{\underline{p}}} \left\{ (n_i + u_i - 1) + \sum_{\delta \in T \cap \Delta_i} n_{\delta} \frac{\theta_i^*}{\delta' \cdot \underline{\theta}^*} \right\}, \quad i=1, \dots, s \quad (5.70)$$

onde $\underline{u} = \mathbf{1}' \underline{u}$. A solução destas equações pode, pois, ser obtida recorrendo ao algoritmo específico EM.

5.3.2. Caso de um padrão de censura em partições

Tal como vimos na secção 3.4 a propósito da estimação por máxima verosimilhança, os procedimentos inferenciais Bayesianos simplificam-se quando \underline{Z} se pode particionar em sub-matrizes indicadoras de partições encaixadas. Para o ilustrarmos, vamos considerar o caso em que $\underline{Z} = (\underline{Z}_2 \underline{Z}_3)$, com \underline{Z}_2 indicando uma partição encaixada na partição indicada por \underline{Z}_3 . Nesta sub-secção retomaremos a notação usada na secção 3.4.

Sendo assim, a densidade (5.64) da distribuição à posteriori

$$D_{s_1}^{s_2+s_3} (\underline{u+N_1}, [Z_2 \ Z_3], \begin{pmatrix} N_2 \\ N_3 \end{pmatrix}) \text{ pode escrever-se como}$$

$$g(\underline{\theta} | N_1, N_2, N_3) = \frac{\prod_{i=1}^s \theta_i^{u_i+n_{1i}-1} \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} (z'_{2hm}\theta)^{n_{2hm}} \prod_{h=1}^{s_3} (z'_{3h}\theta)^{n_{3h}}}{B(\underline{u+N_1}) R_{N_2+N_3}(\underline{u+N_1}, [Z_2 \ Z_3], \begin{pmatrix} -N_2 \\ -N_3 \end{pmatrix})} \quad (5.71)$$

Pelo teorema C.3.2 e nota 2 subsequente, este resultado distribucional sobre $\underline{\theta}$ é equivalente a afirmar que os vectores $\phi_3(\underline{\theta})$, $\{\Gamma_{(2h)}(\underline{\theta}), h=1, \dots, s_3\}$ e $\{\rho_{(2hm)}(\underline{\theta}), m=1, \dots, j_h; h=1, \dots, s_3\}$ são independentemente distribuídos à posteriori segundo as distribuições

$$\left\{ \begin{array}{l} \phi_3(\underline{\theta}) | \{N_t\} \rightsquigarrow D_{s_3} (Z'_3(\underline{u+N_1}) + N_3 + \bar{N}_2) \\ \Gamma_{(2h)}(\underline{\theta}) | \{N_t\} \rightsquigarrow D_{j_h} (\bar{Z}'_{2h}(\underline{u}(\delta_h^3) + N_1(\delta_h^3)) + N_2(h)), h=1, \dots, s_3 \\ \rho_{(2hm)}(\underline{\theta}) | \{N_t\} \rightsquigarrow D_{d_{hm}^2} (\underline{u}(\delta_{hm}^2) + N_1(\delta_{hm}^2)), m=1, \dots, j_h; h=1, \dots, s_3 \end{array} \right. \quad (5.72)$$

onde $\bar{N}_2 = (1'_{-j_h} N_2(h), h=1, \dots, s_3)'$. Esta caracterização da distribuição à posteriori de $\underline{\theta}$ implica que a função R em (5.71) apresente a forma explícita (vide apêndice C.3 a seguir à nota 3)

$$R_{N_2+N_3}(\underline{u+N_1}, [Z_2 \ Z_3], \begin{pmatrix} -N_2 \\ -N_3 \end{pmatrix}) = \frac{B(Z'_3(\underline{u+N_1}) + N_3 + \bar{N}_2)}{B(Z'_3(\underline{u+N_1}))} \prod_{h=1}^{s_3} \frac{B(\bar{Z}'_{2h}(\underline{u}(\delta_h^3) + N_1(\delta_h^3)) + N_2(h))}{B(\bar{Z}'_{2h}(\underline{u}(\delta_h^3) + N_1(\delta_h^3)))}$$

Esta propriedade assegura uma expressão em forma fechada para os momentos generalizados à posteriori de $\underline{\theta}$ (e.g., a fórmula (72) do apêndice C). De um modo alternativo, podemos factorizar a função integranda destes integrais à posteriori segundo um processo similar à factorização (3.36) da verosimilhança e aplicar cabalmente os resultados (5.72). Por exemplo, para o momento misto à posteriori de ordem $\underline{r} = (r_1, \dots, r_s)'$ obtemos

$$E_{\underline{\theta} | \{N_t\}} \left(\prod_{i=1}^s \theta_i^{r_i} \right) = E_{\underline{\theta} | \{N_t\}} \left\{ \prod_{h=1}^{s_3} [\phi_{3h}(\underline{\theta})]^{z'_{3h}\underline{r}} \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} [\Gamma_{(2h)}^m(\underline{\theta})]^{z'_{2hm}\underline{r}} \times \right. \\ \left. \times \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \prod_{i \in \delta_{hm}^2} \left\{ \rho_{(2hm)}(\underline{\theta}) \right\}^{r_i} \right\} = \quad (5.73)$$

$$= \frac{B(\underline{a}_3 + Z'_3 \underline{r})}{B(\underline{a}_3)} \prod_{h=1}^{s_3} \frac{B(\underline{a}_{2h} + \bar{Z}'_{2h} \underline{r}(\delta_h^3))}{B(\underline{a}_{2h})} \prod_{h=1}^{s_3} \prod_{m=1}^{j_h} \frac{B(\underline{a}_{2hm} + \underline{r}(\delta_{hm}^2))}{B(\underline{a}_{2hm})}$$

onde \underline{a}_3 , $\{a_{2h}, h=1, \dots, s_3\}$ e $\{a_{2hm}, m=1, \dots, j_h; h=1, \dots, s_3\}$ são, respectivamente, os parâmetros das distribuições a posteriori de $\phi_3(\theta)$, $\{\Gamma_{(2h)}(\theta)\}$ e $\{\rho_{(2hm)}(\theta)\}$.

Em particular, a média a posteriori de θ_i , $i \in \delta_{hm}^2$ é

$$\begin{aligned}
 E[\theta_i | N_1, N_2, N_3] &= E[\phi_{3h}(\theta) | N_1, N_2, N_3] \quad E[\Gamma_{(2h)}^m(\theta) | N_1, N_2] E[\rho_{(2hm)}^i(\theta) | N_1] \\
 &= \frac{1}{N+u} \left\{ (n_{1i}+u_i) + n_{2hm} \left(\frac{n_{1i}+u_i}{\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)} \right) + n_{3h} \left(\frac{n_{1i}+u_i}{\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)} \right) \right\} \times \\
 &\quad \times \left\{ \frac{n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)}{\sum_{n=1}^{j_h} [n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)]} \right\} \quad (5.74)
 \end{aligned}$$

onde $N = \sum_{t=1}^3 N_t$. Comparando (5.74) com (3.39), vê-se que a média a posteriori de θ se pode calcular da expressão do estimador de máxima verosimilhança substituindo N_1 por N_1+u .

De (5.73) é fácil provar que a variância a posteriori de θ_i , $i \in \delta_{hm}^2$ é dada por

$$\text{var}(\theta_i | N_1, N_2, N_3) = \tilde{\theta}_i \{v_i(N_1+u, N_2, N_3) - \tilde{\theta}_i\} \quad (5.75)$$

onde $\tilde{\theta}_i = E[\theta_i | N_1, N_2, N_3]$

e

$$\begin{aligned}
 v_i(N_1+u, N_2, N_3) &= \frac{1}{N+u+1} \left\{ (n_{1i}+u_i+1) + n_{2hm} \left\{ \frac{n_{1i}+u_i+1}{\left[\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i) \right] + 1} \right\} + n_{3h} \left\{ \frac{n_{1i}+u_i+1}{\left[\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i) \right] + 1} \right\} \right\} \times \\
 &\quad \times \left\{ \frac{n_{2hm} + \left[\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i) \right] + 1}{\left[\sum_{m=1}^{j_h} [n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)] \right] + 1} \right\}
 \end{aligned}$$

Com a densidade a posteriori (5.71) é proporcional à verosimilhança para a "amostra" (N_1+u-1, N_2, N_3) , a moda a posteriori θ_i^* é também definida pelas expressões (3.38) ou (3.39), mas com N_1 substituído por N_1+u-1 , ou seja, as suas componentes θ_i^* para $i \in \delta_{hm}^2$ são dadas por

$$\theta_i^* = \frac{1}{N+u.-s_1} \left\{ (n_{1i}+u_i-1)+n_{2hm} \left(\frac{n_{1i}+u_i-1}{\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i-1)} \right) + n_{3h} \left(\frac{n_{1i}+u_i-1}{\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i-1)} \right) \times \right. \\ \left. \times \left(\frac{n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i-1)}{\sum_{m=1}^{j_h} [n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i-1)]} \right) \right\} \quad (5.76)$$

De acordo com Dickey et al. (1987), outra estimativa Bayesiana de θ pode ser definida através da moda da distribuição à posteriori conjunta de $\phi_3(\theta) \cdot \{\Gamma_{2h}(\theta)\}$ e $\{\rho_{(2hm)}(\theta)\}$. Atendendo a (5.72), esta estimativa, denotada por $\theta_i^{**} = (\theta_i^{**})$, é definida pela expressão (3.38a) com N_1 substituído por N_1+u em (3.38b,c), N_3 por N_3-1-s_3 em (3.38b), N_2 por N_2-1-s_2 em (3.38c) e N_1 por $N_1+u-1-s_1$ em (3.38d). Mais concretamente, para $i \in \delta_{hm}^2$ temos

$$\theta_i^{**} = \frac{n_{3h}^{-1} + \sum_{m=1}^{j_h} [n_{2hm} + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)]}{N+u.-s_3} \times \frac{(n_{2hm}-1) + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)}{\sum_{m=1}^{j_h} [n_{2hm}-1 + \sum_{i \in \delta_{hm}^2} (n_{1i}+u_i)]} \times \frac{n_{1i}+u_i-1}{\sum_{i \in \delta_{hm}^2} (n_{1i}+u_i-1)} \quad (5.77)$$

Os resultados anteriores generalizam-se facilmente ao caso de partições encaixadas em 3 ou mais níveis. No caso particular de Z ser indicadora de uma partição apenas, os resultados anteriores especializam-se nos resultados enunciados explicitamente em Dickey et al. (1987). Concretamente, a distribuição à posteriori $\theta | N_1, N_2 \rightsquigarrow \mathcal{D}_{s_2}^2(u+N_1, Z_2, N_2)$ significa que $\phi_2(\theta) = Z_2' \theta$ e $\rho_{(2j)}(\theta) = \theta_{(2j)} / \phi_{2j}(\theta)$, $j=1, \dots, s_2$ são independentemente distribuídos de acordo com as distribuições

$$\phi_2(\theta) | N_1, N_2 \rightsquigarrow D_{s_2}(Z_2(u+N_1) + N_2) \quad (5.78)$$

$$\rho_{(2j)}(\theta) | N_1 \rightsquigarrow D_{j_2}^2(u + N_1(\delta_j^2)) \quad , \quad j=1, \dots, s_2$$

Os resultados correspondentes a (5.73) - (5.77) são respectivamente

$$E_{\theta | N_1, N_2} \left[\prod_{i=1}^{s_1} \theta_i^{r_i} \right] = \frac{B(Z_2'(u+N_1)+N_2+Z_2'r)}{B(Z_2'(u+N_1)+N_2)} \prod_{j=1}^{s_2} \frac{B(u(\delta_j^2)+N_1(\delta_j^2)+r(\delta_j^2))}{B(u(\delta_j^2)+N_1(\delta_j^2))} \quad (5.79)$$

e para $i \in \delta_j^2$, $j=1, \dots, s_2$

$$\tilde{\theta}_i \equiv E(\theta_i | N_1, N_2) = \frac{1}{N+u} \left\{ (n_{1i} + u_i) + n_{2j} \left(\frac{n_{1i} + u_i}{\sum_{i \in \delta_j^2} (n_{1i} + u_i)} \right) \right\} \quad (5.80)$$

$$\text{var}(\theta_i | N_1, N_2) = \tilde{\theta}_i \left\{ \frac{1}{N+u+1} \left[(n_{1i} + u_i + 1) + n_{2j} \frac{n_{1i} + u_i + 1}{\left[\sum_{i \in \delta_j^2} (n_{1i} + u_i) \right] + 1} \right] - \tilde{\theta}_i \right\} \quad (5.81)$$

$$\theta_i^* = \frac{1}{N+u-s_1} \left\{ (n_{1i} + u_i - 1) + n_{2j} \frac{n_{1i} + u_i - 1}{\sum_{i \in \delta_j^2} (n_{1i} + u_i - 1)} \right\} \quad (5.82)$$

$$\theta_i^{**} = \frac{n_{2j}^{-1} + \sum_{i \in \delta_j^2} (n_{1i} + u_i)}{N+u-s_2} \frac{n_{1i} + u_i - 1}{\sum_{i \in \delta_j^2} (n_{1i} + u_i - 1)} \quad (5.83)$$

O facto das inferências se simplificarem consideravelmente com um padrão de censura encaixado pode ser aproveitado de modo a reduzir o número de expansões multinomiais (veja-se (5.65)) para o cálculo dos momentos à posteriori numa estrutura não encaixada, diminuindo assim o número de computações em (5.69). Este método é particularmente útil em tabelas de contingência quando os dados incompletos formam tabelas marginais suplementares, algumas das quais com a mesma dimensão. Com fins ilustrativos suponhamos que $\underline{Z} = (\underline{Z}_2 \ \underline{Z}_3)$, onde as 2 sub-matrizes indicam partições não encaixadas do conjunto das s_1 categorias.

Expandindo apenas o factor $\prod_{h=1}^{s_3} (z'_{3h} \cdot \theta)^{n_{3h}}$ da densidade à posteriori segundo o esquema descrito em 5.3.1, obtemos

$$\prod_{h=1}^{s_3} (z'_{3h} \cdot \theta)^{n_{3h}} = \sum_{\underline{M}} h(\underline{M}) \prod_{i=1}^{s_1} \theta_i^{m_i}$$

onde agora $\underline{M} = (\underline{M}_h; h=1, \dots, s_3)$, com $\underline{M}_h = (M_{ih}, i=1, \dots, s_1)'$ e $M_{ih} = 0, i \in \delta_h^3$,

é tal que $\underline{1}'_{s_1} \underline{M} = \underline{N}'_3$, $h(\underline{M}) = \prod_{h=1}^{s_3} \binom{n_{3h}}{M_h}$ e $M_{i.} = \sum_{h=1}^{s_3} M_{ih}$, $i=1, \dots, s_1$.

Fazendo $\underline{\bar{M}} = (M_{i.}, i=1, \dots, s_1)'$, a função densidade da distribuição à posteriori $\mathcal{D}_{s_1}^{s_2+s_3}(u+N_1, [Z_2 \ Z_3], \begin{pmatrix} -N_2 \\ -N_3 \end{pmatrix})$ pode exprimir-se pela mistura de funções densidade Dirichlet generalizadas de multiplicidade inferior

$$g(\theta | \underline{N}_1, \underline{N}_2, \underline{N}_3) = \sum_{\underline{M}} \left\{ \frac{h(\underline{M}) B(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) R_{N_2}(\underline{u} + \underline{N}_1 + \underline{\bar{M}}, Z_2, -N_2)}{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) R_{N_2}(\underline{u} + \underline{N}_1 + \underline{\bar{M}}, Z_2, -N_2)} \right\} g(\theta | \underline{N}_1, \underline{N}_2, \underline{\bar{M}}) \quad (5.84)$$

onde $\theta | \underline{N}_1, \underline{N}_2, \underline{\bar{M}} \rightsquigarrow \mathcal{D}_{s_1}^2(\underline{u} + \underline{N}_1 + \underline{\bar{M}}, Z_2, N_2)$.

Os momentos à posteriori definidos em (5.68) passam a ser definidos pelo quociente

$$E_{\theta | \underline{N}_1, \underline{N}_2, \underline{N}_3} \left(\prod_{i=1}^{s_1} \theta_i^{r_i} \right) = \frac{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) + N_2) / B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}))}{\sum_{\underline{M}} h(\underline{M}) B(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) + N_2) / B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}))} \quad (5.85)$$

envolvendo menos termos que aquele definido em (5.69).

5.4. Aplicações

Nesta secção ilustraremos através de dois exemplos algumas das análises Bayesianas descritas neste capítulo. Concretamente, serão comparadas as soluções obtidas pela aplicação das seguintes abordagens:

- Mecanismo de registo não ignorável com uma distribuição à priori do tipo de (5.4) (descrição em 5.2.1 - 5.2.4);
- Mecanismo de registo Bayesianamente ignorável (descrição em 5.3);
- Condicionamento na sub-amostra completamente categorizada (análise "standard" pela teoria inferencial conjugada).

As distribuições à priori nestes dois últimos casos serão naturalmente compatíveis com a distribuição à priori para o vector das probabilidades amostrais usada no 1º caso.

EXEMPLO I: O 1º problema a ser analisado corresponde ao exemplo I da secção 4.5 do capítulo anterior, pelo que os respectivos dados estão registados na tabela 4.5.I.1 aí descrita. Pela discussão havida então sobre o padrão de censura considerado admissível, os vectores das probabilidades condicionais de registo são definidos à luz da notação usada neste capítulo por

$$\beta_{ij} = (\lambda_{ij}^{ij} \quad \lambda_{ij}^{i.} \quad \lambda_{ij}^{.j} \quad \lambda_{ij}^{\delta})', \quad i, j = 1, 2$$

donde

$$\underline{\theta} = (\underline{1}'_4 \otimes \underline{I}_4) \underline{\beta}, \quad \underline{\beta} = (\beta'_{11} \quad \beta'_{12} \quad \beta'_{21} \quad \beta'_{22})'$$

Na análise "clássica" deste exemplo considerou-se apenas a questão de averiguar se os dados são compatíveis com a hipótese de homogeneidade marginal. Aqui continuar-se-á a restringir a análise a tal aspecto, pelo que a função para métrica de interesse é a função $\pi = \theta_{12} - \theta_{21}$. No quadro dos métodos Bayesianos que pretendemos ilustrar, importa, pois, calcular a média e o desvio padrão à posteriori desta função linear.

Neste exemplo, as soluções serão desenvolvidas com base no uso de duas distribuições à priori. Para a 1ª abordagem, considerar-se-á que θ e os vectores β_{ij} são independentemente distribuídos segundo as distribuições de Dirichlet

$$\beta_{ij} | b_{ij}^* \rightsquigarrow D_4(b_{ij}^*), \quad b_{ij}^* = (b_{ij}^{i,j} \quad b_{ij}^{i \cdot} \quad b_{ij}^{\cdot j} \quad b_{ij}^{\delta_0})', \quad i, j = 1, 2$$

$$\theta | b^* \rightsquigarrow D_4([\mathbb{1}'_4 \otimes \mathbb{I}_4] b^*), \quad b^* = (b'_{11} \quad b'_{12} \quad b'_{21} \quad b'_{22})'$$

A 1ª distribuição à priori, denominada priori 1, consiste numa distribuição uniforme trivariada para todo o β_{ij} , resultando então para θ a distribuição Dirichlet simétrica de parâmetro 4.

A 2ª distribuição à priori (priori 2), conceptualmente admitida como traduzindo a crença de algum perito no assunto, visa conferir aos β_{ij} um carácter informativo. Imaginemos então que tal perito espera à priori que, das crianças realmente pertencentes à cela (i,j), para todo o (i,j), 50% são registadas completamente, 20% registadas parcialmente em cada variável e as restantes 10% censuradas totalmente. Além disso, acredita que a precisão destas suas crenças pode ser medida através do valor 10 de $\mathbb{1}'_4 b_{ij}^*$, $\forall (i,j)$. Dentro da família Dirichlet, a opinião à priori desse perito conceptual corresponde a atribuir a todo o β_{ij} a distribuição $D_4(b_{ij}^*)$ com $b_{ij}^* = (5 \ 2 \ 2 \ 1)'$. Assim θ continua sendo caracterizado por uma distribuição Dirichlet simétrica, com a mesma média da priori 1, mas de parâmetro 10 (portanto, com incerteza menor sobre cada elemento de θ).

1. No caso da abordagem que incorpora o mecanismo de registo (e consequentemente, todas as 167 observações), as médias e os desvios padrões à posteriori dos elementos de θ , bem como a covariância à posteriori de θ_{12} e θ_{21} , foram respectivamente calculados de (5.36) e (5.37). As médias e os momentos simples de 2ª ordem à posteriori de cada elemento de θ são dados explicitamente em (5.42) e (5.46) respectivamente. Aqueles resultados, para ambas as distribuições à priori, são reproduzidos na tabela 5.4.I.2, a qual regista ainda a média e o desvio padrão à posteriori de π .

TABELA 5.4.I.2: Resultados para mecanismo não ignorável

	Médias à posteriori				desvios padrões à posteriori			
	θ_{11}	θ_{12}	θ_{21}	π	θ_{11}	θ_{12}	θ_{21}	π
Priori 1	0.331	0.246	0.240	0.006	0.100	0.093	0.080	0.126
Priori 2	0.321	0.246	0.242	0.004	0.077	0.071	0.059	0.097

2. A metodologia que considera ignorável o mecanismo de registo adopta para este o modelo M_1 que estipula que as probabilidades condicionais de registo em qualquer subconjunto δ de categorias só depende de δ e não dos elementos que o compõem. Além disso, supõe a independência à priori entre $\underline{\theta}$ e o vector das probabilidades condicionais de registo distintas. Estas suposições implicam particularmente que as 24 crianças para as quais nada foi registado são ignoradas do ponto de vista das inferências sobre $\underline{\theta}$. As distribuições à priori para $\underline{\theta}$ correspondem àquelas que foram usadas na 1ª metodologia.

Os momentos à posteriori de $\underline{\theta}$ são calculáveis por (5.69), o que requer o cálculo de dois somatórios múltiplos envolvendo cada um cerca de 594×10^3 parcelas (!). No entanto, o facto de um padrão de registo em partições encaixadas em um nível implicar uma simplificação considerável na computação daqueles momentos (veja-se (5.73)) foi aproveitado no sentido de reduzir o número de termos envolvidos em (5.69), conforme foi descrito no final da secção anterior. Assim, expandiu-se o factor da densidade à posteriori relativo à tabela $t=3$, englobando as observações censuradas segundo a variável X_1 e calcularam-se os momentos à posteriori exactos através da correspondente expressão (5.85). Mais pormenorizadamente, através da expressão, cujos somatórios envolvem 364 termos

$$E\left[\prod_{i,j} \theta_{ij}^{r_{ij}} \mid \{N_{\underline{t}}\}\right] = p(\underline{r})/q$$

onde

$$q = \sum_{a=0}^{27} \sum_{b=0}^{12} \binom{27}{a} \binom{12}{b} \frac{B(\underline{u} + \underline{N}_1 + \underline{\bar{M}})}{B(\underline{u} + \underline{N}_1)} \frac{B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}) + \underline{N}_2)}{B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}}))}$$

$$p(\underline{r}) = \sum_{a=0}^{27} \sum_{b=0}^{12} \frac{B(\underline{u} + \underline{N}_1 + \underline{\bar{M}} + \underline{r})}{B(\underline{u} + \underline{N}_1)} \frac{B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}} + \underline{r}) + \underline{N}_2)}{B(Z_2'(\underline{u} + \underline{N}_1 + \underline{\bar{M}} + \underline{r}))}$$

com $\underline{r} = (r_{11} \ r_{12} \ r_{21} \ r_{22})'$, $\underline{N}_1 = (12 \ 4 \ 5 \ 2)'$, $\underline{N}_2 = (50 \ 31)'$, $\underline{M} = (a \ b \ 27-a \ 12-b)'$, $\underline{Z}_2 = \underline{I}_2 \otimes \underline{I}_2$ e \underline{u} é o vector dos parâmetros da distribuição à priori de $\underline{\theta}$.

Os cálculos foram executados em computador através da elaboração de um programa no qual as razões de funções beta indicadas foram avaliadas recorrendo ao resultado expresso na proposição C.1.2 (expressão (23) do apêndice C).

Como o número usado de observações não é pequeno podemos alternativamente obter valores aproximados desses momentos através da aplicação do método sugerido por Tierney and Kadane (1986), o qual tem a vantagem de reduzir consideravelmente o tempo de computação.

Calculou-se ainda a moda à posteriori através da solução iterativa de (5.70), partindo-se dos seguintes valores iniciais:

- vector \underline{p}_1 das proporções observadas na tabela completa;
- estimador de MV, moda à posteriori e média à posteriori de $\underline{\theta}$, excluindo a 3ª tabela, definidos explícita e respectivamente por (3.35), (5.82) e (5.80);
- estimador de MV de $\underline{\theta}$ dado em 4.5 e respectiva média à priori.

Em todos os casos se obtiveram os mesmos valores terminais, quer pelo algoritmo EM quer pelo método de Newton-Raphson. Este último revelou-se constante e incomparavelmente mais rápido que o EM, como se esperava.

A tabela 5.4.I.3 regista as estimativas procuradas. Na 1ª linha do seu corpo, para cada priori, estão os valores exactos das médias e dos desvios padrões, com os respectivos valores aproximados (método de Laplace) entre parênteses. Na 3ª linha surgem as componentes da moda à posteriori de $\underline{\theta}$.

TABELA 5.4.I.3: Resultados para mecanismo ignorável

	médias e moda à posteriori				desvios padrões à posteriori			
	θ_{11}	θ_{12}	θ_{21}	π	θ_{11}	θ_{12}	θ_{21}	π
Priori 1	0.426	0.189	0.238	-0.049	0.052	0.029	0.037	0.069
	(0.427)	(0.190)	(0.239)	(-0.049)	(0.057)	(0.049)	(0.051)	(0.068)
	0.435	0.186	0.241					
Priori 2	0.388	0.207	0.236	-0.030	0.032	0.015	0.021	0.066
	(0.390)	(0.208)	(0.238)	(-0.030)	(0.049)	(0.043)	(0.043)	(0.062)
	0.395	0.206	0.239					

3. Por fim, a restrição aos dados completos e a decorrente aplicação da conhecida metodologia Bayesiana conjugada conduziram aos resultados reproduzidos na tabela 5.4.I.4 segundo uma ordem idêntica à da tabela anterior.

TABELA 5.4.I.4: Resultados para a sub-amostra completa

	médias e moda à posteriori				desvios padrões à posteriori			
	θ_{11}	θ_{12}	θ_{21}	π	θ_{11}	θ_{12}	θ_{21}	π
Priori 1	0.410	0.205	0.231	-0.026	0.078	0.064	0.067	0.104
	0.429	0.200	0.229					
Priori 2	0.349	0.222	0.238	-0.016	0.060	0.052	0.053	0.085
	0.356	0.220	0.237					

4. A análise dos resultados que constam das tabelas anteriores permite constatar os seguintes aspectos:

- Os desvios padrões das funções lineares de θ registados nas tabelas 5.4.I.3 e 5.4.I.4 são menores quando associados à priori 2 e à abordagem com mecanismo ignorável. Este resultado é natural já que essa distribuição à priori traduz opiniões mais precisas e a referida abordagem envolve uma maior informação amostral decorrente do uso de 120 observações adicionais. A comparação com os valores correspondentes para a abordagem com mecanismo não ignorável torna-se difícil devido à incorporação, neste caso, de novos elementos ligados à distribuição à priori para as probabilidades condicionais de registo.

- Os valores das médias à posteriori obtidos por expansões multinomiais e pelo método de Laplace evidenciam uma boa concordância. O método aproximado parece reflectir uma tendência de sobrestimação das médias e dos desvios padrões à posteriori dos elementos de θ .

- No que diz respeito à questão de interesse, a magnitude dos valores da média e do desvio padrão à posteriori de π em todas as abordagens parece apontar para a validade da hipótese de homogeneidade marginal. A comparação desses valores nas tabelas 5.4.I.3 e 5.4.I.4 revela que a inclusão das tabelas suplementares conduz a uma duplicação aproximada da média de π que não parece, contudo, ser suficiente para pôr em causa a referida homogeneidade. Já a primeira metodologia

parece reforçar esta hipótese. Deve ainda notar-se a influência que as distribuições à priori usadas exercem no sentido da homogeneidade marginal, mormente quando se usam apenas os dados completos. De qualquer modo, o uso de uma priori não informativa (e.g., na acepção de Jeffreys) não parece suscitar conclusões diferentes.

Deste modo, estas soluções Bayesianas apontam no mesmo sentido das soluções clássicas descritas no capítulo 4, as quais se apoiam em hipóteses mais restritivas sobre o mecanismo de registo.

EXEMPLO II: Os dados deste problema resultaram de um estudo realizado no Hospital das Clínicas da Universidade de São Paulo, envolvendo 100 crianças com obstrução biliar. A experiência consistia em classificar cada criança de acordo com o resultado positivo (nível 1) ou negativo (nível 2) de dois testes clínicos e com o tipo de obstrução biliar, intra-hepática ($X_3 = 1$) ou extra-hepática ($X_3 = 2$). Contudo, por razões que se desconhecem totalmente, apenas 93 crianças foram classificadas completamente. O resultado X_2 do 2º teste não foi conhecido para as restantes 7 crianças que, por sinal, responderam positivamente ($X_1 = 1$) ao 1º teste. As frequências observadas $\tilde{N}_1 = (n_{ijk}^{(1)})$ e $\tilde{N}_2 = (n_{ijk}^{(2)})$ estão expostas na tabela 5.4.II.1:

TABELA 5.4.II.1: Dados sobre 100 crianças com obstrução biliar

		$X_3 = 1$		$X_3 = 2$		
		X_2		X_2		
$j \backslash i$		1	2	1	2	
X_1	1	28	12	2	9	93
	2	1	4	2	35	
						7

		X_3		
		1	2	
$k \backslash i$		1	2	
X_1	1	7	0	7
	2	0	0	

A análise destes dados pretende responder a questões do tipo: Existe alguma relação de independência entre as 3 variáveis? Como descrever a estrutura de associação na distribuição conjunta de X_1 , X_2 e X_3 ? Para tal, consideraremos as razões de produtos cruzados

$$\alpha^{(i)} = \frac{\theta_{111} \theta_{122}}{\theta_{112} \theta_{121}}, \quad i=1,2; \quad \beta^{(j)} = \frac{\theta_{1j1} \theta_{2j2}}{\theta_{1j2} \theta_{2j1}}, \quad j=1,2; \quad \gamma^{(k)} = \frac{\theta_{11k} \theta_{22k}}{\theta_{12k} \theta_{21k}}, \quad k=1,2$$

com base nas quais se podem exprimir os parâmetros relevantes dos modelos log-linear.

Como não temos qualquer informação sobre as condições da experiência, admitiremos que as configurações admissíveis dos dados se resumem às configurações observadas. Ou seja, cada elemento populacional ou é completamente categorizado ou sofre apenas censura do valor de X_2 . De acordo com o tratamento teórico do problema, o resultado conjunto dos processos de amostragem e de registro é modelado pela distribuição multinomial.

Admitamos ainda que *à priori* as chances em média de classificação completa versus censura em X_2 são de 3:1 para todo o elemento da população em estudo, mas que a precisão desse juízo é maior para as crianças com o mesmo resultado nos dois testes do que para as restantes. Definindo então tal precisão através do valor

$$\frac{1}{2} b_{ijk}^* = \begin{cases} 4, & i = j \\ 2, & i \neq j \end{cases}, \text{ com } b_{ijk}^* = (b_{ijk}^{i.jk} \ b_{ijk}^{i.k})'$$

a distribuição *à priori* (5.4) fica completamente especificada por

$$\beta_{ijk} = (\lambda_{ijk}^{i.jk} \ \lambda_{ijk}^{i.k})' \rightsquigarrow D_2(b_{ijk}^*)$$

onde

$$b_{ijk}^* = \begin{cases} (3 \ 1) & , i=j \\ (3/2 \ 1/2) & , i \neq j \end{cases}$$

e por $\underline{\theta} \rightsquigarrow D_8(\underline{u})$, $\underline{u} = (4 \ 4 \ 2 \ 2 \ 2 \ 2 \ 4 \ 4)'$, onde $\underline{\theta}$ é o vector dos θ_{ijk} ordenados lexicograficamente.

Frise-se que esta escolha para a distribuição *à priori* serve apenas propósitos ilustrativos, não sendo baseada em quaisquer crenças apriorísticas que sejam do nosso conhecimento.

1. Começando com a análise circunscrita ao grupo de crianças que não sofreram censura, a tabela 5.4.II.2 regista, para efeitos de comparação com os resultados obtidos pelas outras metodologias Bayesianas, alguns parâmetros da distribuição *à posteriori* de $\underline{\theta}$ baseada na *priori* acima indicada.

TABELA 5.4.II.2: Parâmetros da distribuição *à posteriori* com os dados completos

	θ_{111}	θ_{112}	θ_{121}	θ_{122}	θ_{211}	θ_{212}	θ_{221}	θ_{222}
média	0.274	0.051	0.120	0.094	0.026	0.034	0.068	0.333
desvio padrão	0.041	0.020	0.030	0.027	0.015	0.017	0.023	0.043
moda	0.267	0.043	0.112	0.086	0.017	0.026	0.060	0.328

A tabela 5.4.II.3 regista algumas características de funções de θ relacionadas com razões de produtos cruzados. Como se sabe, cada razão de produtos cruzados mede a associação condicional entre um determinado par de variáveis da da uma categoria da 3ª variável. A hipótese de inexistência de interacção de 2ª ordem pode traduzir-se, por exemplo, através de $\alpha^{(1)}/\alpha^{(2)} = 1$ ou da correspondente forma logarítmica. A hipótese de independência condicional $X_1 \perp\!\!\!\perp X_2 | X_3$ corresponde a $\alpha^{(1)} = \alpha^{(2)} = 1$, sendo os restantes dois tipos de independência condicional definidos de modo análogo em termos das razões $\{\beta^{(j)}\}$ e $\{\gamma^{(k)}\}$. Os momentos simples destas funções correspondem a momentos mistos generalizados de θ pelo que o seu cálculo não oferece qualquer dificuldade. As 1ªs 3 linhas da tabela 5.4.II.3 mostram a média e o desvio padrão à posteriori bem como a média à priori.

TABELA 5.4.II.3: Resultados para medidas de associação com os dados completos

$\psi(\theta)$	$\alpha^{(1)}$	$\alpha^{(2)}$	$\beta^{(1)}$	$\beta^{(2)}$	$\gamma^{(1)}$	$\gamma^{(2)}$	$\alpha^{(1)}/\alpha^{(2)}$
média à priori	16	16	2.66	2.66	2.66	2.66	7.11
média à posteriori	9.85	7.80	12.80	7.80	5.42	5.57	2.28
desvio padrão à posteriori	12.11	7.78	19.08	5.08	3.92	6.58	4.21
$E(\log\psi(\theta) \underline{n})$	1.81	1.67	1.96	1.83	1.43	1.30	0.14
$\sigma(\log\psi(\theta) \underline{n})$	0.75	0.73	0.88	0.56	0.60	0.86	1.05

Com a finalidade de clarificarmos a resposta às questões de interesse atrás citadas, recorreremos a uma análise assintótica baseada no resultado de que a distribuição à posteriori conjunta dos contrastes $\sum_{i,j,k} c_{ijk}^{\ell} \log \theta_{ijk}$, $\ell = 1, \dots, p$ ($p < 8$) é aproximadamente normal, com médias $\sum_{i,j,k} c_{ijk}^{\ell} \log(n_{ijk}^{(1)} + u_{ijk})$ e covariâncias $\sum_{i,j,k} c_{ijk}^{\ell} c_{ijk}^m (n_{ijk}^{(1)} + u_{ijk})^{-1}$ (veja-se Lindley [1964, teor. 1]) - note-se que a distribuição à posteriori de θ se pode visualizar como resultado da actualização da priori $D_8(0 \underline{1}_8)$ pela amostra de frequências $\underline{N}_1 + \underline{u}$. As duas últimas linhas da tabela 5.4.II.3 mostram precisamente os parâmetros desta distribuição assintótica para o logaritmo (natural) das funções θ indicadas.

De acordo com esse resultado, a interacção de 2ª ordem não é, de modo nenhum, significativa, dado que o valor 0 de $\log(\alpha^{(1)}/\alpha^{(2)})$ está contido no intervalo de credibilidade a 95%. Pelo mesmo raciocínio conclui-se que a única interacção relativa às tabelas bidimensionais condicionais que se mostra não significativa é aquela referente a $\ln \gamma^{(2)}$.

Constatada a não significância da interação de 2ª ordem, interessa verificar as hipóteses de independência condicional, cada uma delas traduzível pelo anulamento simultâneo da interação de 2ª ordem e de uma interação de 1ª ordem. Tais hipóteses podem ser equivalentemente definidas pelo anulamento do logaritmo de duas razões de produtos cruzados. Por exemplo, $X_1 \perp\!\!\!\perp X_2 | X_3$ corresponde a $\log \alpha^{(1)} = \log \alpha^{(2)} = 0$. Observe-se que o resultado distribucional acima indicado implica a independência à posteriori assintótica entre $\alpha^{(1)}$ e $\alpha^{(2)}$ (bem como entre os dois $\beta^{(j)}$ e entre os dois $\gamma^{(k)}$). A aplicação do raciocínio supra-mencionado permite concluir que $\log \alpha^{(1)}$ e $\log \alpha^{(2)}$ conjuntamente são fortemente significativos e que resultados similares ocorrem para os outros pares de razões de produtos cruzados. Estas conclusões decorrem do carácter significativo das interações de 1ª ordem, evidenciável pela análise da distribuição à posteriori assintótica de $\log(\alpha^{(1)} \alpha^{(2)})$, $\log(\beta^{(1)} \beta^{(2)})$ e $\log(\gamma^{(1)} \gamma^{(2)})$. Estes resultados implicam assim a violação de relações de independência mais fortes.

As conclusões anteriores estão em consonância com os resultados das metodologias de MV e de MQP. Observe-se, no entanto, que a distribuição à priori escolhida também contribui de algum modo para tal afinidade dado o seu peso relativo ($24/117 \times 100 \cong 20\%$) e a forte associação entre X_1 e X_2 para cada nível de X_3 que ela traduz.

Em suma, a análise Bayesiana dos dados completos aponta no sentido das seguintes conclusões:

- A chance de um 2º teste positivo para qualquer criança com um 1º teste positivo é significativamente maior do que a mesma chance para uma criança com um 1º teste negativo;
- A chance de obter um 1º teste positivo para uma criança com obstrução intra-hepática é significativamente maior, independentemente do resultado do 2º teste, do que a mesma chance para uma criança com obstrução extra-hepática;
- A chance de uma criança ter obstrução intra-hepática com um 2º teste positivo é maior do que a mesma chance dado um 2º teste negativo, independentemente do 1º teste.

2. Analisaremos agora todos os dados observados no âmbito da metodologia que ignora o mecanismo de registo. Dado que os dados censurados estão configurados numa partição (indicada pela matriz $Z_2 = (I_2 \otimes I_2) \otimes I_2$), as médias e os desvios padrões à posteriori dos elementos de θ são dados em forma explícita por (5.80) e (5.81). Estes valores estão registados na tabela 5.4.II.4, a qual inclui ainda as componentes da moda à posteriori e da "moda" modificada, calculadas respectivamente por (5.82) e (5.83). Como a única componente não nula de N_2 é $n_{11}^{(2)}$, só as

componentes da moda relativas a θ_{111} e θ_{121} sofrem modificação relativamente à moda respeitante aos dados completos. A diferença nos desvios padrões à posteriori também é compreensivelmente diminuta dado o pequeno número de observações parcialmente categorizadas.

TABELA 5.4.II.4: Parâmetros de distribuição à posteriori sob mecanismo ignorável

	θ_{111}	θ_{112}	θ_{121}	θ_{122}	θ_{211}	θ_{212}	θ_{221}	θ_{222}
média	0.297	0.048	0.130	0.089	0.024	0.032	0.065	0.315
desvio padrão	0.042	0.019	0.032	0.025	0.014	0.016	0.022	0.042
moda	0.310	0.043	0.130	0.086	0.017	0.026	0.060	0.328
"moda" modificada	0.305	0.044	0.128	0.089	0.019	0.026	0.065	0.324

O cálculo dos momentos à posteriori das funções multiplicativas de θ indicadas na tabela 5.4.II.3 efectua-se através da expressão (5.79). A tabela 5.4.II.5 regista os valores respectivos da média e do desvio padrão à posteriori. A sua análise revela que apenas os momentos relativos aos $\{\beta^{(j)}\}$ variam face aos respectivos valores da tabela 5.4.II.3. Uma explicação imediata é sugerida pela própria fórmula usada no seu cálculo, já que quando o expoente \underline{r} é tal que $\underline{Z}_2' \underline{r} = 0$, a expressão (5.79) define os respectivos momentos quando se usa apenas a tabela completamente categorizada. Por exemplo, para a média de $\alpha^{(1)}$, $\underline{r} = (1 \ 0 \ -1 \ 0 \ -1 \ 0 \ 1 \ 0)'$ e para a média de $\alpha^{(1)}/\alpha^{(2)}$, $\underline{r} = (1 \ -1 \ -1 \ 1 \ -1 \ 1 \ 1 \ -1)'$. Logo, aquela condição é satisfeita contrariamente ao que sucede para a média de $\beta^{(1)}$, por exemplo, em que $\underline{r} = (1 \ -1 \ 0 \ 0 \ -1 \ 1 \ 0 \ 0)'$.

A coincidência de resultados para as funções mencionadas é, no fundo, consequência das respectivas distribuições à posteriori serem as mesmas nas duas abordagens. De facto, a caracterização (5.78) da distribuição à posteriori de θ , $\mathcal{D}_8^4(\underline{L}, \underline{Z}_2, \underline{N}_2)$, onde $\underline{L} = \underline{N}_1 + \underline{u}$, revela que os vectores $\underline{Z}_2' \theta = (\theta_{1.1} \theta_{1.2} \theta_{2.1} \theta_{2.2})'$ e $\rho_{i.k} = (\rho_{i.k}^j, j=1,2)'$, $i,k=1,2$, onde $\rho_{i.k} = \theta_{ijk}/\theta_{i.k}$, são independentemente distribuídos segundo as distribuições $D_4(\underline{Z}_2' \underline{L} + \underline{N}_2)$ e $D_2((L_{ijk}, j=1,2)')$, respectivamente. Estas distribuições beta para as probabilidades amostrais condicionais a (X_1, X_3) são precisamente as mesmas que se obtêm para os dados completos, onde a distribuição à posteriori de θ é $D_8(\underline{L})$. Ora, qualquer das funções acima citadas, à excepção dos $\{\beta^{(j)}\}$, exprime-se apenas em função dos $\{\rho_{i.k}\}$. Com efeito, é fácil constatar que

$$\alpha^{(k)} = \rho_{1.k}^1 \rho_{2.k}^2 / \rho_{1.k}^2 \rho_{2.k}^1, \quad \gamma^{(i)} = \rho_{i.1}^1 \rho_{i.2}^2 / \rho_{i.1}^2 \rho_{i.2}^1$$

mas

$$\beta^{(j)} = (\rho_{1.1}^j \rho_{2.2}^j / \rho_{1.2}^{(j)} \rho_{2.1}^{(j)}) \times (\theta_{1.1} \theta_{2.2} / \theta_{1.2} \theta_{2.1})$$

TABELA 5.4.II.5: Resultados para medidas de associação sob mecanismo de registo ignorável

	$\alpha^{(1)}$	$\alpha^{(2)}$	$\beta^{(1)}$	$\beta^{(2)}$	$\gamma^{(1)}$	$\gamma^{(2)}$	$\alpha^{(1)} / \alpha^{(2)}$
média à posteriori	9.85	7.80	14.75	8.99	5.42	5.57	2.28
desvio padrão à posteriori	12.11	7.78	21.94	5.83	3.92	6.58	4.21

Em virtude das considerações anteriores, o argumento em torno dos resultados distribucionais assintóticos implica a manutenção das conclusões extraídas da análise dos dados completos sobre a inexistência da interação de 2ª ordem e o carácter significativo da associação condicional de X_1 e X_2 dado X_3 e de X_2 e X_3 dado X_1 .

Sobre a hipótese de independência condicional entre X_1 e X_3 dado X_2 , devemos observar que a variação distribucional nos $\{\beta^{(j)}\}$ implica designadamente que as duas funções $\log \beta^{(j)}$ passem a ser correlacionadas à posteriori. De facto, cada $\log \beta^{(j)}$ exprime-se como uma soma de dois termos independentes, onde um deles (o logaritmo da razão de produtos cruzados da tabela marginal (X_1, X_3)) é comum às duas funções. A análise distribucional do vector $(\log \beta^{(1)}, \log \beta^{(2)})$, baseada no comportamento assintótico de cada termo Dirichlet, à luz do resultado mencionado de Lindley (1964), leva-nos a concluir também pela significância de associação condicional de X_1 e X_3 dado X_2 . Esta conclusão pode ser considerada previsível a partir da constatação do aumento, face aos resultados para os dados completos, da média à posteriori de cada $\beta^{(j)}$ e do aumento quase proporcional do respectivo desvio padrão - note-se que o coeficiente de variação à posteriori é sensivelmente o mesmo.

3. Considerando por fim a abordagem Bayesiana com mecanismo de registo não ignorável, os valores das médias e dos desvios padrões à posteriori dos elementos de $\underline{\theta}$ foram calculados a partir de (5.42) e (5.46) e encontram-se registados na tabela 5.4.II.6:

TABELA 5.4.II.6: Parâmetros da distribuição à posteriori de θ sob mecanismo de registo não ignorável

	θ_{111}	θ_{112}	θ_{121}	θ_{122}	θ_{211}	θ_{212}	θ_{221}	θ_{222}
média	0.296	0.048	0.132	0.089	0.024	0.032	0.065	0.314
desvio padrão	0.045	0.019	0.034	0.025	0.014	0.016	0.022	0.042

A análise destes valores deixa transparecer uma estreita concordância com os da tabela 5.4.II.4. No que concerne às médias, este aspecto poderá ser antecipado se procedermos à comparação de (5.42) e (5.80) e notarmos que

$b_{1j1}^{1.1} / \sum_{j=1,2} b_{1j1}^{1.1}$ não difere substancialmente de $L_{1j1} / \sum_{j=1,2} L_{1j1}$, onde

$$\underline{L} = (L_{ijk}) = \underline{N}_1 + \underline{u}.$$

Os momentos à posteriori das razões de produtos cruzados (e funções delas) foram obtidos por (5.48), ou mais concretamente, por

$$E\left\{ \prod_{i,j,k} \theta_{ijk}^{r_{ijk}} \mid \underline{N}_1, \underline{N}_2 \right\} = \sum_{m=0}^7 \binom{7}{m} \frac{B(\underline{b}^{1.1} + \underline{y}^{1.1})}{B(\underline{b}^{1.1})} \frac{B(\underline{L} + \underline{y}_2 + \underline{r})}{B(\underline{L} + \underline{y}_2)}$$

onde $\underline{b}^{1.1} = (b_{111}^{1.1} \ b_{121}^{1.1})'$, $\underline{y}^{1.1} = (m \ 7-m)'$, $\underline{y}_2 = (m \ 0 \ 7-m \ 0 \ 0 \ 0 \ 0 \ 0)'$ e \underline{r} é o vector dos r_{ijk} em ordenação lexicográfica. A tabela 5.4.II.7 regista as médias e os desvios padrões das funções que têm sido objecto de análise.

TABELA 5.4.II.7: Resultados para medidas de associação sob mecanismo de registo não ignorável

	$\alpha^{(1)}$	$\alpha^{(2)}$	$\beta^{(1)}$	$\beta^{(2)}$	$\gamma^{(1)}$	$\gamma^{(2)}$	$\alpha^{(1)}/\alpha^{(2)}$
média à posteriori	9.85	7.80	14.67	9.10	5.42	5.57	2.28
desvio padrão à posteriori	12.39	7.78	21.87	6.02	4.07	6.58	4.28

Devemos notar desde já que os momentos indicados de $\alpha^{(2)}$ e $\gamma^{(2)}$ são exactamente iguais aos registados na tabela 5.4.II.5 devido a que $B(\underline{L} + \underline{y}_2 + \underline{r}) / B(\underline{L} + \underline{y}_2) = B(\underline{L} + \underline{r}) / B(\underline{L})$, para todo o $m=0,1,\dots,7$. A semelhança estreita dos valores das tabelas 5.4.II.7 e 5.4.II.5 faz-nos crer que se devam manter as conclusões extraídas nas outras abordagens.

Em suma, qualquer das três análises Bayesianas focadas parece apontar no sentido da estrutura de associação entre os resultados dos dois testes clínicos e o tipo de obstrução biliar ser descrita, de forma satisfatória, pela inexistência, unicamente, da interação de 2ª ordem. A

oOo

APÊNDICE A: ALGUNS RESULTADOS SOBRE ESPAÇOS VECTORIAIS

Consideremos uma transformação linear de \mathbb{R}^p em \mathbb{R}^q , $p \leq q$, representada, relativamente a sistemas de coordenadas fixos pela matriz $q \times p$ \underline{A} com vectores coluna \underline{a}_i , $i = 1, \dots, p$. Denotamos o espaço contradomínio de \underline{A} por $\underline{M}(\underline{A})$ i.e., $\underline{M}(\underline{A}) = \{\underline{u} \in \mathbb{R}^q : \underline{u} = \underline{A}\underline{\beta}$ para algum $\underline{\beta} \in \mathbb{R}^p\}$. A sua dimensão, denotada por $\dim \underline{M}(\underline{A})$, é dada pelo número de vectores coluna (ou linha) linearmente independentes da matriz geradora, i.e. $\dim \underline{M}(\underline{A}) = r(\underline{A})$ onde $r(\underline{A})$ representa o posto de \underline{A} . Suponhamos que o subespaço de \mathbb{R}^q gerado por \underline{A} (i.e., pelas colunas de \underline{A}) tenha dimensão r ($r \leq p$). Então, qualquer submatriz de \underline{A} formada por r colunas linearmente independentes constitui uma matriz base de $\underline{M}(\underline{A})$.

O complemento ortogonal de $\underline{M}(\underline{A})$ é denotado por $[\underline{M}(\underline{A})]^\perp$, isto é $[\underline{M}(\underline{A})]^\perp = \{\underline{v} \in \mathbb{R}^q : \underline{v}'\underline{u} = 0, \forall \underline{u} \in \underline{M}(\underline{A})\}$. Constitui um subespaço de \mathbb{R}^q de dimensão igual a $q - \dim \underline{M}(\underline{A}) = q - r$ e o seu complemento ortogonal coincide com $\underline{M}(\underline{A})$.

O espaço nulo de \underline{A} é o subespaço de \mathbb{R}^p definido por $\underline{N}(\underline{A}) = \{\underline{\alpha} \in \mathbb{R}^p : \underline{A}\underline{\alpha} = \underline{0}_{(q)}\}$. A sua dimensão define a chamada nulidade de \underline{A} e é tal que $\dim \underline{N}(\underline{A}) = p - r(\underline{A})$.

Proposição A.1: $[\underline{M}(\underline{A})]^\perp = \underline{N}(\underline{A}')$ e $[\underline{N}(\underline{A})]^\perp = \underline{M}(\underline{A}')$.

Dem: Por definição, $\underline{N}(\underline{A}')$ é o subespaço de \mathbb{R}^p formado por todos os vectores \underline{v} tais que $\underline{A}'\underline{v} = \underline{0}_{(p)}$, i.e., ortogonais a toda a linha de \underline{A}' (coluna de \underline{A}). Supondo $\underline{v} \in \underline{N}(\underline{A}')$, temos então $\underline{v}'\underline{A}\underline{\beta} = 0 \forall \underline{\beta} \in \mathbb{R}^p$, o que significa que

$\underline{v} \in [\underline{M}(\underline{A})]^\perp$. $\underline{N}(\underline{A}')$ e $[\underline{M}(\underline{A})]^\perp$ são assim dois subespaços de \mathbb{R}^p tais que $\underline{N}(\underline{A}') \subset [\underline{M}(\underline{A})]^\perp$. Toda a base de $\underline{N}(\underline{A}')$ pode assim ser completada de modo a obter-se uma base de $[\underline{M}(\underline{A})]^\perp$. Como $\dim \underline{N}(\underline{A}') = p - r(\underline{A}') = p - r(\underline{A}) = \dim [\underline{M}(\underline{A})]^\perp$, qualquer base de $[\underline{M}(\underline{A})]^\perp$ não pode ter mais elementos que qualquer base de $\underline{N}(\underline{A}')$ e portanto, $\underline{N}(\underline{A}') = [\underline{M}(\underline{A})]^\perp$.

Se aplicarmos esta relação a \underline{A}' em vez de \underline{A} , obtemos $[\underline{M}(\underline{A}')]^\perp = \underline{N}(\underline{A})$ donde se obtém a relação desejada, considerando os complementos ortogonais dos dois subespaços. Δ

Saliente-se que o uso deste resultado permite-nos provar que $\underline{M}(\underline{A}'\underline{A}) = \underline{M}(\underline{A}')$, pelo facto de $\underline{A}\underline{\alpha} = \underline{0}_{(q)}$ sse $\underline{A}'\underline{A}\underline{\alpha} = \underline{0}_{(p)}$ para todo o $\underline{\alpha} \in \mathbb{R}^p$.

Proposição A.2: $\underline{M}(\underline{A}) = \underline{N}(\underline{W}')$ onde \underline{W} é uma matriz $q \times (q-r)$ tal que $\underline{A}'\underline{W} = \underline{0}_{(p, q-r)}$, com $r = r(\underline{A})$.

Dem. Sejam \underline{a}_i , $i=1, \dots, r$ r colunas linearmente independentes de \underline{A} e determinemos $q-r$ vectores linearmente independentes de \mathbb{R}^q , $\underline{\omega}_j$, $j=1, \dots, q-r$ de modo que

$a'_{i-j}\omega_j = 0, i=1, \dots, r; j=1, \dots, q-r$. Então, $S = \{\omega_j, j=1, \dots, q-r\}$ constitui uma base do subespaço $N(A')$. Se $\underline{v} \in N(A')$, existe então um vector de \mathbb{R}^{q-r} ,

$$\underline{\beta} = (\beta_1, \dots, \beta_{q-r})' \text{ tal que } \underline{v} = \sum_{j=1}^{q-r} \beta_j \omega_j = \underline{W}\underline{\beta}, \text{ onde } \underline{W} \text{ é a matriz cujas colu-}$$

nas são os elementos de S , satisfazendo portanto a relação $\underline{A}'\underline{W} = \underline{0}_{(p, q-r)}$. Assim, $N(A') \subset M(W)$. Reciprocamente, se \underline{v} pertence ao subespaço $(q-r)$ -dimensional de \mathbb{R}^q gerado por uma matriz \underline{W} de posto completo cujas colunas são ortogonais às colunas de \underline{A} , então

$$\underline{v} = \underline{W}\underline{\beta} \text{ para algum } \underline{\beta} \in \mathbb{R}^{q-r} \implies \underline{A}'\underline{v} = \underline{A}'\underline{W}\underline{\beta} = \underline{0}_{(p)} \implies \underline{v} \in N(A')$$

e por conseguinte, $N(A') = M(W)$. Conclui-se assim que os complementos ortogonais de $N(A')$ e $M(W)$ coincidem, ou seja, pela proposição A.1, $M(A) = N(W')$, com \underline{W} uma matriz base do complemento ortogonal de $M(A)$. Δ

Proposição A.3: O sistema $\begin{cases} \underline{A}\underline{x} = \underline{u} \\ \underline{H}\underline{x} = \underline{0} \end{cases}$, onde \underline{A} e \underline{H} denotam matrizes de dimensão

respectivamente $q \times p$ e $t \times p$, é consistente e tem solução única $\forall \underline{u} \in M(A)$ sse $M(H') \cap M(A') = \{0\}$ e o posto de $\begin{pmatrix} \underline{A} \\ \underline{H} \end{pmatrix}$ é igual a p .

Demonstração: Façamos $\underline{G} = (\underline{A}' \ \underline{H}')'$ e $\underline{u}_0 = (\underline{u}' \ 0)'$ com $\underline{u} \in M(A)$. Da teoria dos sistemas lineares sabemos que o sistema não homogêneo $\underline{G}\underline{x} = \underline{u}_0$ é consistente sse $\underline{u}_0 \in M(G)$ e tem solução única se adicionalmente o posto de \underline{G} é igual a p . Como $N(G')$ é o complemento ortogonal de $M(G)$, então $\underline{u}_0 \in M(G)$ sse $\underline{z}'\underline{u}_0 = 0, \forall \underline{z} \in N(G')$. Dada a partição de \underline{u}_0 , seja $\underline{z} = (\underline{z}'_1 \ \underline{z}'_2)'$, com \underline{z}'_1 e \underline{z}'_2 de dimensões respectivamente $(q \times 1)$ e $(t \times 1)$, a correspondente partição de \underline{z} . Assim $\underline{G}\underline{x} = \underline{u}_0$ é consistente $\forall \underline{u} \in M(A)$ sse $\underline{z}'_1 \underline{u} = 0$ para todo o $\underline{z} = (\underline{z}'_1 \ \underline{z}'_2)'$ tal que $\underline{G}'\underline{z} = \underline{0}_{(p)}$, i.e., tal que $\underline{A}'\underline{z}'_1 = -\underline{H}'\underline{z}'_2$. Note-se que $\underline{A}'\underline{z}'_1 \in M(A')$ e $-\underline{H}'\underline{z}'_2 \in M(H')$, pelo que para todo o $\underline{z} \in N(G')$, $\underline{A}'\underline{z}'_1 \in M(A') \cap M(H')$. Suponhamos que $M(H') \cap M(A') = \{0\}$. Então $\underline{A}'\underline{z}'_1 = \underline{0}_{(p)} \iff \underline{z}'_1 \in N(A') \iff \underline{z}'_1 \underline{u} = 0, \forall \underline{u} \in M(A)$, o que implica, pelo resultado acima, que o sistema $\underline{G}\underline{x} = \underline{u}_0$ é consistente, $\forall \underline{u} \in M(A)$.

Suponhamos agora que $M(H') \cap M(A') \neq \{0\}$, isto é, que existe pelo menos um vector não nulo de \mathbb{R}^p pertencendo a $M(H')$ que é simultaneamente vector de $M(A')$. Sendo $\underline{\lambda}$ um desses vectores, então $\exists \omega_1 \in \mathbb{R}^q: \underline{\lambda} = \underline{A}'\omega_1$ e $\exists \omega_2 \in \mathbb{R}^t: \underline{\lambda} = \underline{H}'(-\omega_2)$. Fazendo $\underline{\omega} = (\omega'_1 \ \omega'_2)'$, então a definição de $\underline{\lambda}$ implica que $\underline{\omega} \in N(G')$. Consideremos $\underline{u}^* \in M(A)$ tal que $\underline{u}^* = \underline{A}\underline{\lambda}$. Ora $\underline{\omega}'(\underline{u}^* 0) = \omega'_1 \underline{A}\underline{\lambda} = \underline{\lambda}'\underline{\lambda} \neq 0$. Ou seja, existe um $\underline{\omega} \in N(G')$ e um $\underline{u}^* \in M(A)$ tal que $\omega'_1 \underline{u}^* \neq 0$, o que significa então que não existe solução para o sistema $\underline{G}\underline{x} = (\underline{u}^* 0)'$. Ficou assim demonstrado que a consistência de $\underline{G}\underline{x} = (\underline{u}' 0)'$, $\forall \underline{u} \in M(A)$ implica $M(H') \cap M(A') = \{0\}$.

O argumento do início da demonstração deixa então claro que a proposição está completamente provada. Δ

Observação: Se o posto de \underline{A} é $r < p \leq q$, o posto de \underline{G} é p sse o posto de \underline{H} é $p-r$, quando o sistema é consistente. Isto significa que $t-(p-r)$ linhas de \underline{H} são linearmente dependentes e assim redundantes. Por este motivo, podemos considerar \underline{H} como uma matriz $(p-r) \times p$ e, deste modo, a proposição estabelece que o sistema $\begin{pmatrix} \underline{A} \\ \underline{H} \end{pmatrix} \underline{x} = \begin{pmatrix} \underline{u} \\ \underline{0} \end{pmatrix}$ é consistente e de solução única, para todo o elemento \underline{u} do subespaço $M(\underline{A})$ r -dimensional sse $M(\underline{H}') \cap M(\underline{A}') = \{ \underline{0} \}$ e \underline{H} é de posto completo.

oOo

APÊNDICE B

APÊNDICE B1 - FUNÇÕES, PARTIÇÕES, ESTATÍSTICAS E σ -ÁLGEBRAS

1) Seja (Ξ, \mathcal{B}) um espaço mensurável e $\psi: \Xi \rightarrow \Psi$ uma função sobrejectiva arbitrária. A função ψ induz em Ξ uma partição (i.e., um conjunto de subconjuntos de Ξ mutuamente exclusivos e cuja união perfaz Ξ), denotada por $\Pi(\psi)$, cujos elementos são $\{\psi^{-1}(\alpha)\} = \{\theta: \psi(\theta) = \alpha\}$, $\alpha \in \Psi$. Por outro lado, cada partição Π pode ser vista como induzida pela função $\phi_{\Pi}: \Xi \rightarrow \Pi$ que atribui a cada θ o elemento de Π contendo θ , denotado por π_{θ} . Qualquer outra função indutora de Π tem o seu contradomínio em correspondência biunívoca com o de ϕ_{Π} . Esta correspondência entre partições e funções leva a que a relação de redução entre cada par de partições esteja associada com a existência de uma função relacionando cada par de funções. Mais explicitamente, se $\Pi < \Pi'$ então $\exists \rho: \phi_{\Pi} = \rho \circ \phi_{\Pi'}$, e reciprocamente, se $\phi = \rho \circ \psi$ então $\Pi(\phi) < \Pi(\psi)$.

2) Se ψ é uma transformação mensurável em relação a uma σ -álgebra fixa, \mathcal{B}_1 , em Ψ , a σ -álgebra induzida por ψ é a sub-álgebra (abreviatura de sub- σ -álgebra) de \mathcal{B} , $\mathcal{B}_{\psi} = \psi^{-1}(\mathcal{B}_1) = \{\psi^{-1}(C): C \in \mathcal{B}_1\}$. Reciprocamente, se \mathcal{C} é uma sub-álgebra de \mathcal{B} , então \mathcal{C} pode ser vista como induzida pela estatística $\psi(\theta) = \theta: (\Xi, \mathcal{B}) \rightarrow (\Xi, \mathcal{C})$.

Se ψ é uma função arbitrária, defina-se no seu contradomínio Ψ a σ -álgebra $\mathcal{B}_1^* = \{C \subset \Psi: \psi^{-1}(C) \in \mathcal{B}\}$. Assim, ψ pode fazer-se corresponder à transformação mensurável $(\Xi, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1^*)$. A σ -álgebra induzida por ψ é então definida por $\mathcal{B}_{\psi}^* = \psi^{-1}(\mathcal{B}_1^*) = \mathcal{B} \cap \psi^{-1}[P(\Psi)]$, onde $P(\Psi)$ é o conjunto de todas as partes de Ψ . Contudo, nem toda a sub-álgebra de \mathcal{B} é induzida por uma função neste sentido. De facto:

Lema B.1.1 (Blackwell): Se \mathcal{C} é uma sub-álgebra própria de \mathcal{B} contendo todos os conjuntos $\{\theta\}$, então \mathcal{C} não pode ser induzida por uma função.

Prova: Suponhamos que $\mathcal{C} = \psi^{-1}(\mathcal{B}_1^*)$ para alguma função sobrejectiva $\psi: \Xi \rightarrow \Psi$. Como $\{\theta\} \in \mathcal{C}$ por hipótese, então existe um $C \in \mathcal{B}_1^*$ tal que $\{\theta\} = \psi^{-1}(C)$. Por definição de \mathcal{B}_1^* , existe um $B \in \mathcal{B}$, tal que $B = \psi^{-1}(C)$, ou seja, tal que $\psi(B) = C$, devido a que ψ é sobrejectiva. Assim, $\{\theta\} = \psi^{-1}[\psi(B)]$. Como $\psi^{-1}[\psi(B)] \supset B$ segue-se que $B = \{\theta\}$, ou seja, $\{\theta\} = \psi^{-1}[\psi(\{\theta\})]$, $\forall \theta \in \Xi$, o que implica que $\psi^{-1}[\psi(B)] = B$, $\forall B \in \mathcal{B}$. Por definição de $\psi^{-1}(\mathcal{B}_1^*)$, segue-se então que $B \in \mathcal{C}$ pelo facto de $\psi^{-1}[\psi(B)] \in \mathcal{C}$. Como $\mathcal{C} \subset \mathcal{B}$ por hipótese, conclui-se que $\mathcal{C} = \mathcal{B}$ o que é uma contradição, já que \mathcal{C} é uma sub-álgebra própria de \mathcal{B} . \square

Um exemplo é dado pela sub-álgebra da σ -álgebra de Borel em \mathbb{R} formada pelos subconjuntos contáveis ou pelos subconjuntos cujos complementares são contáveis. Desta forma, contrariamente à equivalência entre sub-álgebras e estatísticas, a noção de sub-álgebra é mais geral que a de função no sentido em que existem sub-álgebras que não podem ser induzidas por funções pelo modo indicado.

Prop. B.1.2: Seja ψ uma função sobrejectiva entendida como correspondente à transformação $(\mathcal{E}, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1^*)$ e ρ uma função de ψ com contradomínio $\Phi \subset \mathbb{R}$. A função real ρ é \mathcal{B}_1^* -mensurável (i.e. $\rho: (\Psi, \mathcal{B}_1^*) \rightarrow (\Phi, \mathcal{B}_0 \cap \Phi)$ com \mathcal{B}_0 a σ -álgebra de Borel de \mathbb{R}) sse $\rho\psi$ é uma estatística real.

Prova: Sendo $D \in \mathcal{B}_0 \cap \Phi$, $(\rho\psi)^{-1}(D) = \psi^{-1}[\rho^{-1}(D)] \in \mathcal{B}$ se ρ é \mathcal{B}_1^* -mensurável. Por outro lado, se $(\rho\psi)^{-1}(D) \in \mathcal{B}$, $\forall D \in \mathcal{B}_0 \cap \Phi$, conclui-se pela definição de \mathcal{B}_1^* que $\rho^{-1}(D) \in \mathcal{B}_1^*$, c.q.d. Δ

Prop B.1.3: Seja ψ a estatística $(\mathcal{E}, \mathcal{B}) \rightarrow (\Psi, \mathcal{B}_1)$ com $\Psi = \psi(\mathcal{E})$ e ϕ a estatística real $(\mathcal{E}, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_0 \cap \Phi)$ onde $\Phi \subset \mathbb{R}$ e \mathcal{B}_0 é a σ -álgebra dos borelianos de \mathbb{R} . A função ϕ é \mathcal{B}_ψ -mensurável sse existe uma transformação mensurável $\rho: (\Psi, \mathcal{B}_1) \rightarrow (\Phi, \mathcal{B}_0)$ tal que $\phi = \rho\psi$.

Prova: A demonstração da parte suficiente é idêntica à da parte necessária da proposição anterior e a conclusão segue-se da definição de $\mathcal{B}_\psi = \psi^{-1}(\mathcal{B}_1)$.

Para a parte necessária consideremos primeiramente uma função simples \mathcal{B}_ψ -mensurável, i.e., $\phi = \sum_{i=1}^k b_i I_{B_i}$ onde $B_i \in \mathcal{B}_\psi$, $i=1, \dots, k$ formam uma partição de \mathcal{E} .

Por definição de \mathcal{B}_ψ , $\exists C_i \in \mathcal{B}_1: B_i = \psi^{-1}(C_i)$, $i=1, \dots, k$. Seja $\rho_i = I_{C_i}$,

$i=1, \dots, k$ e $\rho = \sum_{i=1}^k b_i I_{C_i}$. Então $\rho\psi = \sum_{i=1}^k b_i (I_{C_i} \circ \psi) = \sum_{i=1}^k b_i I_{\{\psi \in C_i\}} = \sum_{i=1}^k b_i I_{B_i}$ e

ρ é uma função simples definida em Ψ \mathcal{B}_1 -mensurável.

Sendo agora ϕ uma função não negativa \mathcal{B}_ψ -mensurável, sabemos que existe uma sequência crescente de funções simples não negativas $\{\phi_n, n \geq 1\}$ tal que $\phi = \lim_{n \rightarrow \infty} \phi_n$. Pelo argumento anterior, existem funções simples \mathcal{B}_1 -mensuráveis, ρ_n , tal que $\phi_n = \rho_n \psi$, $\forall n \geq 1$. Pela natureza monótona de $\{\phi_n(\theta) = \rho_n[\psi(\theta)], n \geq 1\}$, a sequência $\{\rho_n\}$ é igualmente monótona crescente. Definindo $\rho = \lim_{n \rightarrow \infty} \rho_n$, se o limite existir e $\rho = 0$, no caso contrário, obtém-se $\phi = \rho\psi$ com ρ \mathcal{B}_1 -mensurável.

Finalmente, a proposição mantém-se para qualquer ϕ invocando a conhecida de composição em funções não negativas $\phi = \phi^+ - \phi^-$. Δ

Observações:

i) Esta proposição é igualmente válida se ψ não for sobrejectiva, embora neste caso a função ρ não seja única. Se ϕ admite duas representações diferentes $\rho_1 \circ \psi$ e $\rho_2 \circ \psi$, então ρ_1 e ρ_2 devem concordar em $\psi(\Xi)$ e, portanto, quase em toda a parte relativamente à medida induzida em Ψ a partir da medida eventualmente definida em (Ξ, \mathcal{B}) .

ii) O resultado expresso nas duas proposições anteriores mantêm-se naturalmente para funções ρ com valores em \mathbb{R}^k , $k \geq 1$ com \mathcal{B}_0 designando a σ -álgebra de Borel de \mathbb{R}^k . A proposição B.1.3 mantêm-se ainda válida para qualquer estatística (não necessariamente real) $\phi: (\Xi, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_2)$ desde que \mathcal{B}_2 contenha todos os conjuntos singulares de ϕ (vide Ash [1972, p. 254]). Δ

Dada a correspondência entre partições e funções, a σ -álgebra induzida por uma partição Π , designada por \mathcal{B}_Π , é definida como a σ -álgebra induzida pela função associada ϕ_Π no sentido acima indicado, i.e., $\mathcal{B}_\Pi = \phi_\Pi^{-1}(\mathcal{B}_1^*)$ onde $\mathcal{B}_1^* = \{C \subset \Pi: \phi_\Pi^{-1}(C) \in \mathcal{B}\}$. Dada a definição de ϕ_Π , \mathcal{B}_Π é assim o conjunto dos elementos de \mathcal{B} que se exprimem como uma união de partes de Π . Naturalmente que \mathcal{B}_Π é a σ -álgebra induzida por qualquer função associada a Π . Com efeito, dado $\phi_\Pi: (\Xi, \mathcal{B}) \rightarrow (\Pi, \mathcal{B}_1^*)$ podemos definir uma função bijectiva $\rho: \Pi \rightarrow \Psi$ e uma σ -álgebra em Ψ , \mathcal{B}_1^{**} , tal que $\mathcal{B}_1^{**} = \{A \subset \Psi: \rho^{-1}(A) \in \mathcal{B}_1^*\}$. Pela definição de \mathcal{B}_1^* resulta então que $\mathcal{B}_1^{**} = \{A \subset \Psi: \phi_\Pi^{-1}[\rho^{-1}(A)] \in \mathcal{B}\}$. Fazendo $\psi = \rho \circ \phi_\Pi$ fica definida uma função sobrejectiva $\psi: \Xi \rightarrow \Psi$ cuja σ -álgebra induzida $\mathcal{B}_\psi^* = \psi^{-1}(\mathcal{B}_1^{**})$ coincide com \mathcal{B}_Π .

Se a partição Π é mensurável (i.e., os seus elementos pertencem a \mathcal{B}), a função associada $\phi_\Pi: (\Xi, \mathcal{B}) \rightarrow (\Pi, \mathcal{B}_1^*)$ é agora \mathcal{B} -mensurável em relação a uma σ -álgebra \mathcal{B}_1^* que contém os conjuntos singulares de Π . Em virtude da proposição B.1.3 e da observação ii) subsequente podemos então concluir que:

Prop. B.1.4: Sejam Π_1 e Π_2 duas partições com funções associadas

$\phi_1 \equiv \phi_{\Pi_1}: (\Xi, \mathcal{B}) \rightarrow (\Pi_1, \mathcal{B}_1^*)$ e $\phi_2 \equiv \phi_{\Pi_2}: (\Xi, \mathcal{B}) \rightarrow (\Pi_2, \mathcal{B}_2^*)$. Se Π_1 é mensurável, então ϕ_1 é \mathcal{B}_{Π_2} -mensurável sse existe uma transformação mensurável $\rho: (\Pi_2, \mathcal{B}_2^*) \rightarrow (\Pi_1, \mathcal{B}_1^*)$ tal que $\phi_1 = \rho \circ \phi_2$.

NOTA. A referência explícita a ρ como transformação mensurável pode ser omitida pela definição de ϕ_1 como uma transformação \mathcal{B} -mensurável e pela validade da prop. B.1.2 para funções ρ não necessariamente reais.

3) Vimos acima que toda a partição induz uma sub-álgebra. Contudo, nem toda a sub-álgebra é induzida por uma partição, como o comprova o lema B.1.1. Mas, por outro lado, toda a sub-álgebra induz uma partição. Com efeito, seja \mathcal{C} uma sub-ál-

gebra de \mathcal{B} e defina-se em Ξ a relação binária

$$\theta_1 \tilde{\mathcal{C}} \theta_0 \iff I_{\mathcal{C}}(\theta_1) = I_{\mathcal{C}}(\theta_0), \forall C \in \mathcal{C}$$

onde $I_{\mathcal{C}}(\theta)$ denota a função indicadora do conjunto C .

É fácil verificar que $\tilde{\mathcal{C}}$ traduz uma relação de equivalência. Os pontos θ_1 e θ_0 são equivalentes segundo $\tilde{\mathcal{C}}$ se não são separados por qualquer conjunto de \mathcal{C} . A partição induzida por \mathcal{C} é definida como o conjunto quociente de Ξ segundo $\tilde{\mathcal{C}}$. Designando por π_{θ_0} o elemento de $\Xi/\tilde{\mathcal{C}}$ contendo θ_0 , então $\pi_{\theta_0} \subset C$ ou $\pi_{\theta_0} \subset \Xi - C$ para todo o $C \in \mathcal{C}$, i.e., π_{θ_0} é a intersecção de todos os conjuntos de \mathcal{C} que contêm θ_0 . Por outras palavras, cada elemento de $\Xi/\tilde{\mathcal{C}}$ é um conjunto que não admite subconjuntos próprios \mathcal{C} -mensuráveis, os chamados átomos de \mathcal{C} . Os átomos não são necessariamente elementos de \mathcal{C} dado que podem ser definidos pela intersecção de um número não contável de conjuntos de \mathcal{C} . Contudo, a partição induzida por uma sub-álgebra \mathcal{C} separável, está contida em \mathcal{C} . Com efeito, seja $\{C_n, n \geq 1\}$ um gerador de \mathcal{C} . Então

$$\pi_{\theta_0} = \left(\bigcap_{n: \theta_0 \in C_n} C_n \right) \cap \left(\bigcap_{n: \theta_0 \notin C_n} C_n^c \right)$$

o que ilustra a \mathcal{C} -mensurabilidade de cada átomo.

Naturalmente que a partição induzida por \mathcal{C} , que designaremos por $\Pi(\mathcal{C})$, pode ser induzida por outras sub-álgebras. Por outro lado, a classe de sub-álgebras induzindo uma dada partição Π pode ser vazia. Por exemplo, seja $\Xi = \mathbb{R}$ e $\mathcal{B} = \{\emptyset, \Xi, (-\infty, 0], (0, +\infty)\}$. As únicas sub-álgebras de \mathcal{B} são $\mathcal{C}_1 = \{\emptyset, \Xi\}$ e $\mathcal{C}_2 = \mathcal{B}$. Como $\Pi(\mathcal{C}_1) = \{\Xi\}$ e $\Pi(\mathcal{C}_2) = \{(-\infty, 0], (0, +\infty)\}$, segue-se que $\Pi = \{\{\theta\}, \theta \in \mathbb{R}\}$ não é induzida por qualquer sub-álgebra.

4) As definições de partição induzida por uma sub-álgebra e de sub-álgebra induzida por uma partição permitem estabelecer facilmente que

Prop. B.1.5: i) $\mathcal{C}_1 \subset \mathcal{C}_2 \implies \Pi(\mathcal{C}_1) < \Pi(\mathcal{C}_2)$

$$\text{ii) } \Pi_1 < \Pi_2 \implies \mathcal{B}_{\Pi_1} \subset \mathcal{B}_{\Pi_2}$$

Consideremos agora $\mathcal{C} \subset \mathcal{B}$ e $C \in \mathcal{C}$. Por definição de $\Pi(\mathcal{C})$, qualquer elemento desta partição ou está inteiramente contido em C ou é disjunto de C . Assim, C deve ser uma união de partes de $\Pi(\mathcal{C})$, i.e., $C \in \mathcal{B}_{\Pi(\mathcal{C})}$. Por outro lado, designemos por π_{θ} o elemento de uma dada partição Π_0 contendo θ e por π'_{θ} o elemento de \mathcal{B}_{Π_0} que contém o mesmo ponto θ . Como π'_{θ} é a intersecção de todos aqueles conjuntos de \mathcal{B}_{Π_0} contendo θ , conclui-se que π_{θ} deve estar contido em π'_{θ} (veja-se, p.e., o último exemplo acima indicado), ou seja, π'_{θ} deve ser uma união de partes de Π_0 . Se $\Pi_0 \subset \mathcal{B}$, então π_{θ} é um átomo mensurável de \mathcal{B}_{Π_0} , pelo que $\pi_{\theta} = \pi'_{\theta}$. Deste modo fica demonstrado que

Prop. B.1.6: Sejam Π_0 uma partição de Ξ e \mathcal{C} uma sub-álgebra de \mathcal{B} .

i) $\mathcal{C} \subset \mathcal{B}_{\Pi_0}(\mathcal{C})$

ii) $\Pi(\mathcal{B}_{\Pi_0}) < \Pi_0$, coincidindo as duas partições se Π_0 é \mathcal{B} -mensurável.

A conjugação dos enunciados ii) das duas proposições anteriores permite concluir que:

Corol. B.1.7: A partição \mathcal{B} -mensurável Π_1 é uma redução da partição Π_2 sse

$$\mathcal{B}_{\Pi_1} \subset \mathcal{B}_{\Pi_2} .$$

Deve notar-se que este resultado pode ser também obtido por aplicação da proposição B.1.4; atendendo a que $\Pi_1 < \Pi_2$ sse ϕ_1 for uma função de ϕ_2 , relação esta que doravante denotaremos por $\phi_1 < \phi_2$.

5) Suponhamos agora que (Ξ, \mathcal{B}) é euclideano e nele está definida uma medida de probabilidade ν . Sendo Π_1 e Π_2 duas partições de Ξ , diz-se que Π_1 é ν -essencialmente uma redução de Π_2 - e escreve-se $\Pi_1 < \Pi_2 [\nu]$ - se existe um conjunto $N \in \mathcal{B}$ ν -nulo tal que $\Pi_1 < \Pi_2$ em $\Xi - N$. Do mesmo modo, sendo ϕ_1 e ϕ_2 duas funções definidas em Ξ , diz-se que ϕ_1 é ν -essencialmente uma função de ϕ_2 - o que será denotado por $\phi_1 < \phi_2 [\nu]$ - se existe um conjunto $N \in \mathcal{B}$ ν -nulo tal que $\phi_1 < \phi_2$ em $\Xi - N$. Sendo Π_1 e Π_2 duas partições do espaço euclideano Ξ , a possibilidade de associação dessas partições com funções reais indutoras das σ -álgebras induzidas pelas partições permite-nos usar os resultados de Bahadur (1955) para relacionar os anteriores conceitos de redução ν -essencial com os de inclusão ν -essencial entre as σ -álgebras induzidas (vide definição deste último conceito no ponto 6) deste apêndice. Assim:

Prop. B.1.8: Sejam Π_1 e Π_2 duas partições do espaço euclideano Ξ munido da respectiva σ -álgebra de Borel, \mathcal{B} , e da medida de probabilidade ν .

i) Se $\mathcal{B}_{\Pi_1} \subset \mathcal{B}_{\Pi_2} [\nu]$ e Π_1 é induzida por uma estatística real, então $\Pi_1 < \Pi_2 [\nu]$;

ii) Se $\Pi_1 < \Pi_2 [\nu]$ e Π_2 é induzida por uma estatística real, então $\mathcal{B}_{\Pi_1} \subset \mathcal{B}_{\Pi_2} [\nu]$;

iii) Se Π_1 e Π_2 são induzidas por estatísticas reais, então

$$\Pi_1 < \Pi_2 [\nu] \iff \mathcal{B}_{\Pi_1} \subset \mathcal{B}_{\Pi_2} [\nu].$$

Prova: Sejam ϕ_i , $i=1,2$ funções reais em correspondência biunívoca com ϕ_{Π_i} , $i=1,2$, de modo que as σ -álgebras induzidas por ϕ_i , $i=1,2$ correspondem a \mathcal{B}_{Π_i} , $i=1,2$. Os lemas 6 e 9 de Bahadur (1955) provam respectivamente i) e ii) e, conseqüentemente, iii).

A comparação da proposição B.1.8 com o corolário B.1.7 evidencia que a verificação da equivalência entre os conceitos de " $\subset [v]$ " e de " $\subset [v]$ ", citados exige, em geral, mais requisitos que os exigidos pelos correspondentes conceitos quando definidos exactamente em toda a parte.

6) Seja $(\Omega, \mathcal{F}, \mu)$ um dado espaço de probabilidade. Se $\{\mathcal{F}_\alpha : \alpha \in \Delta\}$ representa uma colecção arbitrária de sub-álgebras de \mathcal{F} , então $\bigvee \mathcal{F}_\alpha$ denota a menor σ -álgebra contendo cada \mathcal{F}_α , $\alpha \in \Delta$, isto é, $\bigvee_{\alpha \in \Delta} \mathcal{F}_\alpha = \sigma(\bigcup_{\alpha \in \Delta} \mathcal{F}_\alpha)$.

Sejam A_1 e A_2 dois conjuntos de \mathcal{F} e f_1 e f_2 duas funções \mathcal{F} -mensuráveis. A relação $A_1 = A_2 [\mu]$ significa que $\mu(A_1 \Delta A_2) = 0$ onde $A_1 \Delta A_2 = (A_1 - A_2) \cup (A_2 - A_1)$. De um modo análogo, $f_1 = f_2 [\mu]$ significa que $\mu(\{\omega \in \Omega : f_1(\omega) \neq f_2(\omega)\}) = 0$. Se em (Ω, \mathcal{F}) estiver definida uma família R de medidas de probabilidade, então $A_1 = A_2 [R]$ e $f_1 = f_2 [R]$ significam que $A_1 \Delta A_2$ e $\{f_1 \neq f_2\}$ são conjuntos nulos para todo o elemento de R (i.e., R -nulos).

Sejam \mathcal{F}_1 e \mathcal{F}_2 duas sub-álgebras de \mathcal{F} . Dizemos que \mathcal{F}_1 está incluída μ -essencialmente (resp. R -essencialmente) em \mathcal{F}_2 - o que denotaremos por $\mathcal{F}_1 \subset \mathcal{F}_2 [\mu]$ (resp. $\mathcal{F}_1 \subset \mathcal{F}_2 [R]$) - se para cada $A_1 \in \mathcal{F}_1$ existe um $A_2 \in \mathcal{F}_2$ tal que $A_1 = A_2 [\mu]$ (resp. $A_1 = A_2 [R]$). Equivalentemente, podemos dizer que para toda a função real $f_1 \in \mathcal{F}_1$ - modo abreviado de referir que f_1 é \mathcal{F}_1 -mensurável - existe uma função real $f_2 \in \mathcal{F}_2$ tal que $f_1 = f_2 [\mu]$ (resp. $f_1 = f_2 [R]$). Frequentemente, esta última asserção é expressa dizendo que toda a função real $f_1 \in \mathcal{F}_1$ é μ -essencialmente (resp. R -essencialmente) \mathcal{F}_2 -mensurável.

O completamento de qualquer sub-álgebra \mathcal{F}_1 é definido por $\overline{\mathcal{F}}_1 = \mathcal{F}_1 \vee \overline{\mathcal{F}}_0$, onde $\overline{\mathcal{F}}_0$ é o completamento da σ -álgebra trivial $\mathcal{F}_0 = \{\emptyset, \Omega\}$, i.e. $\overline{\mathcal{F}}_0 = \{F \in \mathcal{F} : \mu^2(F) = \mu(F)\}$. Dado que $\overline{\mathcal{F}}_0 = \sigma(N)$ onde N é o σ -anel definido por $\{N \in \mathcal{F} : \mu(N) = 0\}$, conclui-se que $\overline{\mathcal{F}}_1 = \sigma(\mathcal{F}_1 \cup \sigma(N)) = \sigma(\mathcal{F}_1 \cup N) = \mathcal{F}_1 \vee N = \{F_1 \Delta N : F_1 \in \mathcal{F}_1, N \in N\}$.

Tendo em conta a definição da igualdade μ -essencial de dois conjuntos, a definição acima de $\overline{\mathcal{F}}_1$ permite constatar que $\overline{\mathcal{F}}_1 = \{F \in \mathcal{F} : F = F_1 [\mu] \text{ para algum } F_1 \in \mathcal{F}_1\}$. Em consequência desta definição, torna-se fácil verificar que $\mathcal{F}_1 \subset \mathcal{F}_2 [\mu]$ se e só se $\mathcal{F}_1 \subset \overline{\mathcal{F}}_2$. Assim, a μ -equivalência entre \mathcal{F}_1 e \mathcal{F}_2 , denotada por $\mathcal{F}_1 = \mathcal{F}_2 [\mu]$, e que significa $\mathcal{F}_1 \subset \mathcal{F}_2 [\mu]$ e $\mathcal{F}_2 \subset \mathcal{F}_1 [\mu]$, corresponde a $\overline{\mathcal{F}}_1 = \overline{\mathcal{F}}_2$. Em particular, para toda a sub-álgebra \mathcal{F}_2 , tem-se $\mathcal{F}_2 = \overline{\mathcal{F}}_2 [\mu]$ e toda a função real f é μ -essencialmente \mathcal{F}_2 -mensurável sse $f \in \overline{\mathcal{F}}_2$.

A definição de $\overline{\mathcal{F}}_1 = \mathcal{F}_1 \vee N$ permite igualmente verificar que $\mathcal{F}_1 \vee (\mathcal{F}_2 \cap \overline{\mathcal{F}}_0) = \overline{\mathcal{F}}_1 \cap \mathcal{F}_2$ quando $\mathcal{F}_1 \subset \mathcal{F}_2$. Ou seja, o completamento de \mathcal{F}_1 com os conjuntos μ -nulos de \mathcal{F}_2 coincide com a σ -álgebra dos conjuntos de \mathcal{F}_2 que são μ -iguais aos conjuntos de \mathcal{F}_1 . Por outras palavras, $\overline{\mathcal{F}}_1 \cap \mathcal{F}_2$ representa a menor sub-álgebra de \mathcal{F}_2 (note-se que $\overline{\mathcal{F}}_1$ não está incluída necessariamente em \mathcal{F}_2) contendo \mathcal{F}_1 e todos os conjuntos μ -nulos de \mathcal{F}_2 . \mathcal{F}_1 diz-se então completa relativamente a \mathcal{F}_2 (ou \mathcal{F}_2 -completa) se $\overline{\mathcal{F}}_1 \cap \mathcal{F}_2 = \mathcal{F}_1$. Considerando $\mathcal{F}_2 = \mathcal{F}$ obtém-se a definição usual de completicidade de qualquer sub-álgebra \mathcal{F}_1 .

APÊNDICE B.2 - ALGUMAS NOTAS SOBRE A ESTRUTURA DO MODELO BAYESIANO

1) Seja (Y, \mathcal{A}, P) o modelo estatístico "clássico" onde P representa a família de medidas de probabilidade P_θ ; $\theta \in \Xi$. Seja \mathcal{B} uma σ -álgebra em Ξ seleccionada de modo que as funções $\theta \rightarrow P_\theta(A)$ para todo o $A \in \mathcal{A}$ sejam \mathcal{B} -mensuráveis. Desta forma, $P_\theta(A): \Xi \times \mathcal{A} \rightarrow [0,1]$ representa uma função de transição de (Ξ, \mathcal{B}) para (Y, \mathcal{A}) .

Denotemos por $\Omega = \Xi \times Y$ o produto cartesiano dos espaços paramétrico e amostral e por $\mathcal{F} = \mathcal{B} \times \mathcal{A}$ a σ -álgebra produto em Ω . Ou seja, \mathcal{F} é a menor σ -álgebra contendo a classe dos rectângulos mensuráveis $S_0 = \{B \times A: B \in \mathcal{B}, A \in \mathcal{A}\}$, que é uma classe π (fechada sob intersecções finitas). Como a família S_1 de uniões finitas de elementos mutuamente disjuntos de S_0 é a menor álgebra contendo S_0 , \mathcal{F} é igualmente a σ -álgebra gerada por S_1 .

O teorema generalizado da medida produto (vide Ash [1972, ps. 97-100]) permite então definir no espaço produto (Ω, \mathcal{F}) uma medida de probabilidade $\mu = \nu \times P_\theta$ definida por

$$\forall F \in \mathcal{F}: \mu(F) = \int_{\Xi} P_\theta(F^\theta) \nu(d\theta) \quad (1)$$

onde F^θ é a secção de F em θ , i.e., $F^\theta = \{y \in Y: \omega = (\theta, y) \in F\}$. A medida produto μ é única no sentido em que é idêntica a qualquer outra medida μ_0 em (Ω, \mathcal{F}) tal que $\mu_0(S) = \mu(S)$, $\forall S \in S_0$.

O espaço de probabilidade assim construído, $(\Omega, \mathcal{F}, \mu)$, é conhecido como modelo estatístico Bayesiano. Nele, as observações y e o parâmetro θ podem ser visualizados através das funções mensuráveis $\bar{y}(\omega) = y: (\Omega, \mathcal{F}) \rightarrow (Y, \mathcal{A})$ e $\bar{\theta}(\omega) = \theta: (\Omega, \mathcal{F}) \rightarrow (\Xi, \mathcal{B})$. As sub-álgebras (designação abreviada de sub- σ -álgebras) de \mathcal{F} induzidas por \bar{y} e por $\bar{\theta}$, denotadas respectivamente por $\mathcal{F}_{\bar{y}}$ e por $\mathcal{F}_{\bar{\theta}}$, são assim as σ -álgebras geradas pelas classes dos cilindros unidimensionais, $\{E \times A: A \in \mathcal{A}\}$ e $\{B \times Y: B \in \mathcal{B}\}$, isto é, $\mathcal{F}_{\bar{y}} = E \times \mathcal{A}$ e $\mathcal{F}_{\bar{\theta}} = \mathcal{B} \times Y$. Notando que $(B \times Y) \cup (E \times A)$, $B \in \mathcal{B}$, $A \in \mathcal{A}$ é a união dos rectângulos mensuráveis disjuntos $B \times A$, $B \times A^c$ e $B^c \times A$, fica evidenciado que $\mathcal{F} \equiv \sigma(S_0) = \mathcal{F}_{\bar{\theta}} \vee \mathcal{F}_{\bar{y}} \equiv \sigma(\{(B \times Y) \cup (E \times A): B \in \mathcal{B}, A \in \mathcal{A}\})$. Doravante, denotaremos por \bar{B} (resp. \bar{A}) o elemento de $\mathcal{F}_{\bar{\theta}}$ (resp. $\mathcal{F}_{\bar{y}}$) que é imagem inversa de $B \in \mathcal{B}$ (resp. $A \in \mathcal{A}$) por $\bar{\theta}$ (resp. \bar{y}), i.e., $\bar{B} = B \times Y$ (resp. $\bar{A} = E \times A$).

Da definição de μ deduz-se que

$$\forall (B \times A) \in S_0, \mu(B \times A) = \int_{\mathcal{B}} P_\theta(A) \nu(d\theta) \quad (2)$$

pelo que $\nu(B) = \mu(\bar{B})$, $\forall B \in \mathcal{B}$, i.e., a medida de probabilidade a priori pode ser identificada com a restrição, $\mu_{\bar{\theta}}$, de μ a $\mathcal{F}_{\bar{\theta}}$. Por outro lado, a medida de probabilidade marginal em (Y, \mathcal{A}) , $P(A) \equiv \int_{\Xi} P_\theta(A) \nu(d\theta)$, identifica-se assim com a restrição $\mu_{\bar{y}}$, de μ a $\mathcal{F}_{\bar{y}}$. A definição de P permite desde já concluir que é dominada pela família P .

Por construção, $P_{\theta}(A)$ define uma função de transição em $\Xi \times \mathbf{A}$. Definindo

$$P_{\bar{\theta}(\omega)}^*(A) = P_{\theta}(A) \quad \text{para } \theta: \bar{\theta}(\omega) = \theta$$

obtemos uma função de transição relacionada em $\Omega \times \mathbf{A}$. A expressão (2) pode assim ser reescrita como

$$\mu(B \times A) = \int_{\bar{B}} P_{\bar{\theta}(\omega)}^*(A) \mu_{\bar{\theta}}(d\omega) \quad (3)$$

Ora, por definição da esperança condicional da função indicadora $I_{\bar{A}}$, $A \in \mathbf{A}$, dado $\mathcal{F}_{\bar{\theta}}$ (apêndice B.3), temos pelo facto de $\bar{B} \cap \bar{A} = B \times A$

$$\forall B \in \bar{B}, \mu(B \times A) \equiv \int_{\bar{B}} I_{\bar{A}}(\omega) \mu(d\omega) = \int_{\bar{B}} \mu(\bar{A} | \mathcal{F}_{\bar{\theta}})(\omega) \mu_{\bar{\theta}}(d\omega) \quad (4)$$

Comparando (3) e (4) obtemos para todo o $A \in \mathbf{A}$

$$P_{\bar{\theta}}^*(A) = \mu(\bar{A} | \mathcal{F}_{\bar{\theta}}) \quad [\mu_{\bar{\theta}}] \quad (5)$$

o que manifesta que $P_{\theta}(A)$ está associada a uma versão regular da restrição a $\mathcal{F}_{\bar{y}}$ da probabilidade condicional dado $\mathcal{F}_{\bar{\theta}}$.

A construção de $P_{\theta}(A)$ como uma função de transição em $\Xi \times \mathbf{A}$ tem assim implícito que o espaço mensurável (Y, \mathbf{A}) é tal que admite a existência de uma probabilidade condicional regular para \bar{y} dado $\mathcal{F}_{\bar{\theta}}$. Este requisito é satisfeito nas situações usuais, onde (Y, \mathbf{A}) é um espaço euclídeano - veja-se o sentido desta designação na sub-secção 2.2.1 - pelo conhecido teorema de Doob (vide p.e. Ash [1972. sec. 6.6]).

Admitindo igualmente que Ξ é euclídeano e \mathbf{B} a respectiva σ -álgebra de Borel, seja então $\nu_{\bar{y}}^*(B)$ uma função de transição em $\Omega \times \mathbf{B}$ definida por

$$\nu_{\bar{y}}^* = \mu(\bar{B} | \mathcal{F}_{\bar{y}}) \quad [\mu_{\bar{y}}] \quad (6)$$

obtida tomando uma versão regular da probabilidade condicional para $\bar{\theta}$ dado $\mathcal{F}_{\bar{y}}$. Dela podemos obter uma função de transição $\nu_y(B)$ de (Y, \mathbf{B}) para (Ξ, \mathbf{B}) definindo

$$\nu_y(B) = \nu_{\bar{y}(\omega)}^*(B) \quad \text{para } \omega: \bar{y}(\omega) = y$$

Por um argumento análogo ao que conduziu a (4) e tendo em conta que $P \equiv \mu_{\bar{y}}$, temos para todo o $B \times A \in S_0$

$$\mu(B \times A) \equiv \int_{\bar{A}} I_{\bar{B}}(\omega) \mu(d\omega) = \int_{\bar{A}} \mu(\bar{B} | \mathcal{F}_{\bar{y}})(\omega) \mu_{\bar{y}}(d\omega) = \int_{\bar{A}} \nu_{\bar{y}(\omega)}^*(B) \mu_{\bar{y}}(d\omega) = \int_{\bar{A}} \nu_y(B) P(dy) \quad (7)$$

Devido à unicidade estabelecida no teorema generalizado da medida produto, conclui-se assim que

$$\forall F \in \mathcal{F}, \mu(F) = \int_{\mathcal{Y}} \nu_y(F^y) P(dy) \quad (8)$$

onde $F^y = \{\theta \in \mathcal{E} : \omega = (\theta, y) \in F\}$ representa a secção de F em y . Ou seja, a medida produto $\mu = \nu \times P_\theta$ admite também a decomposição $\mu = P \times \nu_y$, o que define a regularidade do modelo Bayesiano, para adoptar a terminologia de Florens et al (1983). Assim, sob as asas de um modelo Bayesiano regular, fica definido um novo modelo $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$, onde $\mathcal{Q} = \{\nu_y : y \in \mathcal{Y}\}$ representa a família das medidas de probabilidade à posteriori - o denominado dual Bayesiano de $(\mathcal{Y}, \mathcal{A}, P)$. Neste quadro, de (7) obtemos

$$\nu(B) = \int_{\mathcal{Y}} \nu_y(B) P(dy), \quad \forall B \in \mathcal{B}$$

revelando que $\nu \ll \mathcal{Q}$.

Denotemos por $P_B(\cdot)$, com $B \in \mathcal{B}$ fixo, a medida de probabilidade em $(\mathcal{Y}, \mathcal{A})$ definida em (2). A expressão (7) revela assim que $P_B \ll P$, sendo a respectiva derivada de Radon-Nikodym, $P_B(dy)/P(dy)$, representada por

$$\frac{dP_B(y)}{dP(y)} = \nu_y(B) \quad [P]$$

Reciprocamente, sendo $\nu_A(\cdot)$, com $A \in \mathcal{A}$ fixo, a medida de probabilidade em $(\mathcal{E}, \mathcal{B})$ definida em (7), conclui-se analogamente que $\nu_A \ll \nu$ e

$$\frac{d\nu_A(\theta)}{d\nu(\theta)} = P_\theta(A) \quad [\nu]$$

Assim, numa experiência Bayesiana regular, as funções \mathcal{A} -mensurável, $\nu(\cdot)$, e \mathcal{B} -mensurável, $P(A)$, para cada $B \in \mathcal{B}$ e $A \in \mathcal{A}$ respectivamente, podem ser encaradas como derivadas de Radon-Nikodym em relação a medidas apropriadas.

2) Consideremos agora que $(\mathcal{Y}, \mathcal{A})$ é um espaço euclídeo e que a família \mathcal{P} é dominada por uma medida σ -finita m . Como \mathcal{A} é separável (i.e. gerada por uma classe contável), pode ser escolhida uma versão de dP_θ/dm que é \mathcal{F} -mensurável. Seja $f(\theta, y)$ uma versão da função densidade relativamente a m nessas condições, o que garante a sua \mathcal{A} -mensurabilidade (resp. \mathcal{B} -mensurabilidade) para cada $\theta \in \mathcal{E}$ (resp. para cada $y \in \mathcal{Y}$). Pelo teorema de Fubini, obtemos de (2)

$$\forall (B \times A) \in \mathcal{S}_0, \mu(B \times A) = \int_{B \times A} f(\theta, y) (\nu \times m)(d\theta \times dy) \quad (9)$$

Definamos em $\mathcal{F} = \sigma(\mathcal{S}_0)$ a medida

$$\mu^*(F) = \int_F f(\theta, y) (\nu \times m)(d\theta \times dy)$$

Por (9), $\mu^*(F) = \mu(F)$ em S_0 . Por um argumento em termos do teorema de Dynkin (vide B.3), $\mu^*(F) = \mu(F)$ em F . Isto é, $\mu \ll \nu \times m$ e $d\mu(\theta, y)/d(\nu \times m)(\theta, y) = f(\theta, y)$.

Por outro lado, como $P \ll \mathcal{P}$ segue-se que $P \ll m$ e, pelo teorema de Fubini

$$P(A) = \int_A \left[\int_{\mathcal{E}} f(\theta, y) \nu(d\theta) \right] m(dy) \quad (10)$$

revelando que a função \mathcal{A} -mensurável $\int_{\mathcal{E}} f(\theta, y) \nu(d\theta)$ é uma versão de dP/dm . Consequentemente $\nu \times P \ll \nu \times m$ com

$$\frac{d(\nu \times P)(\theta, y)}{d(\nu \times m)(\theta, y)} = \int_{\mathcal{E}} f(\theta, y) \nu(d\theta) \quad [\nu \times m]$$

Assim, sempre que esta versão for estritamente positiva

$$\frac{d\mu(\theta, y)}{d(\nu \times P)(\theta, y)} = \frac{f(\theta, y)}{\int_{\mathcal{E}} f(\theta, y) \nu(d\theta)} = h(\theta, y) \quad (11)$$

e no conjunto m -excepcional onde é nula, $d\mu/d(\nu \times P)$ pode ser arbitrariamente definida como igual a 1.

Designando por $g(\theta, y)$ qualquer versão de $d\mu/d(\nu \times P)$ resulta pelo teorema de Fubini

$$\forall (B \times A) \in S_0, \mu(B \times A) = \int_B \int_A g(\theta, y) P(dy) \nu(d\theta) = \int_A \int_B g(\theta, y) \nu(d\theta) P(dy) \quad (12)$$

A comparação da 1ª relação de (12) com (2) produz

$$P_\theta(A) = \int_A g(\theta, y) P(dy) \quad [\nu] \quad (13)$$

O conjunto ν -nulo onde (13) pode não se verificar é, em geral, dependente do conjunto A . Mas, como (Y, \mathcal{A}) foi suposto euclideano, pode mostrar-se (vide e.g. Picci [1974]) que a expressão (13) é válida para todo o $A \in \mathcal{A}$ desde que $\theta \in \mathcal{E} - N_1$, onde N_1 é um conjunto ν -nulo independente de A mas possivelmente dependente da versão particular escolhida de $d\mu/d(\nu \times P)$. Resulta então de (13) que P domina a família $\{P_\theta: \theta \in \mathcal{E} - N_1\}$ e, portanto, é ν -equivalente (no sentido de continuidade absoluta mútua) a \mathcal{P} , relação que denotaremos por $\langle \approx \rangle$.

No caso geral, a dominação de \mathcal{P} por P , exigindo que $\forall A \in \mathcal{A}$, $P_\theta(A) = 0$ $[\nu] \implies P_\theta(A) = 0$, $\forall \theta \in \mathcal{E}$, está dependente da medida à priori usada (Florens et al. [1983, sec. 2.3]). Contudo, no caso dominado em questão, podemos definir a função de transição $P_\theta(A)$ de modo que a equivalência entre P e \mathcal{P} se mantenha sem conjuntos excepcionais. Basta atender à 1ª igualdade de (12) e definir

$$P_\theta(A) = \int_A g(\theta, y) P(dy), \quad \forall \theta \in \mathcal{E} \quad (14)$$

onde $g(\theta, y)$, que é \mathcal{B} -mensurável para cada y , pode ser a versão mencionada acima.

Do mesmo modo, pela 2ª igualdade de (12) podemos definir

$$v_y(B) = \int_B g(\theta, y) v(d\theta), \quad \forall y \in Y \quad (15)$$

obtendo assim uma função de transição em $Y \times \mathcal{B}$ que, por (12) e (7), está associada a uma versão regular da restrição a $\mathcal{F}_{\bar{\theta}}$ da probabilidade μ condicional a $\mathcal{F}_{\bar{y}}$ e tal que $\{v_y : y \in Y\} \ll v$.

Em suma, a dominação citada de P quando definida num espaço euclidiano implica a dominação do modelo Bayesiano - entendida no sentido de $\mu \ll v \times P$ - e consequentemente a sua regularidade, independentemente da suposição de $(\mathcal{E}, \mathcal{B})$ ser euclidiano. Quando $(\mathcal{E}, \mathcal{B})$ é igualmente euclidiano, (13) e a expressão correspondente para $v_y(B)$ obtida da comparação da 2ª relação de (12) com (7), revelam que as derivadas de Radon-Nikodym, dP_{θ}/dP e dv_y/dv , coincidem ($v \times P$) - quase em toda a parte.

Suponhamos adicionalmente que $v \ll \lambda$, onde λ é uma medida σ -finita em $(\mathcal{E}, \mathcal{B})$ e seja $q(\theta)$ uma versão da densidade à priori relativamente a λ . Então, para P -essencialmente todo o y

$$\frac{dv_y(\theta)}{d\lambda(\theta)} = \frac{dv_y(\theta)}{dv(\theta)} \frac{dv(\theta)}{d\lambda(\theta)} = q(\theta|y) \quad [\lambda]$$

onde

$$q(\theta|y) = h(\theta, y) q(\theta) = \frac{f(\theta, y) g(\theta)}{\int_{\mathcal{E}} f(\theta, y) g(\theta) d\lambda(\theta)} \quad (16)$$

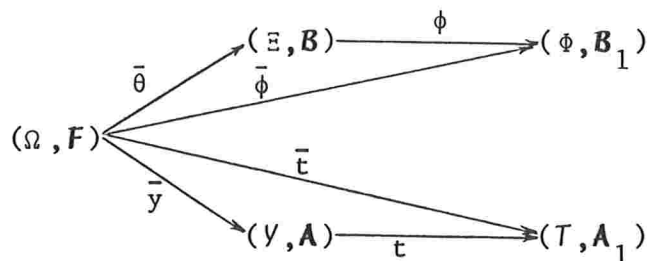
traduz a conhecida expressão para a função densidade à posteriori resultante da aplicação do teorema de Bayes.

Observação: Se usarmos uma medida v σ -finita, mas não finita, o teorema generalizado da medida produto continua sendo aplicável, sendo $\mu = v \times P_{\theta}$ igualmente σ -finita e positiva. Contudo, pode acontecer que a medida de probabilidade preditiva P não seja σ -finita (se o for, não há alterações nos resultados à posteriori) e, nesse caso, a decomposição $\mu = P \times v_y$ já não é válida. Adicionalmente, embora o 2º membro de (15) continue definindo uma função de transição em $Y \times \mathcal{B}$ - e, por isso, possa ser igualmente denominada de medida a posteriori para y fixo - ela já não é finita, com os problemas que tal implica em termos inferenciais. Tais possibilidades levam-nos a que não consideremos aqui distribuições impróprias em $(\mathcal{E}, \mathcal{B})$. Δ

3) Para evitar complicações admitiremos que ambos os espaços mensuráveis (Y, \mathcal{A}) e $(\mathcal{E}, \mathcal{B})$ são euclidianos de modo que as probabilidades em (Ω, \mathcal{F}) condicionais em sub-álgebras de \mathcal{F} admitem versões regulares (vide e.g. Ash [1972, sec. 6.6]).

Consideremos as estatísticas $\phi: (\mathcal{E}, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_1)$ e $t: (Y, \mathcal{A}) \rightarrow (T, \mathcal{A}_1)$. Elas de finem as transformações compostas respectivamente $\mathcal{F}_{\bar{\theta}}$ -mensuráveis e $\mathcal{F}_{\bar{y}}$ -mensuráveis

(prop. B.1.3), $\bar{\phi} = \phi \circ \bar{\theta}$ e $\bar{t} = t \circ \bar{y}$. Sendo \mathcal{B}_ϕ (resp. \mathcal{A}_t) a sub-álgebra induzida por ϕ (resp. t), a sub-álgebra induzida por $\bar{\phi}$ (resp. \bar{t}) é então $\mathcal{F}_{\bar{\phi}} = \mathcal{B}_\phi \times \mathcal{Y}$ (resp. $\mathcal{F}_{\bar{t}} = \mathcal{E} \times \mathcal{A}_t$).



Por definição das probabilidades condicionais em (Ω, \mathcal{F}) dado $\mathcal{F}_{\bar{\phi}}$ e $\mathcal{F}_{\bar{t}}$ respectivamente, temos para todo o $B \in \mathcal{B}$ e $A \in \mathcal{A}$

$$\mu[(B \cap B_1) \times A] = \int_{\bar{B}_1} \mu[B \times A | \mathcal{F}_{\bar{\phi}}](\omega) \mu_{\bar{\phi}}(d\omega), \quad \forall B_1 \in \mathcal{B}_\phi \quad (17)$$

$$\mu[B \times (A \cap A_1)] = \int_{\bar{A}_1} \mu[B \times A | \mathcal{F}_{\bar{t}}](\omega) \mu_{\bar{t}}(d\omega), \quad \forall A_1 \in \mathcal{A}_t \quad (18)$$

onde $\mu_{\bar{\phi}}$ (resp. $\mu_{\bar{t}}$) denota a restrição de μ a $(\Omega, \mathcal{F}_{\bar{\phi}})$ (resp. $(\Omega, \mathcal{F}_{\bar{t}})$). Tomando uma versão regular dessas probabilidades condicionais obtemos funções de transição em $\Phi \times \mathcal{B}$ e $T \times \mathcal{A}$, respectivamente, que designaremos por $\varepsilon_\phi(\mathcal{B})$ e por $\varepsilon_t(\mathcal{A})$.

Adoptando o procedimento de identificar as sub-álgebras de $\mathcal{F}_{\bar{\phi}}$ (resp. $\mathcal{F}_{\bar{t}}$) pelas correspondentes sub-álgebras de $\mathcal{B}(\mathcal{A})$ as restrições de $\varepsilon_\phi(\mathcal{B})$ a $\mathcal{F}_{\bar{\theta}}$ e a $\mathcal{F}_{\bar{y}}$ serão identificadas respectivamente pelas funções de transição $v_\phi(\mathcal{B}): \Phi \times \mathcal{B} \rightarrow [0,1]$ e $P_\phi(\mathcal{A}): \Phi \times \mathcal{A} \rightarrow [0,1]$. Do mesmo modo, $v_t(\mathcal{B})$ e $P_t(\mathcal{A})$ designarão as funções de transição em $T \times \mathcal{B}$ e em $T \times \mathcal{A}$ que correspondem às restrições de $\varepsilon_t(\mathcal{A})$ a $\mathcal{F}_{\bar{\theta}}$ e a $\mathcal{F}_{\bar{y}}$, respectivamente.

Consideremos agora o condicionamento nas sub-álgebras induzidas por $z = (\bar{\phi}, \bar{y})$ (i.e. $\sigma(z) = \mathcal{F}_{\bar{\phi}} \vee \mathcal{F}_{\bar{y}}$) e por $x = (\bar{\theta}, \bar{t})$ (ou seja, $\sigma(x) = \mathcal{F}_{\bar{\theta}} \vee \mathcal{F}_{\bar{t}}$). Por definição da restrição a $\mathcal{F}_{\bar{\phi}}$ (resp. $\mathcal{F}_{\bar{t}}$) da probabilidade condicional dado $\sigma(z)$ (resp. $\sigma(x)$), temos para todo o $B \in \mathcal{B}$ e $A \in \mathcal{A}$

$$\mu[(B \cap B_1) \times A_1] = \int_{\bar{B}_1 \cap \bar{A}_1} \mu[\bar{B} | \sigma(z)](\omega) \mu_z(d\omega), \quad \forall B_1 \in \mathcal{B}_\phi, \quad \forall A_1 \in \mathcal{A} \quad (19)$$

$$\mu[B_1 \times (A \cap A_1)] = \int_{\bar{B}_1 \cap \bar{A}_1} \mu[\bar{A} | \sigma(x)](\omega) \mu_x(d\omega), \quad \forall B_1 \in \mathcal{B}, \quad \forall A_1 \in \mathcal{A}_t \quad (20)$$

onde μ_z e μ_x denotam as restrições de μ a $(\Omega, \sigma(z))$ e a $(\Omega, \sigma(x))$ respectivamente. As σ -álgebras $\sigma(z)$ e $\sigma(x)$ serão futuramente designadas por $\mathcal{B}_\phi \vee \mathcal{A}$ e $\mathcal{B} \vee \mathcal{A}_t$, respectivamente.

Tomando uma versão regular dessas duas probabilidades condicionais obtemos funções de transição em $(\mathcal{Y} \times \Phi) \times \mathcal{B}$ e em $(\mathcal{E} \times T) \times \mathcal{A}$ que designaremos respectivamente por $v_{y,\phi}(\mathcal{B})$ e por $P_{\theta,t}(\mathcal{A})$. Atendendo a (6), (5) e à definição das probabi-

lidades condicionais, $v_y(\cdot | \mathcal{B}_\phi)$ em $(\mathcal{E}, \mathcal{B})$ e $P_\theta(\cdot | \mathcal{A}_t)$ em $(\mathcal{Y}, \mathcal{A})$, os primeiros membros de (19) e (20) podem pôr-se na forma

$$\begin{aligned} \mu[(B \cap B_1) \times A_1] &= \int_{\bar{A}_1} I_{\frac{\cdot}{B \cap B_1}}(\omega) \mu(d\omega) = \int_{A_1} v_y(B \cap B_1) P(dy) = \\ &= \int_{A_1} \left[\int_{B_1} v_y(B | \mathcal{B}_\phi) \tilde{v}_y(d\theta) \right] P(dy) \end{aligned} \quad (21)$$

$$\begin{aligned} \mu[B_1 \times (A \cap A_1)] &= \int_{\bar{B}_1} I_{\frac{\cdot}{A \cap A_1}}(\omega) \mu(d\omega) = \int_{B_1} P_\theta(A \cap A_1) v(d\theta) = \\ &= \int_{B_1} \left[\int_{A_1} P_\theta(A | \mathcal{A}_t) \tilde{P}_\theta(dy) \right] v(d\theta) \end{aligned} \quad (22)$$

onde \tilde{v}_y (resp. \tilde{P}_θ) é a restrição de v_y a $(\mathcal{E}, \mathcal{B}_\phi)$ (resp. de P_θ a $(\mathcal{Y}, \mathcal{A}_t)$).

Quando \mathcal{B}_ϕ e \mathcal{A}_t (ou equivalentemente \mathcal{B}_1 e \mathcal{A}_1) são separáveis - como acontece se ϕ e t são estatísticas reais - existem versões $(\mathcal{B}_\phi \vee \mathcal{A})$ -mensuráveis de $v_y(B | \mathcal{B}_\phi)$, $\forall y \in \mathcal{Y}$, $\forall B \in \mathcal{B}$ e $(\mathcal{B} \vee \mathcal{A}_t)$ -mensuráveis de $P_\theta(A | \mathcal{A}_t)$, $\forall \theta \in \mathcal{E}$, $\forall A \in \mathcal{A}$ (Florens et al. [1983, sec.2.3]). Com versões desse tipo, constata-se que $v_{y,\phi}(\cdot)$ (resp. $P_{\theta,t}(\cdot)$) corresponde essencialmente à distribuição a posteriori de θ dado ϕ (resp. à distribuição amostral de y dado t).

Para o efeito, vamos considerar o caso de $v_{y,\phi}$ (o caso de $P_{\theta,t}$ é perfeitamente análogo pela dualidade existente num modelo Bayesiano regular). Pela construção da medida produto μ , a sua restrição a S_0 , até aqui denotada pelo mesmo símbolo μ , pode ser vista como induzida pelo vector aleatório $u = (\bar{\theta}, \bar{y})$. Assim, $\forall C \in \mathcal{B}$, $D \in \mathcal{A}$

$$\begin{aligned} (\mu u^{-1})(C \times D) &= \mu[u^{-1}(C \times D)] = \mu(\bar{C} \cap \bar{D}) = \\ &= \int_{\bar{D}} \mu(\bar{C} | \mathcal{F}_{\bar{y}})(\omega) \mu_{\bar{y}}(d\omega) = \int_D v_y(C) P(dy) \end{aligned}$$

e portanto

$$(\mu u^{-1})(d\theta \times dy) = v_y(d\theta) P(dy)$$

O 2º membro de (19) transforma-se então pelo teorema de Fubini em

$$\mu[(B \cap B_1) \times A_1] = \int_{A_1} \left[\int_{B_1} v_{y,\phi}(B) \tilde{v}_y(d\theta) \right] P(dy) \quad (23)$$

Por comparação com (21) conclui-se, dada a arbitrariedade de $A_1 \in \mathcal{A}$, que

$$v_y(B \cap B_1) = \int_{B_1} v_{y,\phi}(B) \tilde{v}_y(d\theta) [P] \quad (24)$$

Como por definição da probabilidade a posteriori condicional

$$\nu_y(B \cap B_1) = \int_{B_1} \nu_y(B|B_\phi) \tilde{\nu}_y(d\theta), \quad \forall B_1 \in \mathcal{B}_\phi$$

resulta, dada a bimensurabilidade assumida de $\nu_y(B|B_\phi)$, que $\forall B \in \mathcal{B}$, $\nu_{y,\phi}(B) = \nu_y(B|B_\phi) [\tilde{\nu}_y]$ para P -essencialmente todo o $y \in Y$. Ou seja, a probabilidade à posteriori condicional em ϕ é μ -essencialmente igual à probabilidade condicional em $\mathcal{B}_\phi \vee \mathcal{A}$.

Sob as hipóteses acima admitidas de existência de versões regulares das probabilidades condicionais mencionadas, o teorema generalizado da medida produto garante que as medidas produto condicionais ε_ϕ e ε_t admitem as decomposições

$$\begin{aligned} \varepsilon_\phi &= \nu_\phi \times P_\theta = P_\phi \times \nu_{y,\phi} \quad [\nu_\phi] \\ \varepsilon_t &= \nu_t \times P_{\theta,t} = P_t \times \nu_y \quad [P_t] \end{aligned}$$

traduzindo pois a regularidade dos modelos Bayesianos condicionados $(\Omega, \mathcal{F}, \varepsilon_\phi)$ e $(\Omega, \mathcal{F}, \varepsilon_t)$. Quando (Y, \mathcal{A}) é euclideano, $p \ll m$ e $f(\theta, y)$ é uma versão \mathcal{F} -mensurável de dP_θ/dm , estes dois modelos são dominados - respectivamente por $\nu_\phi \times P$ e por $\nu \times P_t$ - e, consequentemente, regulares, desde que existam versões regulares de ν_ϕ e de P_t , respectivamente (Florens et al. [1983, sec. 1.3.5]). Nesta situação, a família de medidas $\nu_{y,\phi}$ (resp. $P_{\theta,t}$) é dominada por ν_ϕ (resp. P_t).

O procedimento acabado de descrever pode ser aplicado a sub-álgebras $\mathcal{C} \subset \mathcal{B}$ e $\mathcal{D} \subset \mathcal{A}$ sem explicitar se elas são ou não induzidas por estatísticas não triviais. Neste contexto, e frisando a identificação de \mathcal{C} (resp. \mathcal{D}) com $\mathcal{C} \times Y$ (resp. $E \times \mathcal{D}$) - o que leva a que $\mathcal{F}_{\bar{y}} \vee (\mathcal{C} \times Y)$, por exemplo, seja identificada por $\mathcal{A} \vee \mathcal{C}$ - os símbolos $\nu_{\mathcal{C}}$, $\nu_{\mathcal{A} \vee \mathcal{C}}$ e $\nu_{\mathcal{D}}$ denotarão as medidas no espaço paramétrico que se identificam com a restrição a \mathcal{B} das probabilidades condicionais em $\mathcal{C} \vee \mathcal{A} \vee \mathcal{C}$ e \mathcal{D} , respectivamente. A separabilidade de \mathcal{C} garante, como vimos, a interpretação de $\nu_{\mathcal{A} \vee \mathcal{C}}$ como uma medida à posteriori condicional em \mathcal{C} . Analogamente, as medidas no espaço amostral identificadas com a restrição a \mathcal{A} das probabilidades condicionais em \mathcal{D} , $\mathcal{B} \vee \mathcal{D}$ ($\cong \mathcal{F}_{\bar{y}} \vee (E \times \mathcal{D})$) e \mathcal{C} serão designadas por $P_{\mathcal{D}}$, $P_{\mathcal{B} \vee \mathcal{D}}$ e $P_{\mathcal{C}}$, respectivamente. A função $P_{\mathcal{B} \vee \mathcal{D}}^{\bar{\theta}}$ representa uma probabilidade amostral condicionada em \mathcal{D} se \mathcal{D} for separável.

4) Na secção 2.2 aparecem relações que são válidas quase em toda a parte com respeito a uma determinada medida, afigurando-se conveniente que tal validade seja estendida a toda a parte. Importa, pois, impor condições nessa medida de modo a que essa extensão seja válida. Quando a medida envolvida é a probabilidade marginal P , as condições que precisamos estão traduzidas no seguinte conceito de regularidade que é o dual do conceito de Florens et al. [1983] de regularidade de ν para P .

Diz-se que P é regular para a família \mathcal{Q} se para toda a função ϕ real limitada \mathcal{B} -mensurável, i.e., $\phi \in L_\infty(\mathcal{B})$, $E_{\nu_y}(\phi) = 0 [P] \implies E_{\nu_y}(\phi) = 0, \forall y \in Y$.

Esta condição é verificada em inúmeras situações usuais do caso euclidiano:

- Se Y é numerável, então a regularidade de P para \mathcal{Q} é claramente garantida se o conjunto vazio for o único conjunto P -nulo, isto é, se $P(\{y\}) > 0, \forall y \in Y$;
- Se Y é não numerável e \mathcal{Q} é tal que para todo $\phi \in L_\infty(\mathcal{B})$ $E_{\nu_y}(\phi) \equiv f_\phi^*(y)$ é contínua em Y , a regularidade de P é assegurada se P atribuir probabilidade positiva a qualquer aberto de A .

De facto, fixemos $\phi \in L_\infty(\mathcal{B})$ e suponhamos que y_0 é um ponto de Y tal que $f_\phi^*(y_0) \neq 0$. Pela continuidade de $f_\phi^*(y)$ existe um $\varepsilon > 0$ tal que $f_\phi^*(y) \neq 0, \forall y \in V_\varepsilon(y_0)$, onde $V_\varepsilon(y_0)$ designa a bola aberta de centro y_0 e raio ε . A hipótese sobre P implica que $P[V_\varepsilon(y_0)] > 0$, o que significa que é falsa a afirmação de $f_\phi^*(y) = 0 [P]$.

- No contexto anterior, sob a dominação de \mathcal{P} pela medida de Lebesgue em Y , a condição de $E_{\nu_y}(\phi) = \int_{\mathcal{E}} \phi(\theta) f(\theta, y) \nu(d\theta)$ ser contínua em $Y, \forall \phi \in L_\infty(\mathcal{B})$ é verificada, por exemplo, quando a versão F -mensurável $f(\theta, y)$ é contínua em Y para cada θ e dominada para todo o y por uma função de θ ν -integrável. Neste quadro, se $\int_{\mathcal{E}} f(\theta, y) \nu(d\theta)$ for positivo, $\forall y \in Y$, então P é regular para \mathcal{Q} .

Considerando $\phi = I_B, B \in \mathcal{B}, \nu(B) = 0$, a regularidade de P para \mathcal{Q} implica que \mathcal{Q} seja dominada por (e portanto, equivalente a) ν . Por outro lado, no caso de dominação de \mathcal{P} citado em 2), o facto de $\nu(B) = \int_Y \nu_y(B) P(dy)$ e de $\nu_y(B) = \int_B g(\theta, y) \nu(d\theta)$, expressões que traduzem que $\mathcal{Q} \ll \nu$, implica que a definição de regularidade de P para \mathcal{Q} seja verificada para as funções indicadoras \mathcal{B} -mensuráveis.

APÊNDICE B.3 - ALGUNS RESULTADOS SOBRE INDEPENDÊNCIA CONDICIONAL

Seja $(\Omega, \mathcal{F}, \mu)$ um dado espaço de probabilidade e $L_\infty(\mathcal{F})$ o conjunto das variáveis aleatórias (v.a.) - funções reais definidas em Ω \mathcal{F} -mensuráveis - essencialmente limitadas, i.e., $f \in L_\infty(\mathcal{F})$ se $\exists a \in \mathbb{R}$: $\mu(\{\omega \in \Omega : |f(\omega)| \leq a\}) = 1$. O conjunto das v.a. essencialmente limitadas \mathcal{F}_1 -mensuráveis, com \mathcal{F}_1 uma sub-álgebra de \mathcal{F} , será denotado em consonância por $L_\infty(\mathcal{F}_1)$.

- Sendo $f \in L_\infty(\mathcal{F})$, a esperança condicional de f dado \mathcal{F}_1 , representada por $E(f|\mathcal{F}_1)$, é qualquer elemento da classe de μ -equivalência de v.a. de $L_\infty(\overline{\mathcal{F}}_1)$ tal que

$$\forall g \in L_\infty(\overline{\mathcal{F}}_1), \int_{\Omega} f g \, d\mu = \int_{\Omega} E(f|\mathcal{F}_1) g \, d\mu_1$$

onde μ_1 é a restrição de μ a \mathcal{F}_1 . É contudo suficiente aplicar esta definição a $g = I_{F_1}$, $F_1 \in \mathcal{F}_1$ e, nesse sentido, pode provar-se que tal definição é equivalente à definição usual, a qual se reporta a v.a. μ -integráveis (denotadas por $L_1(\mathcal{F})$). Quando $f = I_F$, $F \in \mathcal{F}$, obtém-se a definição de probabilidade condicional $\mu(F|\mathcal{F}_1) = E(I_F|\mathcal{F}_1)$.

- Algumas das propriedades mais conhecidas da esperança condicional estabelecem que:

"Smoothing property": $\forall f \in L_\infty(\mathcal{F}), \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$

$$E\{E(f|\mathcal{F}_2)|\mathcal{F}_1\} = E\{E(f|\mathcal{F}_1)|\mathcal{F}_2\} = E(f|\mathcal{F}_1) \quad [\mu]$$

"Self-adjointness property": $\forall f, g \in L_1(\mathcal{F}), \mathcal{F}_1 \subset \mathcal{F}$

$$E\{fE(g|\mathcal{F}_1)\} = E\{gE(f|\mathcal{F}_1)\}$$

- Seja M uma classe de subconjuntos de Ω . M é um sistema de Dynkin (abreviadamente, sistema-D) se

i) $\Omega \in M$

ii) $\forall F_1, F_2 \in M, F_1 \subset F_2, F_2 - F_1 \in M$

iii) Se $\{F_n, n \geq 1\}$ é uma sequência crescente de conjuntos de M , $\bigcup_{n=1}^{\infty} F_n \in M$

A seguinte proposição, conhecida como teorema de Dynkin, afigura-se bastante útil na demonstração da equivalência de várias relações (vide Ash [1972, p. 168-169] para a sua demonstração):

Prop. B.3.1: Seja S uma classe- π de subconjuntos de Ω (i.e., uma classe fechada sob intersecções finitas). Se M é um sistema-D incluindo S , então M inclui a σ -álgebra gerada por S .

- Sejam \mathcal{F}_i , $i=1,2,3$ sub-álgebras de \mathcal{F} . Diz-se que \mathcal{F}_1 e \mathcal{F}_2 são independentes condicionalmente a \mathcal{F}_3 - e escreve-se $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$ - se

$$\forall f_i \in L_\infty(\mathcal{F}_i), i=1,2, \quad E(f_1 f_2 | \mathcal{F}_3) = E(f_1 | \mathcal{F}_3) E(f_2 | \mathcal{F}_3) \quad [\mu]$$

Pode provar-se que é suficiente a consideração de f_i $i=1,2$ como funções indicadoras de conjuntos de \mathcal{F}_i para a definição do conceito de independência condicional.

Prop. B.3.2: $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$ sse $\forall f_2 \in L_\infty(\mathcal{F}_2), E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) = E(f_2 | \mathcal{F}_3) \quad [\mu]$

Prova: (Parte suficiente) - Sejam $f_i \in L_\infty(\mathcal{F}_i), i=1,2$.

Pela propriedades de "smoothing" e de idempotência (i.e. $E(f | \mathcal{F}_1) = f$ $[\mu]$ se $f \in L_\infty(\mathcal{F}_1)$)

$$E(f_1 \cdot f_2 | \mathcal{F}_3) = E\{E(f_1 f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) | \mathcal{F}_3\} = E\{f_1 E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) | \mathcal{F}_3\} \quad [\mu]$$

Assim

$$E(f_1 f_2 | \mathcal{F}_3) = E\{f_1 E(f_2 | \mathcal{F}_3) | \mathcal{F}_3\} = E(f_1 | \mathcal{F}_3) E(f_2 | \mathcal{F}_3) \quad [\mu]$$

(Parte necessária) - Sejam $F_1 \in \mathcal{F}_1$ e $F_3 \in \mathcal{F}_3$. Então $F_1 \cap F_3 \in \mathcal{F}_1 \vee \mathcal{F}_3$ e, por definição da esperança condicional dado $\mathcal{F}_1 \vee \mathcal{F}_3$ e dado \mathcal{F}_3 resulta, omitindo a explicitação das restrições consideradas de μ nos integrais citados,

$$\int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) d\mu = \int_{F_1 \cap F_3} f_2 d\mu = \int_{F_3} f_2 I_{F_1} d\mu = \int_{F_3} E(f_2 I_{F_1} | \mathcal{F}_3) d\mu$$

Assim, por hipótese e, por definição de $E(I_{F_1} | \mathcal{F}_3)$, resulta

$$\int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) d\mu = \int_{F_3} E(f_2 | \mathcal{F}_3) E(I_{F_1} | \mathcal{F}_3) d\mu = \int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_3) d\mu$$

Seja $M = \{F \in \mathcal{F}_1 \vee \mathcal{F}_3 : \int_F f_2 d\mu = \int_F E(f_2 | \mathcal{F}_3) d\mu\}$ e $S = \{F_1 \cap F_3 : F_1 \in \mathcal{F}_1, F_3 \in \mathcal{F}_3\}$.

Então $S \subset M \subset \mathcal{F}_1 \vee \mathcal{F}_3$ e S é claramente uma classe- π . Pelas propriedades de integral prova-se que M é um sistema-D e portanto, a proposição B.3.1 assegura-nos que $M = \mathcal{F}_1 \vee \mathcal{F}_3$, já que $\mathcal{F}_1 \vee \mathcal{F}_3 = \sigma(S)$. Ou seja, $E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) = E(f_2 | \mathcal{F}_3) \quad [\mu]$ c.q.d. Δ

Dado que toda a versão de $E(f_2 | \mathcal{F}_3)$ é $\bar{\mathcal{F}}_3$ -mensurável e atendendo ao significado da mensurabilidade μ -essencial para v.a. (vide B.1), obtém-se como corolário

Corolário B.3.3: $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$ sse $\forall f_2 \in L_\infty(\mathcal{F}_2), E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) \in \bar{\mathcal{F}}_3$.

No caso particular em que $\mathcal{F}_3 \subset \mathcal{F}_1$, $\mathcal{F}_1 \vee \mathcal{F}_3 = \mathcal{F}_1$ e assim

Corol. B.3.4: Se $F_3 \subset F_1$, então $F_1 \perp\!\!\!\perp F_2 | F_3$ sse $E(f_2 | F_1) = E(f_2 | F_3) [\mu]$, ou equivalentemente, $E(f_2 | F_1) \in \bar{F}_3$, para toda a função $f_2 \in L_\infty(F_2)$.

A proposição seguinte ilustra a propriedade de monotonicidade da independência condicional relativamente a sub-álgebras de F_1 .

Prop. B.3.5: Se $F_1 \perp\!\!\!\perp F_2 | F_3$ e $F_3 \subset F_4 \subset F_1$, então $F_1 \perp\!\!\!\perp F_2 | F_4$.

Prova: Pelo corolário B.3.4, $F_1 \perp\!\!\!\perp F_2 | F_3 \implies E(f_2 | F_1) = E(f_2 | F_3) [\mu]$, $\forall f_2 \in L_\infty(F_2)$. Mas a definição de $F_1 \perp\!\!\!\perp F_2 | F_3$ implica, pelo facto de $F_4 \subset F_1$, que $F_4 \perp\!\!\!\perp F_2 | F_3$, ou seja, pelo corolário B.3.4

$$\forall f_2 \in L_\infty(F_2), E(f_2 | F_4) = E(f_2 | F_3) [\mu]$$

Consequentemente,

$$\forall f_2 \in L_\infty(F_2), E(f_2 | F_1) = E(f_2 | F_4) [\mu] \iff F_1 \perp\!\!\!\perp F_2 | F_4. \quad \Delta$$

- Designemos por $F_1[F_2]$ a σ -álgebra gerada por qualquer versão de $E(f_2 | F_1)$ para toda a função $f_2 \in L_\infty(F_2)$. Como $F_1[F_2] \subset F_1$, por construção de $F_1[F_2]$ temos, de acordo com o corolário B.3.4, $F_1 \perp\!\!\!\perp F_2 | F_1[F_2]$. Seja agora $F_3 \subset F_1$ tal que $F_1 \perp\!\!\!\perp F_2 | F_3$.

Pelo corolário B.3.4., $\forall f_2 \in L_\infty(F_2)$, $E(f_2 | F_1) \in \bar{F}_3$, o que implica $F_1[F_2] \subset \bar{F}_3$, dado que $F_1[F_2]$ é a menor σ -álgebra contendo as funções $E(f_2 | F_1)$. Fica assim demonstrado que:

Prop. B.3.6: Sendo F_i , $i=1,2,3$ sub-álgebras de F tal que $F_3 \subset F_1$, então

i) $F_1 \perp\!\!\!\perp F_2 | F_1[F_2]$

ii) Se $F_1 \perp\!\!\!\perp F_2 | F_3$ então $F_1[F_2] \subset F_3 [\mu]$

O resultado desta proposição significa que $F_1[F_2]$ é essencialmente a menor das sub-álgebras de F_1 condicionalmente às quais F_1 e F_2 são independentes.

- Sendo F_1 e F_2 duas sub-álgebras de F , diz-se que F_1 é fortemente identificável por F_2 (ou que F_2 identifica fortemente F_1), e escreve-se $F_1 \ll F_2$, se

$$\forall f \in L_\infty(\bar{F}_1), E(f | F_2) = 0 [\mu] \implies f = 0 [\mu]$$

A proposição seguinte, que é uma versão abstracta da proposição 7 de Basu and Pereira (1983), mostra que toda a sub-álgebra de F_1 condicionalmente à qual F_1 e F_2 são independentes e fortemente identificável por F_2 coincide essencialmente com $F_1[F_2]$.

Prop. B.3.7: Sejam F_i , $i=1,2,3$ sub-álgebras de F tais que $F_3 \subset F_1$ e $F_1 \perp\!\!\!\perp F_2 | F_3$. Se $F_3 \ll F_2$ então $F_3 = F_1[F_2] [\mu]$.

Prova. Atendendo à proposição B.3.6 temos por um lado $F_1[F_2] \subset \bar{F}_3$ e, por outro, $F_1 \perp\!\!\!\perp F_2 | F_1[F_2]$, o que implica $F_3 \perp\!\!\!\perp F_2 | F_1[F_2]$, dado que $F_3 \subset F_1$, ou seja, pela simetria do conceito de independência condicional $F_2 \perp\!\!\!\perp F_3 | F_1[F_2]$. Sendo então $f \in L_\infty(F_3)$, pela propriedade "smoothing" da esperança condicional e pela proposição B.3.2 temos

$$\begin{aligned} E(f|F_2) &= E\{E(f|F_2 \vee F_1[F_2])|F_2\} \quad [\mu] \\ &= E\{E(f|F_1[F_2])|F_2\} \quad [\mu] \end{aligned}$$

isto é, com $g = f - E(f|F_1[F_2])$, $E(g|F_2) = 0 \quad [\mu]$. Pelo facto de $F_1[F_2] \subset \bar{F}_3$, $g \in \bar{F}_3$ e assim, pela hipótese de $F_3 \ll F_2$, resulta $f = E(f|F_1[F_2]) \quad [\mu]$, o que significa que $f \in \overline{F_1[F_2]}$. Assim, $F_3 = \overline{F_1[F_2]}$ e conseqüentemente, $\bar{F}_3 = \overline{F_1[F_2]}$, o que é equivalente a afirmar que F_3 e $F_1[F_2]$ são μ -essencialmente equivalentes, c.q.d.

△

- Em ordem a evidenciar as implicações estatísticas do conceito de identificabilidade forte e da proposição B.3.7 consideremos que $(\Omega, \mathcal{F}, \mu)$ representa o modelo Bayesiano regular construído como foi indicado em B.2 e seja $F_1 = F_{\bar{y}}$ e $F_2 = F_{\bar{\theta}}$.

Sendo $\mathcal{D} \subset \mathcal{A}$, então por definição de identificabilidade forte

$$E \times \mathcal{D} \ll F_{\bar{\theta}} \iff \forall \bar{t} \in L_\infty(E \times \mathcal{D}), E(\bar{t}|F_{\bar{\theta}}) = 0 \quad [\mu] \implies \bar{t} = 0 \quad [\mu]$$

Mas, pelo que vimos em B.2, $P \equiv \mu_{\bar{y}}$, $\bar{t} \in L_\infty(E \times \mathcal{D})$ sse $\bar{t} = t\bar{y}$ com $t \in L_\infty(\mathcal{D})$ e $E(\bar{t}|F_{\bar{\theta}}) = E_{P_{\bar{\theta}}}(t) \quad [v]$. Assim, num modelo Bayesiano regular

$$E \times \mathcal{D} \ll F_{\bar{\theta}} \text{ sse } \forall t \in L_\infty(\mathcal{D}), E_{P_{\bar{\theta}}}(t) = 0 \quad [v] \implies t = 0 \quad [P]$$

o que revela que o conceito de identificabilidade forte de \mathcal{D} por \mathcal{B} traduz a versão Bayesiana do conceito de σ -álgebra completa limitada. Recorde-se, para o efeito, que no contexto clássico, \mathcal{D} é uma σ -álgebra completa limitada se $\forall t \in L_\infty(\mathcal{D}), E_{P_{\bar{\theta}}}(t) = 0, \forall \theta \in E \implies t = 0 \quad [P]$.

Num contexto Bayesiano diremos, por dualidade com tal definição clássica, que $\mathcal{C} \subset \mathcal{B}$ é uma σ -álgebra completa limitada se $\forall \phi \in L_\infty(\mathcal{C}), E_{v_y}(\phi) = 0, \forall y \in Y$ implica $\phi = 0 \quad [Q]$.

Por repetição do argumento usado acima para definir $E \times \mathcal{D} \ll F_{\bar{\theta}}$, o conceito de identificabilidade forte de $\mathcal{C} \times Y$ por $F_{\bar{y}}$ significa que

$$\begin{aligned} \mathcal{C} \times Y \ll F_{\bar{y}} &\iff \forall \bar{\phi} \in L_\infty(\mathcal{C} \times Y), E(\bar{\phi}|F_{\bar{y}}) = 0 \quad [\mu] \implies \bar{\phi} = 0 \quad [\mu] \\ &\iff \forall \phi \in L_\infty(\mathcal{C}), E_{v_y}(\phi) = 0 \quad [P] \implies \phi = 0 \quad [v] \end{aligned}$$

Recordando o conceito de regularidade de P para \mathcal{Q} (vide 4) em B.2),
 $E_{\nu}(\phi) = 0 [P] \iff E_{\nu}(\phi) = 0, \forall y \in \mathcal{Y}$ e, por outro lado, tal conceito implica
 que $\mathcal{Q} \ll \nu$. Desta forma, fica estabelecido que

Prop. B.3.8: Num modelo Bayesiano regular, sob a regularidade de P para \mathcal{Q} ,
 $\mathcal{C} \subset \mathcal{B}$ é fortemente identificável por \mathbf{A} sse \mathcal{C} é uma σ -álgebra completa limitada.

O dual desta proposição para sub-álgebras de \mathbf{A} (Florens et al [1983, teor. 5.3.23]), significa assim que o conceito clássico de σ -álgebra \mathcal{D} (resp. estatística t) completa limitada coincide com a sua versão Bayesiana - definida por $E \times \mathcal{D} \ll F_{\theta}$ (resp. $E \times \mathbf{A}_t \ll F_{\theta}$) - quando ν é regular para P .

oOo

APÊNDICE B.4 - O DUAL DO TEOREMA DE FACTORIZAÇÃO "CLÁSSICO": DEMONSTRAÇÕES

Por definição, a σ -álgebra $\mathbf{C} \subset \mathbf{B}$ é suf $(\mathbf{B}, \mathcal{Q})$ se para todo o $B \in \mathbf{B}$ existe uma função real $\psi_B \in \mathbf{C}$ tal que

$$\nu_y(B \cap C) = \int_C \psi_B d\nu_y, \quad \forall C \in \mathbf{C}, \quad \forall y \in \mathcal{Y} \quad (1)$$

i.e., $\psi_B = E_y(I_B | \mathbf{C}) [\nu_y]$, $\forall y \in \mathcal{Y}$, onde E_y designa a esperança matemática relativamente à medida ν_y .

Seja $\mathcal{Q} \ll \nu_{y_0}$ e ϕ_y uma versão fixa de $d\nu_y/d\nu_{y_0}$ para cada $y \in \mathcal{Y}$.

Prova do teorema 2.13

Parte necessária: Por definição de ϕ_y e pelo facto de $\psi_B = E_{y_0}(I_B | \mathbf{C}) [\nu_{y_0}]$

$$\nu_y(B \cap C) = \int_C I_B d\nu_y = \int_C I_B \phi_y d\nu_{y_0} \quad (2)$$

$$\int_C \psi_B d\nu_y = \int_C \psi_B \phi_y d\nu_{y_0} = \int_{\Xi} E_{y_0}(I_B | \mathbf{C}) \phi_y I_C d\nu_{y_0} \quad (3)$$

Pela propriedade de auto-adjunção da esperança condicional (vide B.3)

$$\int_C \psi_B d\nu_y = \int_{\Xi} I_B E_{y_0}[\phi_y I_C | \mathbf{C}] d\nu_{y_0} = \int_C I_B E_{y_0}(\phi_y | \mathbf{C}) d\nu_{y_0}$$

Assim, da hipótese de \mathbf{C} ser suf $(\mathbf{B}, \mathcal{Q})$, segue-se que para cada $B \in \mathbf{B}$

$$\int_C I_B \phi_y d\nu_{y_0} = \int_C I_B E_{y_0}(\phi_y | \mathbf{C}) d\nu_{y_0} \quad (4)$$

para todo o $C \in \mathbf{C}$ e $y \in \mathcal{Y}$. Fazendo $C = \Xi$, resulta

$$\nu_y(B) \equiv \int_B \phi_y d\nu_{y_0} = \int_B E_{y_0}(\phi_y | \mathbf{C}) d\nu_{y_0}, \quad \forall B \in \mathbf{B}$$

o que implica que $\phi_y = E_{y_0}(\phi_y | \mathbf{C}) [\nu_{y_0}]$, $\forall y \in \mathcal{Y}$, ou seja, por $\mathcal{Q} \ll \nu_{y_0}$

$$\phi_y = E_{y_0}(\phi_y | \mathbf{C}) [\mathcal{Q}], \quad \forall y \in \mathcal{Y}$$

Isto revela que cada ϕ_y é \mathcal{Q} -essencialmente \mathbf{C} -mensurável como pretendíamos.

Parte suficiente: Note-se em primeiro lugar que toda a função real \mathcal{Q} -essencialmente \mathbf{C} -mensurável é \mathcal{Q} -essencialmente igual a qualquer versão da sua esperança condicional em \mathbf{C} relativamente a qualquer membro de \mathcal{Q} . Assim, em particular, $\forall y \in \mathcal{Y}$.

$$\phi_y \in \mathbf{C}[\mathcal{Q}] \implies \phi_y = E_{y_0}(\phi_y | \mathbf{C}) [\mathcal{Q}]$$

o que implica para todo o $y \in Y$

$$\int_C I_B \phi_y dv_{y_0} = \int_C I_B E_{y_0}(\phi_y | \mathcal{C}) dv_{y_0}, \quad \forall C \in \mathcal{C}$$

ou seja (4). Por (2) e pela propriedade de auto-adjunção do operador $E_{y_0}(\cdot | \mathcal{C})$

$$v_y(B \cap C) = \int_C E_{y_0}(I_B | \mathcal{C}) \phi_y dv_{y_0} = \int_C E_{y_0}(I_B | \mathcal{C}) dv_y$$

Seja ψ_B uma versão de $E_{y_0}(I_B | \mathcal{C})$. Então ψ_B é uma função real \mathcal{C} -mensurável e, por definição de $E_y(I_B | \mathcal{C})$, $\psi_B = E_y(I_B | \mathcal{C})[v_y] \quad \forall y \in Y$, o que comprova a suficiência de \mathcal{C} com respeito a $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$.

O enunciado fica completamente demonstrado atendendo a que $\phi_y \in \mathcal{C}[\mathcal{Q}]$, $\forall y \in Y$ sse $\sigma\{\phi_y : y \in Y\} \subset \mathcal{C}[\mathcal{Q}] \quad \Delta$

Prova do lema 2.1

Como $\mathcal{Q} \subset \tilde{\mathcal{Q}}$, segue-se que $\mathcal{C} \text{ suf } (\mathcal{B}, \tilde{\mathcal{Q}}) \implies \mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$. Suponhamos agora que ψ_B é para cada $B \in \mathcal{B}$ a versão comum de $E_{v_y}(I_B | \mathcal{C})$, para todo o $v_y \in \mathcal{Q}$ e seja

$$v^* = \sum_{i=1}^{\infty} a_i v_{y_i} \in \tilde{\mathcal{Q}}. \quad \text{Como}$$

$$v^*(B \cap C) = \int_C I_B dv^* = \sum_{i=1}^{\infty} a_i \int_C I_B dv_{y_i} = \sum_{i=1}^{\infty} a_i \int_C \psi_B dv_{y_i} = \int_C \psi_B dv^*$$

para todo o $C \in \mathcal{C}$ e $v^* \in \tilde{\mathcal{Q}}$, fica provada a parte necessária.

Prova do teorema 2.14

Sendo v^* o elemento de $\tilde{\mathcal{Q}}$ equivalente a \mathcal{Q} , consideremos $\mathcal{Q}_1 = \mathcal{Q}U\{v^*\}$. Pelo lema 2.1, $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$ sse $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q}_1)$. Como \mathcal{Q}_1 é auto-dominada por v^* , o teorema 2.13 permite-nos concluir que $\mathcal{C} \text{ suf } (\mathcal{B}, \mathcal{Q})$ sse para cada $y \in Y$, dv_y/dv^* é \mathcal{Q} -essencialmente \mathcal{C} -mensurável, i.e., sse $\sigma\{dv_y/dv^* : y \in Y\}$ está incluída \mathcal{Q} -essencialmente em \mathcal{C} , c.q.d. Δ

Prova do teorema 2.15 (dual do teorema de factorização clássico)

Parte necessária. Sendo $v^* \in \tilde{\mathcal{Q}}$ a combinação convexa dominadora de \mathcal{Q} , a suficiência de \mathcal{C} com respeito a $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$ implica pelo teorema 2.14 que dv_y/dv^* é \mathcal{Q} -essencialmente \mathcal{C} -mensurável para todo o $y \in Y$. Escolhendo uma versão \mathcal{C} -mensurável de dv_y/dv^* , denotada por g_y , e sendo h uma versão de $dv^*/d\lambda$ (note-se que, como $v^* \ll \mathcal{Q}$, $v^* \ll \lambda$), obtemos para a função densidade à posteriori

$$q_y = \frac{dv_y}{d\lambda} = \frac{dv_y}{dv^*} \cdot \frac{dv^*}{d\lambda} = g_y \cdot h \quad [\lambda]$$

Parte suficiente: Sendo $\nu^* = \sum_{i=1}^{\infty} a_i \nu_{y_i}$, pela factorização assumida

$$\frac{d\nu^*}{d\lambda} = \sum_{i=1}^{\infty} a_i \frac{d\nu_{y_i}}{d\lambda} = \left(\sum_{i=1}^{\infty} a_i g_{y_i} \right) h$$

Definamos para todo o $y \in \mathcal{Y}$ a função $g'_y = \frac{g_y}{\sum_{i=1}^{\infty} a_i g_{y_i}} I_E$ onde $E = \{ \theta : \sum_{i=1}^{\infty} a_i g_{y_i}(\theta) > 0 \}$.

Essa função é claramente \mathcal{C} -mensurável e, por definição, $E^c = E - E$ é ν^* -nulo, por conseguinte, \mathcal{Q} -nulo. Assim, para todo o $B \in \mathcal{B}$

$$\begin{aligned} \nu_y(B) &= \nu_y(B) - \nu_y(B \cap E^c) = \nu_y(B \cap E) = \int_{B \cap E} q_y d\lambda \\ &= \int_{B \cap E} \frac{g_y}{\sum_{i=1}^{\infty} a_i g_{y_i}} \left(\sum_{i=1}^{\infty} a_i g_{y_i} \right) h d\lambda \\ &= \int_B g'_y d\nu^* \end{aligned}$$

o que revela que a função \mathcal{C} -mensurável g'_y é uma versão de $d\nu_y/d\nu^*$ para todo o $y \in \mathcal{Y}$. O teorema 2.14 garante então a suficiência de \mathcal{C} com respeito a $(\mathcal{E}, \mathcal{B}, \mathcal{Q})$.

△

APÊNDICE C

APÊNDICE C.1. - ALGUMAS NOTAS SOBRE AS FAMÍLIAS DE DISTRIBUIÇÕES
DIRICHLET E DIRICHLET-MULTINOMIAL

A família de distribuições Dirichlet é uma família definida no simplex probabilístico de k ($k \geq 2$) variáveis $\underline{\theta} = (\theta_1 \dots \theta_k)'$, com cada $\theta_i \geq 0$ e $\sum_{i=1}^k \theta_i = 1$, denotado por S_{k-1} , e parametrizada por um vector $\underline{b} = (b_1 \dots b_k)' \in \mathbb{R}_+^k$.

Trata-se assim de uma distribuição $(k-1)$ -variada (o caso $k=2$ define a conhecida família de distribuições beta), que designaremos por $\underline{\theta} | \underline{b} \rightarrow D_k(\underline{b})$. A função densidade de probabilidade expressa em termos de qualquer subconjunto de $k-1$ variáveis θ_i tem a forma

$$g(\underline{\theta} | \underline{b}) = \frac{1}{B(\underline{b})} \prod_{i=1}^k \theta_i^{b_i-1}, \quad \underline{\theta} \in S_{k-1} \quad (1)$$

onde $B(\underline{b}) = \frac{\prod_{i=1}^k \Gamma(b_i)}{\Gamma(b_{\cdot})}$, $b_{\cdot} = \sum_{i=1}^k b_i$, é a função beta completa generalizada,

da, muitas vezes designada como o integral completo de Dirichlet. Um membro especial desta família é a distribuição uniforme contínua em S_{k-1} , correspondente a $\underline{b} = \underline{1}_k$.

Devemos salientar que os resultados considerados mais conhecidos sobre a família Dirichlet são aqui simplesmente registados. A respectiva demonstração pode ser vista em Wilks (1962, sec. 7.7).

A distribuição $\underline{\theta} | \underline{b} \rightarrow D_k(\underline{b})$ pode ser encarada como a distribuição de $k-1$ variáveis θ_i definidas por $\theta_i = \frac{\gamma_i}{\sum_{j=1}^k \gamma_j}$ onde γ_i , $i=1, \dots, k$ têm

distribuições gama independentes $\Gamma(b_i, 1)$ - ou alternativamente, com $\gamma_i | b_i \sim \chi^2(2b_i)$, $i=1, \dots, k$, independentes.

Um vector $\underline{\theta}$, nas condições acima, é limitado e, por isso, a distribuição Dirichlet é completamente determinada pelos seus momentos. O momento misto generalizado Dirichlet é calculado por

$$E_{\underline{\theta}|\underline{b}} \left(\prod_{i=1}^k \theta_i^{c_i} \right) = \frac{B(\underline{b} + \underline{c})}{B(\underline{b})} \quad (2)$$

onde $\underline{c} = (c_1, \dots, c_k)'$ tem coordenadas não negativas. A expressão (2) ainda é aplicável a $\underline{c} \in \mathbb{R}^k$, desde que $\underline{b} + \underline{c} \in \mathbb{R}_+^k$. Por escolha apropriada de \underline{c} podemos obter

$$E(\theta_i) = \frac{b_i}{b.}, \quad i = 1, \dots, k$$

$$\text{Var}(\theta_i) = \frac{b_i(b.-b_i)}{b.^2(b.+1)}, \quad i = 1, \dots, k$$

$$\text{Cov}(\theta_i, \theta_j) = -\frac{b_i b_j}{b.^2(b.+1)}, \quad i \neq j$$

O vector média e a matriz de covariâncias de $\underline{\theta}$ podem então ser expressões por

$$E(\underline{\theta}) = \underline{b}|b. \quad (3)$$

$$\text{Var}(\underline{\theta}) = \frac{1}{b.(b.+1)} (D_{\underline{b}} - \frac{\underline{b}\underline{b}'}{b.}) \quad (4)$$

onde $D_{\underline{b}}$ é a matriz diagonal de ordem k com os elementos de \underline{b} na diagonal principal.

Sendo $(\alpha_1, \dots, \alpha_\ell)$ uma permutação de ℓ índices ($1 \leq \ell \leq k-1$) do conjunto $\{1, 2, \dots, k\}$, tem-se que

$$(\theta_{\alpha_1} \dots \theta_{\alpha_\ell} | 1 - \sum_{j=1}^{\ell} \theta_{\alpha_j}) | \underline{b} \rightsquigarrow D_{\ell+1} [(b_{\alpha_1} \dots b_{\alpha_\ell} \quad b.-\sum_{j=1}^{\ell} b_{\alpha_j})] \quad (5)$$

Em particular, as distribuições marginais univariadas são

$$(\theta_i, 1-\theta_i) | \underline{b} \rightsquigarrow D_2 [(b_i \quad b.-b_i)], \quad i = 1, \dots, k \quad (6)$$

o que corresponde, na notação mais usual, às distribuições beta,

$$\theta_i | \underline{b} \rightarrow B_e(b_i, b.-b_i).$$

Demonstra-se em seguida uma importante caracterização da distribuição de Dirichlet, a qual constitui uma generalização do lema 1 de Basu and Pereira (1982). Dickey et al. (1987) enunciam a condição necessária (lema 4.2) apresentando-a como uma versão dos resultados de Wilks (1962).

Para o efeito, seja $\{\delta_1, \delta_2, \dots, \delta_\ell\}$, $\ell \geq 2$, uma partição do conjunto dos índices $1, 2, \dots, k$, indicado pela matriz $k \times \ell$ Z . Ou seja, $Z = (z_1, z_2, \dots, z_\ell)$ é uma matriz de zeros e uns, sem linhas nulas e com as colunas z_j ortogonais definidas de modo que $z_{ji} = 1$ se $i \in \delta_j$ e $z_{ji} = 0$, no caso contrário.

Teorema C.1.1. Seja $\underline{\theta} = (\theta_1, \dots, \theta_k)'$ com $\theta_i \geq 0$ e $\mathbf{1}'\underline{\theta} = 1$ e $\underline{\theta}(\delta_j)$, $j=1, \dots, \ell$ o vector dos elementos θ_i tais que $i \in \delta_j$, de dimensão $d_j \times 1$ (d_j é o cardinal de δ_j), onde $\{\delta_1, \dots, \delta_\ell\}$ define uma partição de $\{1, \dots, k\}$ representada pela matriz $Z = (z_1, \dots, z_\ell)$. Definindo $\underline{\phi} = (\phi_1, \dots, \phi_\ell)'$ e $\rho(j) = \underline{\theta}(\delta_j) / \phi_j = (\rho_i(j), i \in \delta_j)'$, $j=1, \dots, \ell$, então

$$\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b}) \text{ sse } \begin{cases} \underline{\phi} | \underline{b} \rightsquigarrow D_\ell(Z' \underline{b}) \\ \rho(j) | \underline{b} \rightsquigarrow D_{d_j}(\underline{b}(\delta_j)), j=1, \dots, \ell \\ \underline{\phi} \perp \perp \rho(1) \perp \perp \dots \perp \perp \rho(\ell) | \underline{b} \end{cases}$$

onde $\underline{b} = (b_1, \dots, b_k)'$ e $\underline{b}(\delta_j)$ é o vector dos b_i 's definido em correspondência com $\underline{\theta}(\delta_j)$.

Demonstração - (Parte necessária): Supondo $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$, sejam γ_i , $i=1, \dots, k$ variáveis aleatórias independentemente distribuídas como

$\Gamma(b_i, 1)$ tais que $\theta_i = \gamma_i / \sum_{j=1}^k \gamma_j$. Definamos $\underline{\alpha} = (\alpha_1, \dots, \alpha_\ell)'$ e $\underline{\alpha} = Z' \underline{\gamma}$ onde

$\underline{\gamma} = (\gamma_1, \dots, \gamma_k)'$. Então $\phi_j = \alpha_j / \sum_{m=1}^{\ell} \alpha_m$, ou seja, $\underline{\phi}$ é um vector de coordenadas $\alpha_j / \sum_{m=1}^{\ell} \alpha_m$, $j=1, \dots, \ell$, onde os $\{\alpha_j = z_j' \underline{\gamma}\}$ são independentemente

distribuídos como $\Gamma(z_j' \underline{b}, 1)$. Por conseguinte, $\underline{\phi} | \underline{b} \rightsquigarrow D_\ell(Z' \underline{b})$.

De acordo com o mesmo argumento, o facto das componentes de $\rho(j), \forall j$, poderem ser definidas por $\gamma_i / \sum_{m \in \delta_j} \gamma_m$, $i \in \delta_j$ implica que, fixado \underline{b} , os vec-

tores $\{\rho(j), j=1, \dots, \ell\}$ são independentemente distribuídos segundo as

distribuições $D_{d_j}(\underline{b}(\delta_j))$, $j=1, \dots, \ell$.

Dado que (1) se pode escrever como

$$g(\underline{\theta} | \underline{b}) = \frac{\Gamma(\sum_{j=1}^{\ell} z_j' \cdot \underline{b})}{\prod_{j=1}^{\ell} \Gamma(b_j)} \prod_{j=1}^{\ell} \prod_{i \in \delta_j} \theta_i^{b_i - 1} \quad (7)$$

e que a função densidade de ϕ é

$$g_0(\phi | \underline{b}) = \frac{1}{B(\underline{z}' \underline{b})} \prod_{j=1}^{\ell} \phi_j^{z_j' \cdot \underline{b} - 1}, \quad \phi \in \mathcal{S}_{\ell-1} \quad (8)$$

resulta que a função densidade de θ condicional em ϕ é definida por

$$g^*(\theta | \phi; \underline{b}) = \prod_{j=1}^{\ell} [g_j(\underline{\theta}_{(\delta_j)} / \phi_j) \cdot \phi_j^{-(d_j - 1)}] \quad (9)$$

onde

$$g_j(\underline{\theta}_{(\delta_j)} / \phi_j) = \frac{1}{B(b_{(\delta_j)})} \prod_{i \in \delta_j} (\theta_i / \phi_j)^{b_i - 1} \quad (10)$$

Note-se que (9) define uma distribuição $(k-\ell)$ -variada enquanto que (7) traduz uma distribuição $(k-1)$ -variada. Como a transformação $\underline{\theta}_{(\delta_j)} = \phi_j \underline{\rho}_{(j)}$, $j=1, \dots, \ell$, encarada com ϕ fixo e como uma relação entre $k-\ell$ variáveis θ_i e ρ_j^i , tem como matriz jacobiana a matriz diagonal em blocos de blocos diagonais $\phi_j \cdot I_{d_j - 1}$, $j=1, \dots, \ell$, o respectivo jacobiano é $\prod_{j=1}^{\ell} \phi_j^{d_j - 1}$. Deste modo, a função densidade de $\rho_{(1)}, \dots, \rho_{(\ell)}$ da do ϕ é

$$h(\rho_{(1)}, \dots, \rho_{(\ell)} | \phi; \underline{b}) = \prod_{j=1}^{\ell} g_j(\rho_{(j)}) \quad (11)$$

onde $g_j(\rho_{(j)})$ representa a função densidade marginal de $\rho_{(j)}$. Este resultado reflete então que $\phi \perp \perp \rho_{(1)} \perp \perp \dots \perp \perp \rho_{(\ell)} | \underline{b}$, o que termina a demonstração da parte necessária.

(Parte suficiente): Sendo dados os resultados distribucionais sobre ϕ e $\rho_{(j)}$, $j=1, \dots, \ell$, a função densidade de θ pode exprimir-se por

$$g(\theta | \underline{b}) = g_0(\phi | \underline{b}) p(\theta | \phi; \underline{b}), \quad \phi = \underline{z}' \underline{\theta} \quad (12)$$

onde $g_0(\cdot)$ é definida em (8). A função $p(\cdot|\cdot)$ determina-se a partir de (11), tendo em conta que o jacobiano da transformação $(\varrho_{(1)}, \dots, \varrho_{(\ell)}) \rightarrow \underline{\theta}$

com $\underline{\phi}$ fixo é dado por $\prod_{j=1}^{\ell} \phi_j^{-(d_j-1)}$, obtendo-se naturalmente

$p(\underline{\theta}|\underline{\phi}; \underline{b}) = g^*(\underline{\theta}|\underline{\phi}; \underline{b})$. Deste modo, (12) pode escrever-se como

$$g(\underline{\theta}|\underline{b}) = g_0(\underline{z}'\underline{\theta}|\underline{b}) \prod_{j=1}^{\ell} [g_j(\underline{\theta}(\delta_j)/z_j'\underline{\theta}) \cdot (z_j'\underline{\theta})^{-(d_j-1)}] \quad (13)$$

expressão que corresponde, por (8) e (10), a uma factorização de (7). Assim, $\underline{\theta}|\underline{b} \rightsquigarrow D_k(\underline{b})$ como queríamos demonstrar. \square

Um outro resultado importante concernente à distribuição Dirichlet, útil na prática para o cálculo dos seus momentos e aplicável às distribuições Dirichlet generalizadas (vide C.3) é enunciado em Dickey (1983, teor. 2.1). Em ordem a demonstrá-lo, consideremos o modelo de urna de Pólya, em cuja urna há c_i bolas da cor i , $i=1, \dots, k$ ($k \geq 2$) e onde a bola obtida em cada extracção aleatória simples é repostada juntamente com s ($s > 0$) bolas da mesma cor.

Sendo $\underline{x} = (x_1, \dots, x_k)'$ o vector representando o número de bolas de cada cor obtidas em n extracções, é fácil constatar que a sua função de probabilidade é dada por

$$f(\underline{x}|n, \{c_i\}) = \binom{n}{\underline{x}} \times \frac{\prod_{i=1}^k c_i(c_i+s) \dots [c_i+(x_i-1)s]}{c \cdot (c+s) \dots [c+(n-1)s]} \quad (14)$$

onde $\sum_{i=1}^k x_i = n$, $\binom{n}{\underline{x}} = \frac{n!}{\prod_{i=1}^k x_i!}$ e $c = \sum_{i=1}^k c_i$. O segundo factor de (14),

que denotaremos por $p_n(\underline{x})$, indica a probabilidade de se obter numa dada ordem, qualquer que ela seja, x_i bolas da cor i , $i=1, \dots, k$, em n extracções aleatórias individuais.

A expressão (14) define a denominada distribuição de Pólya $(k-1)$ -variada. Fazendo $b_i = \frac{c_i}{s}$, $i=1, \dots, k$ e tendo em conta que

$a(a+1) \dots (a+m-1) = \frac{\Gamma(a+m)}{\Gamma(a)}$, $a \in \mathbb{R}_+$, $m \in \mathbb{N}$, resulta

$$f(\underline{x}|n, \underline{b}) = \binom{n}{\underline{x}} \frac{B(\underline{b}+\underline{x})}{B(\underline{b})} \quad (15)$$

onde $\underline{b} = (b_1, \dots, b_k)'$. Sendo $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$, por (2)

$$f(\underline{x} | n, \underline{b}) = E_{\underline{\theta} | \underline{b}} \left(\binom{n}{\underline{x}} \prod_{i=1}^k \theta_i^{x_i} \right) \quad (16)$$

o que evidencia que a distribuição de Pólya (k-1)-variada é a mistura da distribuição $\underline{x} | n, \underline{\theta} \rightsquigarrow M_k(n, \underline{\theta})$ com a referida distribuição Dirichlet para $\underline{\theta}$. Esta mistura, onde as coordenadas de \underline{b} não são necessariamente números racionais, é vulgarmente designada por distribuição Dirichlet-Multinomial, escrevendo-se $\underline{x} | n, \underline{b} \rightsquigarrow DM_k(n, \underline{b})$.

Num contexto Bayesiano, a distribuição de Pólya é assim a distribuição preditiva a priori de um vector de frequências multinomial sob a distribuição a priori conjugada.

Seja \underline{y}_j o vector (k×1) indicando a cor saída na extração j (j=1, ..., n), i.e., $\underline{y}_j = \underline{e}_i$ onde \underline{e}_i é o vector com todas as coordenadas nulas exceptuando a i-ésima que é 1, se for extraída uma bola da cor i.

Então $\underline{x} = \sum_{j=1}^n \underline{y}_j$ e $\underline{y}_j | \underline{\theta}$, j=1, ..., n $\overset{iid}{\rightsquigarrow} M_k(1, \underline{\theta})$, pelo que

$$f(\underline{y}_1, \dots, \underline{y}_n | \underline{\theta}) = \prod_{j=1}^n (\underline{y}_j' \cdot \underline{\theta}) = \prod_{i=1}^k \theta_i^{x_i}. \quad \text{Quando } \underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b}), \text{ temos}$$

$$\underline{y}_j | \underline{b} \rightsquigarrow DM_k(1, \underline{b}), \quad j=1, \dots, n \quad (17)$$

e

$$f(\underline{y}_1, \dots, \underline{y}_n | \underline{b}) = \frac{B(\underline{b} + \underline{x})}{B(\underline{b})}$$

O momento misto Dirichlet em (15) pode então exprimir-se por

$$\begin{aligned} E_{\underline{\theta} | \underline{b}} \left(\prod_{i=1}^k \theta_i^{x_i} \right) &\equiv p_n(\underline{x}) \\ &= E_{\underline{\theta} | \underline{b}} \left(\prod_{j=1}^n (\underline{y}_j' \cdot \underline{\theta}) \right) \end{aligned} \quad (18)$$

Como $p_n(\underline{x})$ não depende da ordem em que ocorrem as x_i bolas, $i=1, \dots, k$, vamos admitir que nas primeiras x_1 extracções saíram bolas da cor 1; nas seguintes x_2 , bolas da cor 2, e assim sucessivamente, até se perfazerem as n extracções. Deste modo

$$y_{\underline{j}}' \theta = \begin{cases} e_{\underline{1}}' \theta = \theta_1, & j = 1, \dots, x_1 \\ e_{\underline{2}}' \theta = \theta_2, & j = x_1 + 1, \dots, x_1 + x_2 \\ \dots \dots \dots k-1 \\ e_{\underline{k}}' \theta = \theta_k, & j = \sum_1 x_i + 1, \dots, n \end{cases}$$

e $p_n(\underline{x})$ em (15) é, com $b_{\cdot} = \underline{1}' b$, o produto

$$p_n(\underline{x}) = \frac{b_{\cdot}}{b_{\cdot}} \frac{b_{\cdot} + 1}{b_{\cdot} + 1} \dots \frac{b_{\cdot} + x_1 - 1}{b_{\cdot} + x_1 - 1} \frac{b_{\cdot}}{b_{\cdot} + x_1} \frac{b_{\cdot} + 1}{b_{\cdot} + x_1 + 1} \dots \frac{b_{\cdot} + x_2 - 1}{b_{\cdot} + x_1 + x_2 - 1} \dots \frac{b_{\cdot} + x_k - 1}{b_{\cdot} + n - 1} \quad (19)$$

Em virtude de

$$\theta_{\underline{j}} | y_{\underline{j}}, j \in \Delta \rightsquigarrow D_k(b_{\cdot} + \sum_{j \in \Delta} y_{\underline{j}}) \quad (20)$$

onde Δ é qualquer subconjunto dos índices $1, 2, \dots, n$, segue-se que a distribuição preditiva à posteriori de $y_{\underline{m}}$, $m \notin \Delta$ é

$$y_{\underline{m}} | y_{\underline{j}}, j \in \Delta \rightsquigarrow DM_k(1, b_{\cdot} + \sum_{j \in \Delta} y_{\underline{j}}) \quad (21)$$

Deste modo, quando $y_{\underline{m}} = e_{\underline{i}}$, $i = 1, \dots, k$

$$\begin{aligned} E(y_{\underline{m}}' \theta | y_{\underline{j}}, j \in \Delta) &= \frac{b_{\underline{i}} + \sum_{j \in \Delta} y_{\underline{j}i}}{b_{\cdot} + \sum_{i=1}^k \sum_{j \in \Delta} y_{\underline{j}i}} \equiv \frac{B(b_{\cdot} + \sum_{j \in \Delta} y_{\underline{j}} + e_{\underline{i}})}{B(b_{\cdot} + \sum_{j \in \Delta} y_{\underline{j}})} \\ &= f_{y_{\underline{m}} | y_{\underline{j}}, j \in \Delta}(e_{\underline{i}}) \end{aligned} \quad (22)$$

Em face deste resultado, o factor m ($m = 1, \dots, n$) de $p_n(\underline{x})$ em (19) pode exprimir-se como a média à posteriori do respectivo $\theta_{\underline{i}}$ traduzido por $y_{\underline{m}}' \theta$, ou alternativamente, como a probabilidade preditiva à posteriori do valor de $y_{\underline{m}}$, em ambos os casos com os valores de $y_{\underline{j}}$, $1 \leq j < m$ tomados como "dados". Assim

Proposição C.1.2 : Todo o momento misto Dirichlet, ao ser expresso por (18), pode ser calculado através do seguinte produto de n factores envolvendo sucessivas médias Dirichlet ou probabilidades Dirichlet-Multinomiais

$$\begin{aligned}
 E_{\theta|b} \left[\prod_{i=1}^k \theta_i^{x_i} \right] &= E_{\theta|b} \left[\prod_{j=1}^n (y'_j \cdot \theta) \right] \\
 &= E_{\theta|b} (\theta_1) \quad E_{\theta|y(1)} (\theta_1) \dots E_{\theta|y(x_1-1)} (\theta_1) \quad E_{\theta|y(x_1)} (\theta_2) \quad E_{\theta|y(x_1+1)} (\theta_2) \quad \dots \\
 &\quad \dots E_{\theta|y(x_1+x_2-1)} (\theta_2) \quad \dots E_{\theta|y(x_k)} (\theta_k) \\
 &= f_{y_1} (e_1) f_{y_2|y(1)} (e_1) \dots f_{y_{x_1}|y(x_1-1)} (e_1) \quad f_{y_{x_1+1}|y(x_1)} (e_2) \quad f_{y_{x_1+2}|y(x_1+1)} (e_2) \quad \dots \\
 &\quad \dots f_{y_{x_1+x_2}|y(x_1+x_2-1)} (e_2) \quad \dots f_{y_n|y(n-1)} (e_k) \quad (23)
 \end{aligned}$$

onde $y_{(\ell)} = (y_1, \dots, y_\ell)$, $1 \leq \ell < n$. \square

A visualização da distribuição Dirichlet-Multinomial como uma mistura permite derivar facilmente algumas das suas propriedades através de propriedades correspondentes das distribuições misturada e misturadora. No que concerne aos primeiros momentos de $x|n, b \rightsquigarrow DM_k(n, b)$, obtêm-se

$$\begin{aligned}
 E(x_i) &= nb_i/b, \quad i=1, \dots, k \\
 \text{Var}(x_i) &= \frac{n+b}{b \cdot (b+1)} n \left(b_i - \frac{b_i^2}{b} \right), \quad i=1, \dots, k \\
 \text{Cov}(x_i, x_j) &= \frac{n+b}{b \cdot (b+1)} n \left(-\frac{b_i b_j}{b} \right), \quad i \neq j
 \end{aligned}$$

pelo que

$$E(\underline{x}) = n\underline{b}/b. \equiv E(\omega) \quad (24)$$

$$\text{Var}(\underline{x}) = \frac{n+b.}{b.(b.+1)} n \left(\underline{D}_b - \frac{\underline{b}\underline{b}'}{b.} \right) \equiv \frac{n+b.}{b.(b.+1)} \text{Var}(\omega)$$

onde $\omega | n, \underline{b} \rightsquigarrow M_k(n, \underline{b}/b.)$.

Relativamente às distribuições marginais, sendo $(\alpha_1, \dots, \alpha_\ell)$ uma permutação de ℓ ($1 \leq \ell \leq k-1$) índices de $\{1, 2, \dots, k\}$ tem-se que

$$(x_{\alpha_1} \dots x_{\alpha_\ell} | n - \sum_{i=1}^{\ell} x_{\alpha_i}) | n, \underline{b} \rightsquigarrow DM_{\ell+1} [n, (b_{\alpha_1} \dots b_{\alpha_\ell} b. - \sum_{i=1}^{\ell} b_{\alpha_i})] \quad (26)$$

donde, em particular,

$$(x_i | n - x_i) | n, \underline{b} \rightsquigarrow DM_2 [n, (b_i b. - b_i)] \quad , i=1, \dots, k \quad (27)$$

o que traduz as vulgarmente designadas distribuições beta-binomiais para as variáveis x_i .

Pelo facto de

$$(x_{\alpha_1} \dots x_{\alpha_\ell} | \sum_{i=1}^{\ell} x_{\alpha_i}; n, \theta \rightsquigarrow M_\ell [\sum_{i=1}^{\ell} x_{\alpha_i}, (\theta_{\alpha_1} \dots \theta_{\alpha_\ell}) / \sum_{i=1}^{\ell} \theta_{\alpha_i}]$$

e de

$$(\theta_{\alpha_1} \dots \theta_{\alpha_\ell} | \sum_{i=1}^{\ell} \theta_{\alpha_i} | b \rightsquigarrow D_\ell [(b_{\alpha_1} \dots b_{\alpha_\ell})]$$

resulta

$$(x_{\alpha_1} \dots x_{\alpha_\ell} | \sum_{i=1}^{\ell} x_{\alpha_i}; n, \underline{b} \rightsquigarrow DM_\ell [\sum_{i=1}^{\ell} x_{\alpha_i}, (b_{\alpha_1} \dots b_{\alpha_\ell})] \quad (28)$$

Basu and Pereira (1982, teor. 1) estabelecem uma caracterização da distribuição $\underline{x} | n, \underline{b} \rightsquigarrow DM_k(n, \underline{b})$ com base numa partição de \underline{x} em dois sub-vectores componentes. Na análise Bayesiana do problema geral de dados incompletos do Capítulo 5 necessitamos de uma generalização dessa caracterização, que demonstraremos em seguida.

Teorema C.1.3 : Seja $\underline{x} = (x_1, \dots, x_k)'$ um vector inteiro não negativo com $1' \underline{x} = n$ fixo e $\underline{x}_{(\delta_j)}$, $j=1, \dots, \ell$ ($\ell \geq 2$) o vector das variáveis x_i tais que $i \in \delta_j$, de dimensão $d_j \times 1$ (d_j é o cardinal de δ_j), onde $\{\delta_1, \dots, \delta_\ell\}$ define uma partição de $\{1, \dots, k\}$ indicada pela matriz $k \times \ell$ $\underline{Z} = (z_1, \dots, z_\ell)$. Definindo $\underline{v} = (v_1 \dots v_\ell) = \underline{Z}' \underline{x}$, então

$$\underline{x} | n, \underline{b} \rightsquigarrow DM_k(n, \underline{b}) \quad \text{sse} \quad \begin{cases} \underline{v} | n, \underline{b} \rightsquigarrow DM_\ell(n, \underline{Z}' \underline{b}) \\ \underline{x}_{(\delta_j)} | v_j, \underline{b} \rightsquigarrow DM_{d_j}(v_j, \underline{b}_{(\delta_j)}), \quad j=1, \dots, \ell \\ \underline{x}_{(\delta_1)} \perp\!\!\!\perp \underline{x}_{(\delta_2)} \perp\!\!\!\perp \dots \perp\!\!\!\perp \underline{x}_{(\delta_\ell)} | \underline{v} \end{cases}$$

onde $\underline{b} = (b_1, \dots, b_k) \in \mathbb{R}_+^k$ e $\underline{b}_{(\delta_j)}$, $j=1, \dots, \ell$ é definido de modo consentâneo com $\underline{x}_{(\delta_j)}$.

Demonstração - (Parte necessária): Supondo $\underline{x} | n, \underline{b} \rightsquigarrow DM_k(n, \underline{b})$, seja $\underline{\theta} = (\theta_1, \dots, \theta_k)'$ o vector com distribuição $D_k(\underline{b})$ tal que $\underline{x} | n, \underline{\theta} \rightsquigarrow M_k(n, \underline{\theta})$.

Como consequência das propriedades das distribuições multinomial e Dirichlet temos

$$\begin{aligned} \underline{v} | n, \underline{\theta} \rightsquigarrow M_\ell(n, \underline{Z}' \underline{\theta}) ; \quad \underline{Z}' \underline{\theta} | \underline{b} \rightsquigarrow D_\ell(\underline{Z}' \underline{b}) \\ \underline{x}_{(\delta_j)} | \underline{z}_j' \underline{x}; \underline{\theta} \rightsquigarrow M_{d_j}(\underline{z}_j' \underline{x}, \underline{\theta}_{(\delta_j)} / \underline{z}_j' \underline{\theta}), \quad \underline{\theta}_{(\delta_j)} = (\theta_i, i \in \delta_j)', \quad j=1, \dots, \ell \\ \frac{\underline{\theta}_{(\delta_j)}}{\underline{z}_j' \underline{\theta}} | \underline{b} \rightsquigarrow D_{d_j}(\underline{b}_{(\delta_j)}), \quad j=1, \dots, \ell \end{aligned}$$

o que implica que

$$\underline{v} | n, \underline{b} \rightsquigarrow DM_\ell(n, \underline{Z}' \underline{b})$$

e

$$\underline{x}_{(\delta_j)} | v_j, \underline{b} \rightsquigarrow DM_{d_j}(v_j, \underline{b}_{(\delta_j)}), \quad j=1, \dots, \ell$$

como pretendíamos. Finalmente, de (15) e da função densidade de \underline{v} , concluímos que a função densidade de $(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)})$ condicionada em $\underline{v} = \underline{Z}' \underline{x}$ é

$$h(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)} | \underline{V}; \underline{b}) = \prod_{j=1}^{\ell} \left(\binom{\underline{v}_j}{\underline{x}_{(\delta_j)}} \frac{B(\underline{b}_{(\delta_j)} + \underline{x}_{(\delta_j)})}{B(\underline{b}_{(\delta_j)})} \right)$$

o que mostra que $\underline{x}_{(\delta_1)} \perp\!\!\!\perp \dots \perp\!\!\!\perp \underline{x}_{(\delta_\ell)} | \underline{V}$.

(Parte suficiente): Denotando $g(\cdot)$ e $g_0(\cdot)$ as funções densidade de $(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)})$ e de \underline{V} e pelo facto de

$$g(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)} | \underline{n}, \underline{b}) = g_0(\underline{V} | \underline{n}, \underline{b}) h(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)} | \underline{V}; \underline{b}) \text{ para } \underline{V} = \underline{Z}' \underline{x}$$

resulta pela hipótese que

$$g(\underline{x}_{(\delta_1)}, \dots, \underline{x}_{(\delta_\ell)} | \underline{n}, \underline{b}) = \binom{\underline{n}}{\underline{Z}' \underline{x}} \frac{B(\underline{Z}'(\underline{b} + \underline{x}))}{B(\underline{Z}' \underline{b})} \prod_{j=1}^{\ell} \left(\binom{\underline{z}'_j \underline{x}}{\underline{x}_{(\delta_j)}} \frac{B(\underline{b}_{(\delta_j)} + \underline{x}_{(\delta_j)})}{B(\underline{b}_{(\delta_j)})} \right)$$

Como o 2º membro desta relação corresponde a uma factorização de (15) o teorema fica completamente demonstrado. \square

o|o
~

APÊNDICE C.2. - FUNÇÕES HIPERGEOMÉTRICAS

C.2.1. Séries hipergeométricas

Sendo $a \in \mathbb{R}$ e $m-1 \in \mathbb{N}$

$$(a, m) = a(a+1) \dots (a+m-1)$$

define o factorial ascendente através do denominado símbolo de Appell. Para os outros valores inteiros de m considera-se

$$(a, 0) = 1, \quad \forall a \in \mathbb{R}$$

$$(a, -m) = [(a-1)(a-2) \dots (a-m)]^{-1}, \quad a \neq 1, 2, \dots, m, \quad m \in \mathbb{N}$$

Note-se que quando $a = -n$, $n \in \mathbb{N}$

$$(-n, m) = 0 \quad \text{para } 0 \leq n < m$$

Definindo o coeficiente binomial para $a \in \mathbb{R}$ por

$$\binom{a}{n} = \begin{cases} \frac{a(a-1)\dots(a-n+1)}{n!}, & n \in \mathbb{N} \\ 1 & , n = 0 \end{cases}$$

segue-se imediatamente que

$$\binom{a}{n} = (-1)^n \frac{(-a, n)}{n!}$$

Uma série hipergeométrica é uma série de potências $\sum_{m=0}^{\infty} A_m z^m$ tal que A_{m+1}/A_m é uma função racional de m . O exemplo mais conhecido é o da série hipergeométrica de Gauss, vulgarmente denotada por

$${}_2F_1(a, b; c; z) = \sum_{m=0}^{\infty} \frac{(a, m)(b, m)}{(c, m)} \frac{z^m}{m!}, \quad c \neq 0, -1, -2, \dots$$

O uso do teste da razão permite concluir que ela converge absolutamente em $(-1, 1)$ e diverge no exterior deste intervalo. Nos extremos do intervalo de convergência pode convergir ou divergir, dependendo dos valores dos parâmetros.

Note-se que esta s erie inclui como caso particular ($a=1, b=c$) a s erie geom etrica; da  o nome a ela atribuido por Gauss. Por outro lado,   f acil constatar que a fun ao geradora de probabilidade da distribu ao hipergeom trica $X \rightsquigarrow \text{Hpg}(N_1, N_2, n)$  , com $N = N_1 + N_2$

$$\prod_X(z) \equiv E(z^X) = \sum_{x=\max(0, n-N_2)}^{\min(N_1, n)} \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}} z^x = C {}_2F_1(-n, -N_1; N_2 - n + 1; z)$$

onde $C = \binom{N_2}{n} / \binom{N}{n}$. Esta s erie termina no $(n+1)$ - esimo termo j  que $(-n, x) = 0$ para $x > n$.

A classe de s eries hipergeom tricas (simples) pode definir-se por

$${}_pF_q(\underline{a}; \underline{c}; z) = \sum_{m=0}^{\infty} \frac{\prod_{i=1}^p (a_i, m)}{\prod_{i=1}^q (c_i, m)} \frac{z^m}{m!}, \quad c_i \notin \{0, -1, -2, \dots\}, i=1, \dots, q$$

onde $\underline{a} = (a_1, \dots, a_p)$ e $\underline{c} = (c_1, \dots, c_q)$.

Como

$$\frac{A_{m+1}}{A_m} = \frac{m^{p-q-1} \prod_{i=1}^p (1+a_i/m)}{1+1/m \prod_{i=1}^q (1+c_i/m)}$$

o teste da raz o mostra que se $p=q+1$ a s erie converge absolutamente para $|z| < 1$ e diverge para $|z| > 1$.   o que acontece com a s erie ${}_2F_1$ e tamb m com a s erie binomial

$${}_1F_0(a; z) = \sum_{m=0}^{\infty} (a, m) \frac{z^m}{m!} = (1-z)^{-a}, \quad |z| < 1$$

Quando $p < q+1$, $\lim_{m \rightarrow \infty} \left| \frac{A_{m+1}}{A_m} z \right| = 0$ pelo que a s erie converge absolutamente para qualquer $z \in \mathbb{R}$, definindo portanto uma fun ao inteira. Um exemplo imediato   a fun ao exponencial

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!} = {}_0F_0(z)$$

Desta representação em série podemos deduzir que

$$g(z) \equiv \int_0^z e^{-t^2} dt = \sum_{m=0}^{\infty} \frac{(-1)^m z^{2m+1}}{(2m+1)m!} = z {}_1F_1(1/2; 3/2; -z^2)$$

atendendo a que $(2m+1)^{-1} = (1/2, m)/(3/2, m)$. A função erro $g(z)$ é assim um caso especial da série hipergeométrica confluyente

$${}_1F_1(a; c; z) = \sum_{m=0}^{\infty} \frac{(a, m)}{(c, m)} \frac{z^m}{m!}, \quad c \notin \{0, -1, -2, \dots\}, \quad z \in \mathbb{R}$$

Uma série hipergeométrica dupla é uma série

$$\sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} A_{m_1, m_2} z_1^{m_1} z_2^{m_2}$$

tal que $A_{m_1+1, m_2}/A_{m_1, m_2}$ e $A_{m_1, m_2+1}/A_{m_1, m_2}$ são funções racionais de m_1 e m_2 . De modo idêntico se definem as séries hipergeométricas múltiplas de mais de 2 variáveis.

Uma função hipergeométrica é precisamente a soma de uma série hipergeométrica no interior do respectivo círculo de convergência, sendo usada a mesma designação para o seu prolongamento analítico fora do círculo de convergência.

Muitas das mais importantes funções especiais da matemática aplicada podem ser representadas por integrais de funções elementares em relação a medidas de Dirichlet, e há vantagens significativas em defini-las deste modo em vez de através de séries de potências hipergeométricas. Por exemplo, a função hipergeométrica ${}_1F_1(b_1; b_1+b_2; z)$, $b_1, b_2 > 0$, traz a função geradora de momentos da distribuição beta

$$d\lambda_{(b_1, b_2)}(\theta) = [\theta^{b_1-1} (1-\theta)^{b_2-1} / B(b_1, b_2)] d\theta. \quad \text{Com efeito,}$$

$${}_1F_1(b_1; b_1+b_2; z) = \sum_{m=0}^{\infty} \frac{(b_1, m)}{(b_1+b_2, m)} \frac{z^m}{m!} = \sum_{m=0}^{\infty} \frac{z^m}{m!} \frac{(b_1, m) (b_2, 0)}{(b_1+b_2, m)}$$

Atendendo a que

$$(a, m) = \frac{\Gamma(a+m)}{\Gamma(a)}, \quad m \in \mathbb{Z}, \quad a, a+m \notin \{0, -1, -2, \dots\}$$

e a propriedades da distribuição beta e das séries de potências uniformemente convergentes, temos sucessivamente

$$\begin{aligned} {}_1F_1(b_1; b_1+b_2; z) &= \sum_{m=0}^{\infty} \frac{z^m}{m!} \frac{B(b_1+m, b_2)}{B(b_1, b_2)} \\ &= \sum_{m=0}^{\infty} \frac{z^m}{m!} \int_0^1 \theta^m d\lambda_{(b_1, b_2)}(\theta) = \\ &= \int_0^1 \sum_{m=0}^{\infty} \frac{(z\theta)^m}{m!} d\lambda_{(b_1, b_2)}(\theta) = E_{\theta|b_1, b_2}(e^{z\theta}) \end{aligned}$$

Outro exemplo importante é o da série hipergeométrica de Gauss ${}_2F_1(a, b_1; b_1+b_2; z)$, $b_1, b_2 > 0$, $a \in \mathbb{R}$, que se pode exprimir como uma média beta de uma função potência:

$$\begin{aligned} {}_2F_1(a, b_1; b_1+b_2; z) &= \sum_{m=0}^{\infty} \frac{(a, m)(b_1, m)}{(b_1+b_2, m)} \frac{z^m}{m!}, \quad |z| < 1 \\ &= \sum_{m=0}^{\infty} (a, m) \frac{z^m}{m!} \int_0^1 \theta^m d\lambda_{(b_1, b_2)}(\theta) = \int_0^1 \sum_{m=0}^{\infty} (a, m) \frac{(z\theta)^m}{m!} d\lambda_{(b_1, b_2)}(\theta) \\ &= E_{\theta|b_1, b_2} [(1-z\theta)^{-a}], \quad |z| < 1 \quad (29) \end{aligned}$$

Note-se que neste caso a representação integral tem sentido para todo o $a \in \mathbb{R}$ a não ser que $z \geq 1$. Por isso, (29) define o prolongamento analítico da função hipergeométrica a $\{z \in \mathbb{R}: z < 1\}$.

Estes dois exemplos ilustram a importância das funções exponencial e potência na geração de funções especiais, particularmente de tipo hipergeométrico.

C.2.2. Funções hipergeométricas R de Carlson

Seja $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \mathbb{R}_+^k (\mathbb{R}_+ = (0, \infty))$ com $\sum_{i=1}^k \theta_i = 1$ um vector tal que $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$, onde $\underline{b} = (b_1, \dots, b_k) \in \mathbb{R}_+^k$ e denote-se por $\lambda_{\underline{b}}$ a respectiva medida de Dirichlet definida no simplex $(k-1)$ -dimensional, S_{k-1} , de \mathbb{R}^k . Seja $f(x)$ uma função real definida num intervalo aberto I de \mathbb{R} , mensurável e integrável relativamente a $\lambda_{\underline{b}}$.

Sendo $\underline{z} = (z_1, \dots, z_k)' \in \mathbb{R}^k$, o produto interno $\underline{z}'\underline{\theta}$ representa quer uma combinação convexa de $\{z_1, \dots, z_k\}$ quer uma função linear de $k-1$ variáveis θ_i . Deste último ponto de vista, $\underline{z}'\underline{\theta}$ transforma S_{k-1} no invólucro convexo dos k pontos z_i - denotado por $I(\underline{z})$ - ou seja, no intervalo $[\min_i z_i, \max_i z_i]$.

Supondo que $I(\underline{z}) \subset I$, Carlson (1977, cap. 5) obtém uma transformada de f através de

$$F(\underline{b}; \underline{z}) = \int_{S_{k-1}} f(\underline{z}'\underline{\theta}) d\lambda_{\underline{b}}(\underline{\theta}) \equiv E_{\underline{\theta} | \underline{b}}^{(k-1)} [f(\underline{z}'\underline{\theta})] \quad (30)$$

onde o índice superior no valor esperado indica a dimensionalidade do respectivo integral. Deve-se observar que o tratamento de Carlson envolve funções, variáveis e parâmetros complexos mas restringir-nos-emos ao caso real por ser suficiente para as aplicações estatísticas.

A nova função F de k variáveis z_i e k parâmetros b_i é assim uma média Dirichlet de f sobre $I(\underline{z})$. Note-se que se $f(x)$ é contínua em I , $f(\underline{z}'\underline{\theta})$ é, para cada \underline{z} , limitada em S_{k-1} , o que assegura a existência de F . Contudo, para tal não se afigura necessária a continuidade de $f(x)$.

Pelas propriedades da distribuição de Dirichlet, F é invariante face a permutações simultâneas das variáveis e parâmetros, e tem dimensionalidade reduzida quando algumas das coordenadas de \underline{z} são iguais. Na verdade, se δ for um subconjunto de p elementos de $\{1, \dots, k\}$ tal que $z_i = c, i \in \delta$, considerando $\underline{z}^* = (z_i, i \notin \delta; c)'$, $\underline{\theta}^* = (\theta_i, i \notin \delta; \sum_{i \in \delta} \theta_i)'$ e $\underline{b}^* = (b_i, i \notin \delta; \sum_{i \in \delta} b_i)'$, obtém-se

$$F(\underline{b}; \underline{z}) = E_{\underline{\theta} | \underline{b}}^{(k-1)} [f(\underline{z}'\underline{\theta})] = E_{\underline{\theta}^* | \underline{b}^*}^{(k-p)} [f(\underline{z}^*\underline{\theta}^*)] = F(\underline{b}^*; \underline{z}^*) \quad (31)$$

Se adicionalmente $z_i = d$, $i \in \delta$, F é definida por um integral unidimensional, já que

$$F(\underline{b}; \underline{z}) = F\left(\sum_{i \in \delta} b_i, \sum_{i \in \delta} b_i; d, c\right)$$

No caso extremo em que $z_i = c$, $\forall i$, então

$$F(\underline{b}; c \mathbf{1}_k) = f(c) \quad (32)$$

Como exemplos de médias Dirichlet, se tomarmos $f(x) = e^x$ com $I = \mathbb{R}$, obtemos a função S de Carlson

$$S(\underline{b}, \underline{z}) = E_{\underline{\theta} | \underline{b}} [e^{\underline{z}' \underline{\theta}}] \quad , \quad \underline{z} \in \mathbb{R}^k$$

a qual, com \underline{b} fixo, define a função geradora de momentos da distribuição Dirichlet de parâmetro \underline{b} .

Considerando $f(x) = x^a$ com $I = \mathbb{R}$ se $a \in \mathbb{N}$ e I um intervalo aberto de $\mathbb{R} - \{0\}$, se $a \in \mathbb{R} - \mathbb{N}$, a correspondente função F é a denominada função hipergeométrica R de Carlson denotada por

$$R_a(\underline{b}, \underline{z}) = E_{\underline{\theta} | \underline{b}} (\underline{z}' \underline{\theta})^a \quad , \quad \underline{z} \in I^k \quad (33)$$

Note-se, contudo, que o integral é bem definido quando \underline{z} é um vector de 0's e 1's - traduzindo um momento generalizado de $\underline{z}' \underline{\theta} | \underline{b} \rightsquigarrow \text{Be}(\underline{z}' \underline{b} (\mathbf{1}'_k - \underline{z}') \underline{b})$ - desde que $\underline{z}' \underline{b} + a > 0$.

Quando $a = n \in \mathbb{N}$, a respectiva função hipergeométrica é um polinómio homogêneo de grau n nas variáveis $\{z_i\}$ no mesmo sentido em que R_a é uma função homogênea de grau a . De facto, pelo teorema multinomial

$$R_n(\underline{b}, \underline{z}) = E_{\underline{\theta} | \underline{b}} (\underline{z}' \underline{\theta})^n = E_{\underline{\theta} | \underline{b}} \left\{ \sum_{\underline{m}; \underline{m} = n} \binom{n}{\underline{m}} \prod_{i=1}^k (z_i \theta_i)^{m_i} \right\}$$

onde

$$\underline{m} = (m_1, \dots, m_k)' \quad , \quad m_i = \mathbf{1}'_k \underline{m} \quad \text{e} \quad \binom{n}{\underline{m}} = \frac{n!}{\prod_{i=1}^k m_i!}$$

Como

$$E_{\underline{\theta} | \underline{b}} \left[\prod_{i=1}^k \theta_i^{m_i} \right] = \frac{B(\underline{b} + \underline{m})}{B(\underline{b})} = \frac{\prod_{i=1}^k (b_i, m_i)}{(b, \underline{m})} \quad ,$$

onde $b_i = \mathbf{1}'_k \underline{b}_k$, segue-se que

$$R_n(\underline{b}, \underline{z}) = \frac{n!}{(\underline{b}, n)} \sum_{\underline{m}: \underline{m} = n} \left[\prod_{i=1}^k (b_i, m_i) \right] \prod_{i=1}^k \frac{z_i^{m_i}}{m_i!} \quad (34)$$

o que traduz uma função polinomial dos $\{z_i\}$. A homogeneidade é de grau n , já que

$$\forall c \in \mathbb{R} - \{0\}, R_n(\underline{b}, c\underline{z}) = E_{\theta | \underline{b}} (c\underline{z}'\theta)^n = c^n R_n(\underline{b}, \underline{z})$$

É fácil verificar que o teorema binomial para $(\underline{z}'\theta)^n = [1 - (\underline{1}_k - \underline{z})'\theta]^n$ implica um teorema correspondente para $R_n(\underline{b}, \underline{z})$ expresso por

$$R_n(\underline{b}, \underline{z}) = \sum_{m=0}^n \frac{(-n, m)}{m!} R_m(\underline{b}, \underline{1}_k - \underline{z}) \quad (35)$$

Se alguns $\{z_i\}$ são nulos, o nº de termos da soma (34) diminui, conduzindo a que R_n se possa calcular à custa de um integral (33) com dimensionalidade reduzida. Com efeito, seja δ um subconjunto de $\{1, \dots, k\}$ tal que $z_i = 0, i \notin \delta$. Todos os termos do somatório em (34) são nulos à excepção daqueles para os quais $m_i = 0, i \notin \delta$ (estamos assumindo, como é vulgar, que $0^0 = 1$). Assim, (34) converte-se em

$$\begin{aligned} R_n(b_1, \dots, b_k; z_1, \dots, z_k) &= \frac{n!}{\left(\sum_{i=1}^k b_i, n \right)} \sum_{\substack{\{m_i: i \in \delta\} \\ \sum_{i \in \delta} m_i = n}} \left[\prod_{i \in \delta} (b_i, m_i) \right] \prod_{i \in \delta} \frac{z_i^{m_i}}{m_i!} \\ &= \frac{\left(\sum_{i \in \delta} b_i, n \right)}{\left(\sum_{i=1}^k b_i, n \right)} R_n(b_i, i \in \delta; z_i, i \in \delta) \end{aligned} \quad (36)$$

Se adicionalmente $z_i = 1, i \in \delta$, então, por (32)

$$R_n(b_1, \dots, b_k; \underline{z}) = \frac{\left(\sum_{i \in \delta} b_i, n \right)}{\left(\sum_{i=1}^k b_i, n \right)} = \frac{\Gamma\left(\sum_{i \in \delta} b_i + n\right) \Gamma\left(\sum_{i \notin \delta} b_i\right)}{\Gamma\left(\sum_{i \in \delta} b_i\right) \Gamma\left(\sum_{i \notin \delta} b_i + n\right)} \quad (37)$$

onde $\underline{z} = (z_i, i=1, \dots, k)$ tal que $z_i = 1, i \in \delta$ e $z_i = 0, i \notin \delta$.

Dickey (1983) visualizou $R_n(\underline{b}, \underline{z})$ como a função geradora de probabilidades da distribuição Dirichlet-Multinomial de parâmetros (n, \underline{b}) . Este resultado decorre da representação polinomial (34) a qual traduz que

$$R_n(\underline{b}, \underline{z}) = E_{\underline{m} | n, \underline{b}} \left[\prod_{i=1}^k z_i^{m_i} \right], \text{ com } \underline{m} | n, \underline{b} \rightsquigarrow DM_k(n, \underline{b})$$

Assim, os momentos de qualquer distribuição DM podem ser calculados da função hipergeométrica R_n .

As funções R_n desempenham um papel importante já que através delas se podem exprimir muitas funções F . No caso de $f(x)$ ser analítica em $I \ni I(z)$, a sua série de Taylor em torno de $y \in I$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(y)}{n!} (x-y)^n$$

converge uniformemente em qualquer subconjunto compacto de I , logo em $I(z)$. Consequentemente, a série

$$f(z' \theta) = \sum_{n=0}^{\infty} \frac{f^{(n)}(y)}{n!} (z' \theta - y)^n$$

para $z \in I^k$ fixo converge uniformemente em S_{k-1} . A integração termo a termo produz então

$$F(b, z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(y)}{n!} R_n(b, z - y 1_k) \quad (38)$$

A média Dirichlet de uma função $f: I \rightarrow R$, expandível numa série de potências, tem assim uma representação em série com os mesmos coeficientes mas cujas potências são substituídas pelos polinômios homogêneos R_n .

No caso de $f(x) = (1-x)^a$, $a \in R - N$, considerando $I = (-1, 1)$, $z = I^k$ e $y = 0$, obtemos da série binomial e de (34)

$$R_a(b, 1_k - z) = \sum_{n=0}^{\infty} \frac{(-a, n)}{n!} R_n(b, z) \quad (39)$$

$$= \sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \frac{(-a, m_{\cdot}) \prod_{i=1}^k (b_i, m_i)}{(b_{\cdot}, m_{\cdot})} \prod_{i=1}^k \frac{z_i^{m_i}}{m_i!} \quad (40)$$

Note-se que a representação polinomial (39) ainda é válida quando z é um vector de 0's e 1's uma vez que $0 < z' \theta < 1$, assegurando a expansão $(1 - z' \theta)^a = \sum_{n=0}^{\infty} [(-a, n)/n!] (z' \theta)^n$ e a convergência uniforme da série. Neste caso, a série simplifica-se por (37) para

$$R_a(b, 1_k - z) = \sum_{n=0}^{\infty} \frac{(-a, n) \left(\sum_{i \in \delta} b_i, n \right)}{\left(\sum_{i=1}^k b_i, n \right) n!} = {}_2F_1 \left(-a, \sum_{i \in \delta} b_i; \sum_{i=1}^k b_i; 1 \right) \quad (41)$$

onde δ indica o conjunto de índices i tal que $z_i = 1$.

A série hipergeométrica múltipla (40) representa, dependendo da dimensão k , várias funções hipergeométricas especiais. Por exemplo, vimos

já em C.2.1 que, para $k=2$, $R_a(b_1, b_2; 1-z, 1) = {}_2F_1(-a, b_1; b_1+b_2; z)$.

Até aqui, as funções hipergeométricas R (bem como todas as outras funções F) foram definidas com base na representação integral (33), o que exigia que $\underline{b} \in \mathbb{R}_+^k$. Todavia, a sua representação em série permite prolongá-las analiticamente a $\underline{b} \in \mathbb{R}^k$.

Por exemplo, de (34) obtém-se

$$(\underline{b}, n)R_n(\underline{b}, \underline{z}) = n! \sum_{\underline{m}: \underline{m} = n} \prod_{i=1}^k [(b_i, m_i) \frac{z_i^{m_i}}{m_i!}] \quad (42)$$

Como o polinômio em \underline{b} e \underline{z} no 2º membro tem sentido para $\underline{b} \in \mathbb{R}^k$, ele define o prolongamento analítico de $(\underline{b}, n)R_n(\underline{b}, \underline{z})$ a $(\underline{b}, \underline{z}) \in \mathbb{R}^{2k}$. Assim, desde que $\underline{b} \notin \{0, -1, -2, \dots, -n+1\}$ de modo a evitar $(\underline{b}, n) = 0$, a expressão (34) passa a definir $R_n(\underline{b}, \underline{z})$ para $(\underline{b}, \underline{z}) \in \mathbb{R}^{2k}$.

Consideremos o caso particular em que $b_i = 0$, $i \notin \delta$ onde δ é um subconjunto próprio de $\{1, \dots, k\}$. Como para $i \notin \delta$, (b_i, m_i) é igual a 1 se $m_i = 0$ e igual a 0 se $m_i \neq 0$, os únicos termos não nulos do somatório em (42) são aqueles para os quais $m_i = 0$, $i \notin \delta$. Assim de (42) resulta quando $(\underline{b}, n) \neq 0$

$$R_n(\underline{b}; \underline{z}) = R_n(b_i, i \in \delta; z_i, i \in \delta)$$

onde \underline{b} é tal que $b_i = 0$, $i \notin \delta$. Em suma, na expressão de $R_n(\underline{b}, \underline{z})$, os b_i 's nulos podem ser omitidos concomitantemente com os correspondentes z_i 's desde que pelo menos um b_i seja não nulo.

Em virtude do prolongamento de $R_n(\underline{b}, \underline{z})$ a expressão (35), por exemplo, mantém-se válida para qualquer $\underline{b} \in \mathbb{R}^k$ tal que $(\underline{b}, n) \neq 0$. Em particular, se os $\{b_i\}$ forem inteiros não positivos, fazendo $n = -b$, obtém-se

$$R_{-b}(\underline{b}, \underline{z}) = \sum_{m=0}^{-b} \frac{(b, m)}{m!} R_m(\underline{b}, \underline{1}_k - \underline{z}) \quad (44)$$

uma vez que $(b, -b) \neq 0$.

O desenvolvimento em série das funções F em termos dos polinômios R_n permite então prolongá-las igualmente por analiticidade. Em particular, o facto de

$$R_n(\underline{b}, \underline{z}) / \Gamma(b) = (\underline{b}, n)R_n(\underline{b}, \underline{z}) / \Gamma(b+n)$$

juntamente com a expansão em série (recorde-se (39))

$$R_a(\underline{b}, \underline{z}) = \sum_{n=0}^{\infty} [(-a, n)/n!] R_n(\underline{b}, \underline{1}_k - \underline{z}), \quad |1-z_i| < 1, \quad \forall i \quad (45)$$

permite considerar (45), aliado a (35), como o prolongamento analítico de $R_a(\underline{b}, \underline{z})$ em (33) a $\underline{b} \in \mathbb{R}^k$ tal que $\underline{b} \notin \{0, -1, -2, \dots\}$, desde que \underline{z} verifique a restrição mencionada - veja-se Carlson (1977, sec. 6.3).

Em virtude de (43) temos igualmente

$$R_a(\underline{b}, \underline{z}) = R_a(b_i, i\delta; z_i, i\delta), \quad |1-z_i| < 1, \quad \forall i \quad (46)$$

onde \underline{b} é tal que $b_i = 0$, $i \notin \delta$ e $b \notin \{0, -1, -2, \dots\}$.

Vai considerar-se agora a dedução de importantes relações que permitem calcular as funções R_a para valores restritos do parâmetro a .

Se $\underline{b} \in \mathbb{R}^k$ e $(\omega, \underline{z}) \in \mathbb{R}^{k+1}$ tal que $|\omega(1-z_i)| < 1, \forall i$, a expansão binomial de $[1-\omega(1-z_i)]^{-b_i}, i=1, \dots, k$ é

$$[1-\omega(1-z_i)]^{-b_i} = \sum_{m_i=0}^{\infty} \omega^{m_i} \binom{m_i}{b_i, m_i} (1-z_i)^{m_i} / (m_i!), \quad i=1, \dots, k$$

pelo que o produto das k séries se pode exprimir na série múltipla

$$\prod_{i=1}^k [1-\omega(1-z_i)]^{-b_i} = \sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \omega^{\sum_{i=1}^k m_i} \prod_{i=1}^k \left[\frac{\binom{m_i}{b_i, m_i}}{m_i!} (1-z_i)^{m_i} \right], \quad |\omega(1-z_i)| < 1 \quad \forall i$$

Dada a convergência absoluta desta série, podemos rearranjá-la em qualquer ordem e designadamente em

$$\begin{aligned} \prod_{i=1}^k [1-\omega(1-z_i)]^{-b_i} &= \sum_{n=0}^{\infty} \sum_{\{m_i\}: m_i = n} \omega^n \prod_{i=1}^k \left[\frac{\binom{m_i}{b_i, m_i}}{m_i!} (1-z_i)^{m_i} \right] \\ &= \sum_{n=0}^{\infty} \omega^n \frac{\binom{b, n}}{n!} R_n(\underline{b}, \underline{1}_k - \underline{z}), \quad b \notin \{0, -1, \dots, -n+1\} \end{aligned} \quad (47)$$

devido à expressão (42) e respectiva condição de validade.

Se adicionalmente $b \notin \{0, -1, -2, \dots\}$ e $|1-z_i| < 1, \forall i$, a expressão (45) traduz para $a = -b$, que

$$R_{-b}(\underline{b}, \underline{z}) = \sum_{n=0}^{\infty} \frac{\binom{b, n}}{n!} R_n(\underline{b}, \underline{1}_k - \underline{z}) \quad (48)$$

o que corresponde a (47) com $\omega = 1$. Assim

$$R_{-b}(\underline{b}, \underline{z}) = \prod_{i=1}^k z_i^{-b_i}, \quad \begin{aligned} &b \notin \{0, -1, -2, \dots\} \\ &|1-z_i| < 1, i=1, \dots, k \end{aligned} \quad (49)$$

Naturalmente que quando os $\{b_i\}$ são inteiros negativos, a expressão (49), cujo 1º membro corresponde então à soma (44), é válida sem as restrições supra-mencionadas.

A expressão (47) permite ainda calcular R_a , para uma determinada gama de valores de a , através de um integral simples. Para o efeito, seja \underline{z} tal que $|1-z_i| < 1, \forall i$. Por (45)

$$R_a(\underline{b}, \underline{z}) = \sum_{n=0}^{\infty} \frac{(b_., n)}{n!} R_n(\underline{b}, \underline{1}_k - \underline{z}) \frac{(-a, n)}{(b_., n)}$$

Se $-b_ < a < 0$, então $(-a, n)/(b_., n) = B(-a+n, b_+a)/B(-a, b_+a)$ é o momento de ordem n de uma variável ω , digamos, com distribuição $Be(\underline{c})$, onde $\underline{c} = (-a, b_+a)$, pelo que

$$R_a(\underline{b}, \underline{z}) = \sum_{n=0}^{\infty} \int_0^1 \omega^n \frac{(b_., n)}{n!} R_n(\underline{b}, \underline{1}_k - \underline{z}) d\lambda_{\underline{c}}(\omega)$$

Ora foi visto que a série (47) converge se $|\omega| < (|1-z_i|)^{-1}$, $\forall i$ e portanto converge uniformemente para $0 \leq \omega \leq 1$. Permutando então os símbolos de somatório e de integral e tendo em conta (47), obtemos a denominada identidade de Picard

$$R_a(\underline{b}, \underline{z}) = E_{\omega | \underline{c}} \left[\prod_{i=1}^k (\omega z_i + 1 - \omega)^{-b_i} \right], \quad \begin{array}{l} -b_ < a < 0 \\ |1-z_i| < 1, \forall i \end{array} \quad (50)$$

Note-se que de (50) podemos obter (49) fazendo $a \rightarrow -b_$, implicando que ω convirja para a distribuição degenerada no ponto 1. Carlson (1977, sec. 6.8) mostra ainda que o 2º membro de (50) constitui o prolongamento analítico de $R_a(\underline{b}, \underline{z})$, sob as restrições em \underline{b} e a citadas, a $\underline{z} \in \mathbb{R}_+^k$, com base no qual prova a validade de (49) sob a restrição $b_ \notin \{0, -1, -2, \dots\}$ e a condição menos restritiva de $\underline{z} \in \mathbb{R}_+^k$.

C.2.3. Funções hipergeométricas R de Carlson

Carlson (1971) generalizou as funções F através da consideração de uma forma bilinear em dois vectores Dirichlet. Mais concretamente, seja $\underline{Z} = (z_{ij})$ uma matriz $k \times \ell$ de colunas \underline{z}_j , $j=1, \dots, \ell$. Seja ainda $\underline{\theta} = (\theta_1, \dots, \theta_k)'$ e $\underline{\phi} = (\phi_1, \dots, \phi_\ell)'$ dois vectores aleatórios com valores nos simplex de dimensão $k-1$ (S_{k-1}) e $\ell-1$ ($S_{\ell-1}$), respectivamente, tal que $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$ e $\underline{\phi} | \underline{c} \rightsquigarrow D_\ell(\underline{c})$, onde $\underline{b} = (b_1, \dots, b_k)' \in \mathbb{R}_+^k$ e $\underline{c} = (c_1, \dots, c_\ell) \in \mathbb{R}_+^\ell$.

A forma bilinear, $\underline{\theta}' \underline{Z} \underline{\phi} = \sum_{i=1}^k \sum_{j=1}^{\ell} \theta_i z_{ij} \phi_j$ e as formas lineares $\underline{z}_j' \underline{\theta} = \sum_{i=1}^k \theta_i z_{ij}$, $j=1, \dots, \ell$ são ambas combinações convexas situadas no invólucro convexo de $\{z_{ij}, i=1, \dots, k; j=1, \dots, \ell\}$, denotado por $I(\underline{Z})$. Se f é uma função analítica de variável real definida num domínio $D \supset I(\underline{Z})$.

Carlson define a média Dirichlet dupla

$$F(\underline{b}, \underline{Z}, \underline{c}) = E_{\underline{\theta} | \underline{b}}^{(k-1)} E_{\underline{\phi} | \underline{c}}^{(\ell-1)} f(\underline{\theta}' \underline{Z} \underline{\phi}) \quad (51)$$

Note-se que

$$F(\underline{b}, \underline{Z}, \underline{c}) = f(d) \quad \text{se} \quad \underline{Z} = d \mathbf{1}_{k \times \ell}'$$

e quando $\ell=1$ ($\underline{Z} = \underline{z}_1$)

$$F(\underline{b}, \underline{z}_1, \underline{c}) = F(\underline{b}; \underline{z}_1) \quad (52)$$

evidenciando que F é uma generalização de F .

Várias propriedades de F implicam propriedades correspondentes de F (veja-se Carlson [1971, pg 422]).

Quando $f(x)$ é a função potência x^a , a transformada F converte-se na função hipergeométrica dupla denotada por

$$R_a(\underline{b}, \underline{Z}, \underline{c}) = E_{\underline{\theta} | \underline{b}}^{(k-1)} E_{\underline{\phi} | \underline{c}}^{(\ell-1)} (\underline{\theta}' \underline{Z} \underline{\phi})^a, \quad D = \mathbb{R}_+, \quad a \in \mathbb{R}_+ \quad (53)$$

Do facto de $\underline{\theta}' \underline{Z} \underline{\phi} = \underline{\phi}' \underline{Z}' \underline{\theta}$ segue-se a propriedade de simetria face à transposição

$$R_a(\underline{b}, \underline{Z}, \underline{c}) = R_a(\underline{c}, \underline{Z}', \underline{b})$$

Fazendo $\underline{x} = \underline{Z}' \underline{\theta}$ obtém-se de (53), tendo em conta (33),

$$R_a(\underline{b}, \underline{Z}, \underline{c}) = E_{\underline{\theta} | \underline{b}}^{(k-1)} [R_a(\underline{c}, \underline{x})] \quad (54)$$

Vimos em C.2.2. que se $a = -c_j = -\sum_{j=1}^{\ell} c_j$, a função $R_a(\underline{c}, \underline{x})$ pode ser prolongada analiticamente por (49) a $\mathbb{C} \mathbb{R}^{\ell}$ desde que $c_j \notin \{0, -1, -2, \dots\}$ e $0 < x_j < 2$, $j=1, \dots, \ell$. Esta restrição em $\underline{x} = \underline{Z}' \underline{\theta}$ é satisfeita se cada coluna de \underline{Z} é um vector de 0's e 1's, embora, como vimos em C.2.2., seja dispensável (assim como a restrição em c_j) quando os $\{c_j\}$ são inteiros negativos. Em suma

$$R_{-c}(\underline{b}, \underline{Z}, \underline{c}) = E_{\underline{\theta} | \underline{b}} \left[\prod_{j=1}^{\ell} (z_j' \underline{\theta})^{-c_j} \right], \quad \begin{array}{l} c_j \notin \{0, -1, -2, \dots\} \\ |1 - z_j' \underline{\theta}| < 1, \quad \forall j \end{array} \quad (55)$$

define o prolongamento analítico de R_{-c} a $\mathbb{C} \mathbb{R}^{\ell}$ tal que $c_j \notin \{0, -1, -2, \dots\}$.

Esta expressão permite verificar imediatamente que R_{-c} é não só homogênea de grau $-c$ em \underline{Z} (propriedade correspondente à homogeneidade de R_a em (53)) mas também homogênea de grau $-c_i$ na coluna i de \underline{Z} , quando $\{c_j\}$ são negativos

$$R_{-c}(\underline{b}; \underline{z}_1, \dots, \alpha \underline{z}_i, \dots, \underline{z}_\ell; \underline{c}) = \alpha^{-c_i} R_{-c}(\underline{b}, \underline{Z}, \underline{c}), \quad \alpha \in \mathbb{R}$$

Por outro lado, definindo $\underline{c}^* = (\underline{c}' d - c, \dots)'$ com d apenas restringido

por $d \in \{0, -1, -2, \dots\}$ e $\underline{z}^* = (\underline{z} \ 1_k)$, mostra-se facilmente por (55) que

$$R_{-c}(\underline{b}, \underline{z}, c) = R_{-d}(\underline{b}, \underline{z}^*, c^*) \quad (56)$$

o que significa que R_{-c} pode ser indexada por infinitos valores $-d$, Escolhendo $d=b$, obtemos uma forma não indexada ("bare form"), a qual possui propriedades adicionais e com base na qual se podem obter generalizações da identidade de Picard e exprimir várias funções hipergeométricas simples e duplas (vide Carlson [1971, secs. 4 e 5] e Dickey [1983, sec. 6]).

Note-se ainda que o anulamento de alguns elementos de \underline{c} possibilita a sua remoção, juntamente com as colunas correspondentes de \underline{z} , da expressão de R quer em (54) - atendendo a (46) - quer em (55). Por exemplo, se $\Delta_0 = \{j: c_j = 0\}$, então de (55)

$$R_{-c}(\underline{b}, \underline{z}, c) = R_{-c}(\underline{b}, \bar{\underline{z}}, \bar{\underline{c}}) \quad (57)$$

onde $\bar{\underline{c}} = (c_j, j \notin \Delta_0)'$ e $\bar{\underline{z}} = (z_j, j \notin \Delta_0)$

A função R_{-c} é expressável em forma fechada no importante caso especial em que \underline{z} é indicadora duma partição dos índices $\{1, \dots, k\}$ em ℓ partes, Quando os $\{c_j\}$ são inteiros negativos, a expressão (55) mostra, tendo em consideração que, para tal matriz, $\underline{z}'\theta | \underline{b} \rightsquigarrow D_\ell(\underline{z}'\underline{b})$, que R_{-c} é um momento misto Dirichlet. Concretamente, nesse caso teremos, com $\underline{\phi} = (\phi_1, \dots, \phi_\ell) = \underline{z}'\underline{\theta}$,

$$R_{-c}(\underline{b}, \underline{z}, c) = E_{\underline{\phi} | \underline{b}} \left[\prod_{j=1}^{\ell} \phi_j^{-c_j} \right] = \frac{B(\underline{z}'\underline{b} - \underline{c})}{B(\underline{z}'\underline{b})} \quad (58)$$

Esta expressão ainda é válida para $\underline{c} \in \mathbb{R}^k$, desde que $c_j \in \{0, -1, -2, \dots\}$ e $\underline{z}'\underline{b} - \underline{c} \in \mathbb{R}_+^\ell$.

A forma polinomial $R_{\underline{n}}(\underline{b}, \underline{z}, c)$ é ainda interpretada por Dickey (1983) como a função geradora de probabilidades da distribuição preditiva das frequências de uma tabela de contingência $k \times \ell$ sob o modelo de independência.

Para o verificarmos, seja $\underline{N} = \{n_{ij}\}$ o vector das $k\ell$ frequências, com $1'_{k\ell} \underline{N} = \underline{n}$, tal que $\underline{N} | \underline{n}; \underline{\theta}, \underline{\phi} \rightsquigarrow M_{k\ell}(\underline{n}, \underline{\phi} \otimes \underline{\theta})$, onde $\underline{\theta}$ e $\underline{\phi}$ são respectivamente as probabilidades marginais das linhas e das colunas, e \otimes é o produto de Kronecker à esquerda,

Admita-se como distribuição à priori

$$\begin{aligned} \underline{\theta} | \underline{b} &\rightsquigarrow D_k(\underline{b}) \\ \underline{\phi} | \underline{c} &\rightsquigarrow D_\ell(\underline{c}) \end{aligned} \quad \underline{\theta} \perp \underline{\phi} | \underline{b}, \underline{c}$$

Denotando $\underline{N}_{*} = \{n_{i.}\}$ e $\underline{N}_{.} = \{n_{.j}\}$ os vectores das frequências marginais, respectivamente, das linhas e das colunas, constata-se facilmente que

$$\begin{aligned} \theta | N \rightsquigarrow D_k(b+N_{*}) \\ \phi | N \rightsquigarrow D_\ell(c+N_{*}) \end{aligned} \quad \theta \prod \phi | N$$

A distribuição preditiva *à priori* é

$$f(N|n; \underline{b}, \underline{c}) = \binom{n}{N} E_{\theta|b} \left[\prod_{i=1}^k \theta_i^{n_i} \right] E_{\phi|c} \left[\prod_{j=1}^{\ell} \phi_j^{n_j} \right] = \binom{n}{N} \frac{B(\underline{b}+N_{*})}{B(\underline{b})} \frac{B(\underline{c}+N_{*})}{B(\underline{c})}$$

Fazendo $\underline{z}' = (z_{ij}, i=1, \dots, k; j=1, \dots, \ell)$, a função geradora de probabilidades da distribuição marginal de N é

$$\begin{aligned} E_{N|n; \underline{b}, \underline{c}} \left[\prod_{i,j} z_{ij}^{n_{ij}} \right] &= E_{(\theta, \phi) | \underline{b}, \underline{c}} E_{N|n; \theta, \phi} \left[\prod_{i,j} z_{ij}^{n_{ij}} \right] \\ &= E_{\theta|b} E_{\phi|c} \left\{ \sum_{N:1, \dots, k, \ell} \binom{n}{N} \prod_{i,j} (\theta_i z_{ij} \phi_j)^{n_{ij}} \right\} \\ &= E_{\theta|b} E_{\phi|c} \left[\sum_{i,j} \theta_i z_{ij} \phi_j \right]^n \end{aligned}$$

Considerando o vector \underline{z} organizado na matrix $k \times \ell$ $Z = (z_{ij})$, obtém-se então como pretendíamos

$$\begin{aligned} E_{N|n; \underline{b}, \underline{c}} \left[\prod_{i,j} z_{ij}^{n_{ij}} \right] &\equiv \sum_{N:1, \dots, k, \ell} \binom{n}{N} \frac{B(\underline{b}+N_{*})}{B(\underline{b})} \frac{B(\underline{c}+N_{*})}{B(\underline{c})} \prod_{i,j} z_{ij}^{n_{ij}} \\ &= E_{\theta|b} E_{\phi|c} (\theta' Z \phi)^n \equiv \mathcal{R}_n(\underline{b}, Z, \underline{c}) \end{aligned}$$

ficando igualmente definida a representação polinomial de \mathcal{R}_n . Note-se que a interpretação probabilística de \mathcal{R}_n referida em C.2,2, sai como caso particular ($\ell=1$) deste resultado.

A construção da função \mathcal{R} de argumento matricial deixa perceber o processo de criação de funções hipergeométricas múltiplas. A correspondente forma polinomial pode interpretar-se probabilisticamente à luz da teoria da inferência Bayesiana conjugada em tabelas de contingência multidimensionais sob o modelo de independência completa (para detalhes sobre estas funções vide Carlson [1971, sec. 7] e Dickey [1983, sec. 10]).

DISTRIBUIÇÕES DE DIRICHLET GENERALIZADAS

Com base na função hipergeométrica de argumento vectorial R, Dickey (1983) definiu uma generalização da distribuição $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$ caracterizada pela função densidade no simplex S_{k-1}

$$g(\underline{\theta} | \underline{b}, \underline{z}, c) = \frac{\prod_{i=1}^k \theta_i^{b_i-1}}{B(\underline{b})} \frac{(z' \underline{\theta})^{-c}}{R_{-c}(\underline{b}, \underline{z})} \quad , \quad (59)$$

$\underline{b} \in \mathbb{R}_+^k$; $c \in \mathbb{R}$; $\underline{z} \in \mathbb{R}_+^k$, onde $R_{-c}(\underline{b}, \underline{z}) = E_{\underline{\theta} | \underline{b}} (z' \underline{\theta})^{-c}$, sendo o valor esperado relativo à distribuição $\underline{\theta} | \underline{b} \rightsquigarrow D_k(\underline{b})$.

As funções densidade desta família de distribuições, que denotaremos por $\underline{\theta} | \underline{b}, \underline{z}, c \rightsquigarrow \mathcal{D}_k^1(\underline{b}, \underline{z}, -c)$ - note-se que $D_k(\underline{b}) \sim \mathcal{D}_k^1(\underline{b}, \underline{z}, 0) \forall \underline{z}$ - são homogêneas de grau 0 em \underline{z}

$$g(\underline{\theta} | \underline{b}, \lambda \underline{z}, c) = g(\underline{\theta} | \underline{b}, \underline{z}, c) \quad , \quad \forall \lambda \in \mathbb{R}_+$$

dado que $R_{-c}(\underline{b}, \lambda \underline{z}) = \lambda^{-c} R_{-c}(\underline{b}, \underline{z})$.

Os momentos de $\mathcal{D}_k^1(\underline{b}, \underline{z}, -c)$ são iguais a momentos Dirichlet multiplicados por um quociente de funções hipergeométricas. Com efeito, o momento geral de ordem (\underline{r}, s) , onde $\underline{r} = (r_1, \dots, r_k)'$ é tal que $\underline{b} + \underline{r} \in \mathbb{R}_+^k$, é

$$\begin{aligned} E_{\underline{\theta} | \underline{b}, \underline{z}, c} \left[\prod_{i=1}^k \theta_i^{r_i} (z' \underline{\theta})^s \right] &= \frac{B(\underline{b} + \underline{r})}{B(\underline{b}) R_{-c}(\underline{b}, \underline{z})} \int_{S_{k-1}} (z' \underline{\theta})^{-c+s} d\lambda_{(\underline{b} + \underline{r})}(\underline{\theta}) \\ &= \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{R_{-c+s}(\underline{b} + \underline{r}, \underline{z})}{R_{-c}(\underline{b}, \underline{z})} \end{aligned} \quad (60)$$

É fácil constatar que a família $\mathcal{D}_k^1(\underline{b}, \underline{z}, -c)$ é fechada relativamente à amostragem multinomial. Na verdade, se a verosimilhança da amostra é proporcional a $\prod_{i=1}^k \theta_i^{r_i}$, então por (60) a função preditiva à priori é proporcional (com a mesma constante de proporcionalidade da verosimilhança) a

$$E_{\underline{\theta} | \underline{b}, \underline{z}, c} \left[\prod_{i=1}^k \theta_i^{r_i} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{R_{-c}(\underline{b} + \underline{r}, \underline{z})}{R_{-c}(\underline{b}, \underline{z})}$$

pelo que a distribuição à posteriori é $\underline{\theta} | \underline{r} \rightsquigarrow \mathcal{D}_k^1(\underline{b} + \underline{r}, \underline{z}, -c)$.

Outra sub-família de \mathcal{D}_k^1 é aquela onde $c = \underline{b} = \underline{1}'_k \underline{b}$ e a respectiva classe de funções densidade pode exprimir-se, devido a (59) e (49), por

$$g(\underline{\theta} | \underline{b}, \underline{z}) = [B(\underline{b})]^{-1} \prod_{i=1}^k \theta_i^{b_i-1} (\underline{z}' \underline{\theta})^{-b} \cdot \prod_{i=1}^k z_i^{b_i}, \quad \underline{b}, \underline{z} \in \mathbb{R}_+^k \quad (61)$$

Como Dickey (1983) refere, esta família pode ser encarada como resultante da família $\phi | \underline{b} \rightsquigarrow D_k(\underline{b})$ através da transformação $\theta_i = z_i^{-1} \phi_i / \sum_{j=1}^k z_j^{-1} \phi_j$, $i=1, \dots, k$, o que por sua vez deixa claro que $D_k(\underline{b}) \sim \mathcal{D}_k^1(\underline{b}, \lambda \underline{1}_k, -\underline{b})$, $\forall \lambda > 0$.

Os seus momentos podem obter-se de (60), pelo que

$$E_{\underline{\theta} | \underline{b}, \underline{z}} \left[\prod_{i=1}^k \theta_i^{r_i} (\underline{z}' \underline{\theta})^s \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \prod_{i=1}^k z_i^{-r_i} R_{-\underline{b} + \underline{s}}(\underline{b} + \underline{r}, \underline{z}) \quad (62)$$

Quando $s = -\underline{1}'_k \underline{r} = -r$, (62) é definida em forma explícita, já que pela transformação acima, $\underline{z}' \underline{\theta} = \left(\sum_{j=1}^k z_j^{-1} \phi_j \right)^{-1}$, e assim,

$$E_{\underline{\theta} | \underline{b}, \underline{z}} \left[\prod_{i=1}^k \theta_i^{r_i} (\underline{z}' \underline{\theta})^s \right] = E_{\underline{\phi} | \underline{b}} \left[\prod_{i=1}^k z_i^{-r_i} \phi_i^{r_i} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \prod_{i=1}^k z_i^{-r_i} \quad (63)$$

A demonstração acima do fecho da família $\mathcal{D}_k^1(\underline{b}, \underline{z}, -c)$ relativamente à amostragem multinomial deixa antever que quando c está relacionado com \underline{b} , essa propriedade deixa de se verificar. De facto, como \underline{b} é actualizado para $\underline{b} + \underline{r}$, o parâmetro c teria que sofrer uma transformação correspondente para que a referida família fosse conjugada da amostragem multinomial, o que não acontece, como vimos atrás. Logo, a família $\mathcal{D}_k^1(\underline{b}, \underline{z}, -\underline{b})$, em particular, não partilha dessa propriedade de fecho, contrariamente às famílias Dirichlet e $\mathcal{D}_k^1(\underline{b}, \underline{z}, -c)$ com c distinto de \underline{b} .

O facto da expressão (55) representar uma generalização de $R_{-c}(\underline{b}, \underline{z})$ levou Dickey (1983) a definir uma extensão da família $\mathcal{D}_k^1(\underline{b}, \underline{z}, c)$ caracterizada pela família de funções densidade em S_{k-1}

$$g(\underline{\theta} | \underline{b}, \underline{Z}, c) = \frac{\prod_{i=1}^k \theta_i^{b_i-1}}{B(\underline{b})} \frac{\prod_{j=1}^{\ell} (z_j' \underline{\theta})^{-c_j}}{R_{-c}(\underline{b}, \underline{Z}, c)} \quad (64)$$

onde $\underline{Z} = (z_1 \dots z_{\ell})$ é uma matriz $k \times \ell$ de elementos positivos e $\underline{c} \in \mathbb{R}^{\ell}$ com $c_j \notin \{0, -1, -2, \dots\}$ - esta última restrição em \underline{c} desaparece se os $\{c_j\}$ são inteiros não positivos.

Esta família de distribuições será denotada por $\mathcal{D}_k^\ell(\underline{b}, \underline{Z}, -\underline{c})$ e, tal como sucede para a sub-família $\mathcal{D}_k^1(\underline{b}, \underline{z}, -\underline{c})$, as correspondentes funções densidade são homogêneas de grau 0 em cada vector \underline{z}_j de \underline{Z} , i.e.

$$g(\underline{\theta} | \underline{b}; \lambda_1 \underline{z}_1, \dots, \lambda_\ell \underline{z}_\ell; \underline{c}) = g(\underline{\theta} | \underline{b}; \underline{z}_1, \dots, \underline{z}_\ell; \underline{c}), \quad \lambda_j > 0, \quad \forall j=1, \dots, \ell$$

$$\text{em virtude de } \mathcal{R}_{-\underline{c}}(\underline{b}; \lambda_1 \underline{z}_1, \dots, \lambda_\ell \underline{z}_\ell; \underline{c}) = \prod_{j=1}^{\ell} \lambda_j^{-\underline{c}_j} \mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c}).$$

Os momentos da distribuição $\underline{\theta} | \underline{b}, \underline{Z}, \underline{c} \rightsquigarrow \mathcal{D}_k^\ell(\underline{b}, \underline{Z}, -\underline{c})$ são analogamente proporcionais a razões de funções \mathcal{R} . Na realidade, seja $\underline{Y} = (\underline{y}_1 \dots \underline{y}_p)$ a matriz $k \times p$ tal que $\underline{y}_m \in \mathbb{R}_+^k$, $m = 1, \dots, p$, e $\underline{v} = (v_1, \dots, v_p)'$ e $\underline{s} = (s_1, \dots, s_\ell)'$ vectores de \mathbb{R}^p e \mathbb{R}^ℓ , respectivamente. Considere-se

$$\underline{z}^* = (\underline{z} \ \underline{y}) = (\omega_1, \dots, \omega_{\ell+p}), \quad \omega_j = \begin{cases} \underline{z}_j, & j=1, \dots, \ell \\ \underline{y}_{j-\ell}, & j=\ell+1, \dots, \ell+p \end{cases}$$

$$\underline{a} = (a_1, \dots, a_{\ell+p})', \quad a_j = \begin{cases} \underline{c}_j - s_j, & j=1, \dots, \ell \\ -v_{j-\ell}, & j=\ell+1, \dots, \ell+p \end{cases}$$

Então, com \underline{r} encarado como em (60),

$$E_{\underline{\theta} | \underline{b}, \underline{Z}, \underline{c}} \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{j=1}^{\ell} (\underline{z}_j' \cdot \underline{\theta})^{s_j} \prod_{m=1}^p (\underline{y}_m' \cdot \underline{\theta})^{-v_m} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b}) \mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})} E_{\underline{\theta} | \underline{b} + \underline{r}} \left[\prod_{j=1}^{\ell+p} (\omega_j' \cdot \underline{\theta})^{a_j} \right] \quad (65)$$

$$= \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{\mathcal{R}_{-\underline{a}}(\underline{b} + \underline{r}, \underline{z}^*, \underline{a})}{\mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})}$$

uma vez admitido que $\underline{a}_j = \underline{c}_j - s_j - v_j = \underline{c}_j - \underline{1}'_j \underline{s} - \underline{1}'_j \underline{v} \notin \{0, -1, -2, \dots\}$ - esta restrição adicional é omitida se \underline{c} é um vector inteiro negativo e \underline{s} e \underline{v} vectores inteiros positivos.

Desta expressão para os momentos generalizados de $\mathcal{D}_k^\ell(\underline{b}, \underline{Z}, -\underline{c})$ obtêm-se como casos particulares

$$E_{\underline{\theta} | \underline{b}, \underline{Z}, \underline{c}} \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{j=1}^{\ell} (\underline{z}_j' \cdot \underline{\theta})^{s_j} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{\mathcal{R}_{-\underline{c} + \underline{s}}(\underline{b} + \underline{r}, \underline{Z}, \underline{c} - \underline{s})}{\mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})} \quad (65a)$$

$$E_{\underline{\theta} | \underline{b}, \underline{Z}, \underline{c}} \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{m=1}^p (\underline{y}_m' \cdot \underline{\theta})^{v_m} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{\mathcal{R}_{-\underline{c} + \underline{v}}(\underline{b} + \underline{r}, \underline{z}^*, (\underline{c}' - \underline{v}')')}{\mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})} \quad (65b)$$

$$E_{\underline{\theta} | \underline{b}, \underline{Z}, \underline{c}} \left[\prod_{i=1}^k \theta_i^{r_i} \right] = \frac{B(\underline{b} + \underline{r})}{B(\underline{b})} \frac{\mathcal{R}_{-\underline{c}}(\underline{b} + \underline{r}, \underline{Z}, \underline{c})}{\mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})} \quad (65c)$$

Esta última fórmula, com os $\{r_i\}$ traduzindo as frequências amostrais num esquema amostral bernoulliano multivariado, serve-nos para mostrar que a distribuição à posteriori, $D_k^{\ell}(b+r, Z, -c)$, ainda é um membro da família $D_k^{\ell}(b, Z, -c)$ de distribuições à priori. De novo, se nesta família considerarmos c relacionado com b , por exemplo, $c_i = b_i$ (caso em que a função R em (64) está na forma não indexada), a classe correspondente de distribuições já não é fechada relativamente ao referido esquema de amostragem.

Quando Z é indicadora de uma partição, o momento generalizado definido pela fórmula (65a) é expressável, devido a (58), na forma fechada

$$E_{\theta|b, Z, c} \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{j=1}^{\ell} (z_j' \theta)^{s_j} \right] = \frac{B(b+r)}{B(b)} \frac{B[Z'(b+r)-(c-s)]}{B[Z'(b+r)]} \frac{B(Z'b)}{B(Z'b-c)} \quad (66)$$

Designando $\{\delta_1, \dots, \delta_{\ell}\}$ a partição de $\{1, \dots, k\}$ indicada por Z e sendo $\{b_{(\delta_j)}\}$ e $\{r_{(\delta_j)}\}$ definidos como indicado em C.1, a expressão (66) pode ainda ser posta na forma

$$E_{\theta|b, Z, c} \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{j=1}^{\ell} (z_j' \theta)^{s_j} \right] = \frac{B[(Z'b-c)+(Z'r+s)]}{B(Z'b-c)} \prod_{j=1}^{\ell} \frac{B(b_{(\delta_j)}+r_{(\delta_j)})}{B(b_{(\delta_j)})} \quad (67)$$

Neste contexto, a distribuição $\theta|b, Z, c \rightsquigarrow D_k^{\ell}(b, Z, -c)$ admite uma caracterização similar à da distribuição Dirichlet consubstanciada no teorema C.1.1., como consta no seguinte teorema:

Teorema C.3.1.: Seja Z a matriz $k \times \ell$ indicadora de uma partição $\{\delta_1, \dots, \delta_{\ell}\}$ dos índices $\{1, \dots, k\}$ e denote-se $\phi = (\phi_1, \dots, \phi_{\ell})' = Z'\theta$ e $\rho_{(j)} = \theta_{(\delta_j)}/\phi_j$, $j=1, \dots, \ell$, onde $\theta_{(\delta_j)}$ é o vector dos elementos do conjunto $\{\theta_i: i \in \delta_j\}$ de cardinalidade d_j . Então

$$\theta|b, Z, c \rightsquigarrow D_k^{\ell}(b, Z, -c) \quad \text{sse} \quad \begin{cases} \phi|b, Z, c \rightsquigarrow D_{\ell}(Z'b-c) \\ \rho_{(j)}|b \rightsquigarrow D_{d_j}(b_{(\delta_j)}), \quad j=1, \dots, \ell \\ \phi \perp \perp \rho_{(1)} \perp \perp \dots \perp \perp \rho_{(\ell)} | b, Z, c \end{cases} \quad (68)$$

Dem. Supondo que $\theta|b, Z, c \rightsquigarrow D_k^{\ell}(b, Z, -c)$, o recurso a (66) permite-nos obter

$$E_{\theta|b, Z, c} \left[\prod_{j=1}^{\ell} (z_j' \theta)^{s_j} \right] = \frac{B(Z'b-c+s)}{B(Z'b-c)}$$

Mas, por (2), o 2º membro desta relação traduz o momento misto de ordens s_1, \dots, s_{ℓ} do vector ϕ quando distribuído segundo a distribuição $D_{\ell}(Z'b-c)$. Como ϕ é um vector limitado, a sua distribuição é unicamente determinada pelos seus momentos e, por conseguinte, a sua função densidade é dada por

$$g_0(\underline{\phi}|\underline{b}, \underline{z}, \underline{c}) = \frac{1}{B(\underline{z}'\underline{b}-\underline{c})} \prod_{j=1}^{\ell} \phi_j z_j^{b-c_j-1}, \quad \underline{\phi} \in S_{\ell-1}$$

Deste resultado e de (64), tendo em consideração (58), podemos então concluir que a função densidade condicional de $\underline{\theta}$ dado $\underline{\phi}$ é

$$g^*(\underline{\theta}|\underline{\phi}; \underline{b}) = \prod_{j=1}^{\ell} \left\{ \left(\frac{1}{B(\underline{b}(\delta_j))} \prod_{i \in \delta_j} (\theta_i / \phi_j)^{b_i-1} \right) \phi_j^{-(d_j-1)} \right\}$$

resultado exactamente análogo a (9). Assim, a demonstração do teorema C.1.1 permite-nos provar as implicações restantes da parte necessária.

A demonstração da proposição recíproca segue igualmente os passos usados no teorema C.1.1, tendo em vista que (64) admite a factorização

$$g(\underline{\theta}|\underline{b}, \underline{z}, \underline{c}) = g_0(\underline{z}'\underline{\theta}|\underline{b}, \underline{z}, \underline{c}) g^*(\underline{\theta}|\underline{z}'\underline{\theta}; \underline{b}) \quad \Delta$$

NOTA 1: Dada esta caracterização, o teorema C.1.1. surge como seu corolário, fazendo $\underline{c} = \underline{0}$. Δ

Uma outra situação especial ocorre quando a matriz $k \times \ell$ \underline{Z} é indicadora de partições encaixadas em dois níveis (vide secção 3.4 do texto). Nesse caso, \underline{Z} pode ser particionada na sub-matriz $\underline{Z}_1 = (z_{1h}, h=1, \dots, \ell_1)$, indicadora de uma partição $\{\delta_h^1, h=1, \dots, \ell_1\}$ dos índices $\{1, \dots, k\}$, e numa sub-matriz \underline{Z}_2 de dimensão $k \times \ell_2$ ($\ell_2 = \ell - \ell_1$). Esta, por sua vez, pode ser particionada nas sub-matrizes $\underline{Z}_{2h} = (z_{2hm}, m=1, \dots, j_h)$, $h=1, \dots, \ell_1$, onde cada \underline{Z}_{2h} indica uma partição $\{\delta_{hm}^2, m=1, \dots, j_h\}$ do conjunto δ_h^1 , pelo que $\sum_{m=1}^{j_h} z_{2hm} = z_{1h}$. Neste contexto mais geral, a distribuição $\underline{\theta}|\underline{b}, \underline{z}, \underline{c} \rightsquigarrow \mathcal{D}_k^\ell(\underline{b}, \underline{z}, -\underline{c})$ admite ainda uma caracterização, conforme se descreve em seguida:

Teorema C.3.2: Seja $\underline{Z} = (\underline{Z}_1 \underline{Z}_2)$ a matriz $k \times \ell$ indicadora de partições encaixadas em dois níveis, no sentido acima descrito, e denote-se

$$\underline{\phi}_1 = (\phi_{1h}, h=1, \dots, \ell_1)' = \underline{z}_1' \underline{\theta} \quad \text{e} \quad \sigma_{(1h)} = (\sigma_{(1h)}^1, i \in \delta_h^1)' = \underline{\theta}_{(\delta_h^1)} / \phi_{1h}, h=1, \dots, \ell_1,$$

onde $\underline{\theta}_{(\delta_h^1)} = (\theta_i, i \in \delta_h^1)'$ de dimensão $d_h^1 \times 1$ e $\{\delta_h^1, h=1, \dots, \ell_1\}$ é uma partição do conjunto $\{1, 2, \dots, k\}$. Seja ainda, em correspondência com o particionamento de \underline{Z} e de \underline{Z}_2 , $\underline{c} = (\underline{c}_1' \underline{c}_2')$, com $\underline{c}_1' = (c_{1h}, h=1, \dots, \ell_1)$,

$$\underline{c}_2' = (c_{21}' \dots c_{2h}' \dots c_{2\ell_2}') \quad \text{e} \quad c_{2h}' = (c_{2hm}, m=1, \dots, j_h), h=1, \dots, \ell_1 \quad \text{e defina-se}$$

$$\bar{c}_2 = (c_{2h}, h=1, \dots, \ell_1)', \quad \text{onde} \quad c_{2h} = \sum_{m=1}^{j_h} c_{2hm}. \quad \text{Então}$$

$$\theta | \underline{b}, \underline{Z}, \underline{c} \rightsquigarrow \mathcal{D}_k^\ell(\underline{b}, [\underline{Z}_1 \underline{Z}_2], \begin{pmatrix} -\underline{c}_1 \\ -\underline{c}_2 \end{pmatrix}) \text{ sse } \begin{cases} \phi_1 | \underline{b}, \underline{Z}_1, \underline{c} \rightsquigarrow D_{\ell_1} (Z_1' \underline{b} - \underline{c}_1 - \underline{c}_2) \\ \sigma_{(1h)} | \underline{b}, \underline{Z}_2, \underline{c}_2 \rightsquigarrow \mathcal{D}_{d_h^1}^{j_h}(\underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, -\underline{c}_{2h}), h=1, \dots, \ell_1 \\ \phi_1 \parallel \sigma_{(11)} \parallel \dots \parallel \sigma_{(1\ell_1)} | \underline{b}, \underline{Z}, \underline{c} \end{cases} \quad (69)$$

onde $\bar{z}_{2h} = (\bar{z}_{2hm}, m=1, \dots, j_h)$ é a matriz $d_h^1 \times j_h$ obtida de Z_{2h} por omissão das linhas nulas correspondentes a $i\phi_h^1$, e $\underline{b}_{(\delta_h^1)}$ é definido analogamente a $\theta_{(\delta_h^1)}$.

Dem. Suponhamos primeiramente que θ tem a distribuição \mathcal{D}_k^ℓ mencionada, cuja função densidade se pode exprimir, em virtude do particionamento referido de \underline{Z} , por

$$g(\theta | \underline{b}, \underline{Z}, \underline{c}) = \frac{\prod_{i=1}^k \theta_i^{b_i-1} \prod_{h=1}^{\ell_1} (z_{1h}' \theta)^{-c_{1h}} \prod_{h=1}^{\ell_1} \prod_{m=1}^{j_h} (z_{2hm}' \theta)^{-c_{2hm}}}{B(\underline{b})} \mathcal{R}_{-\underline{c}}(\underline{b}, [\underline{Z}_1 \underline{Z}_2], \begin{pmatrix} \underline{c}_1 \\ \underline{c}_2 \end{pmatrix})$$

Por (65a) temos com $\underline{s} = (s_1, \dots, s_{\ell_1})'$

$$E_{\theta | \underline{b}, \underline{Z}, \underline{c}} \left[\prod_{h=1}^{\ell_1} (z_{1h}' \theta)^{s_h} \right] = \frac{\mathcal{R}_{-\underline{c} + \underline{s}}(\underline{b}, \underline{Z}, \underline{c} - \underline{s}_*)}{\mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c})}$$

onde \underline{s}_* é o vector \underline{s} acrescentado com $\ell - \ell_1$ zeros.

Ora, o produto dos ℓ factores na densidade de θ pode ainda escrever-se como

$$\prod_{h=1}^{\ell_1} (z_{1h}' \theta)^{-c_{1h}} \prod_{h=1}^{\ell_1} \left\{ \prod_{m=1}^{j_h} \left(\frac{z_{2hm}' \theta}{z_{1h}' \theta} \right)^{-c_{2hm}} \right\} = \prod_{h=1}^{\ell_1} \phi_{1h}^{-c_{1h}} \prod_{h=1}^{\ell_1} \left\{ \prod_{m=1}^{j_h} (\bar{z}_{2hm}' \sigma_{(1h)})^{-c_{2hm}} \right\}$$

Como pelo teorema C.1.1.

$$\theta | \underline{b} \rightsquigarrow D_k(\underline{b}) \Rightarrow \begin{cases} \phi_1 | \underline{b} \rightsquigarrow D_{\ell_1} (Z_1' \underline{b}) \\ \sigma_{(1h)} | \underline{b} \rightsquigarrow D_{d_h^1}(\underline{b}_{(\delta_h^1)}), h=1, \dots, \ell_1 \\ \phi_1 \parallel \sigma_{(11)} \parallel \dots \parallel \sigma_{(1\ell_1)} | \underline{b} \end{cases}$$

segue-se que

$$\begin{aligned} \mathcal{R}_{-\underline{c}}(\underline{b}, \underline{Z}, \underline{c}) &= E_{\phi_1 | \underline{b}} \left[\prod_{h=1}^{\ell_1} \phi_{1h}^{-c_{1h}} \prod_{h=1}^{\ell_1} \left\{ E_{\sigma_{(1h)} | \underline{b}} \left[\prod_{m=1}^{j_h} (\bar{z}_{2hm}' \sigma_{(1h)})^{-c_{2hm}} \right] \right\} \right] \\ &= [B(Z_1' \underline{b} - \underline{c}_1 - \underline{c}_2) / B(Z_1' \underline{b})] \prod_{h=1}^{\ell_1} \mathcal{R}_{-\underline{c}_{2h}}(\underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, \underline{c}_{2h}) \end{aligned}$$

e analogamente

$$R_{-c,+s}(\underline{b}, \underline{Z}, \underline{c}-\underline{s}_*) = \frac{B(Z'_1 \underline{b}-\underline{c}_1-\bar{c}_2+s)}{B(Z'_1 \underline{b})} \prod_{h=1}^{\ell_1} R_{-c_{2h}}(\underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, c_{2h})$$

Assim

$$E_{\theta} | \underline{b}, \underline{Z}, \underline{c} \left[\prod_{h=1}^{\ell_1} (z'_{1h} \theta)^{s_h} \right] = \frac{B(Z'_1 \underline{b}-\underline{c}_1-\bar{c}_2+s)}{B(Z'_1 \underline{b}-\underline{c}_1-\bar{c}_2)}$$

Como o 2º membro reflecte o momento misto de ordens s_1, \dots, s_{ℓ_1} do vector limitado ϕ_1 quando distribucionalmente é $D_{\ell_1}(Z'_1 \underline{b}-\underline{c}_1-\bar{c}_2)$, fica provado que a distribuição marginal de ϕ_1 é

$$g_0(\phi_1 | \underline{b}, \underline{Z}_1, \underline{c}) = \frac{\prod_{h=1}^{\ell_1} z'_{1h} \underline{b}-c_{1h}-c_{2h}^{-1}}{B(Z'_1 \underline{b}-\underline{c}_1-\bar{c}_2)}, \quad \phi_1 \in S_{\ell_1-1}$$

Consequentemente, a distribuição condicional de θ dado ϕ_1 fica definida por

$$g^*(\theta | \phi_1; \underline{b}, \underline{Z}_2, \underline{c}_2) = \prod_{h=1}^{\ell_1} \left\{ \frac{\prod_{i=1}^{b_i-1} (\theta_i / \phi_{1h})^{i-1}}{B(\underline{b}_{(\delta_h^1)})} \cdot \frac{\prod_{m=1}^{j_h} \left[\frac{\bar{z}'_{2hm} \theta_{(\delta_h^1)}}{\phi_{1h}} \right]^{-c_{2hm}}}{R_{-c_{2h}}(\underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, c_{2h})} \right\} \phi_{1h}^{-(d_h^1-1)}$$

A repetição, de novo, dos argumentos usados na demonstração do teorema C.1.1 permite-nos provar as restantes implicações do enunciado. Δ

NOTA 2: Como \bar{z}_{2h} é indicadora de uma partição de δ_h^1 , $h=1, \dots, \ell_1$, cada um dos vectores $\sigma_{(1h)}$ está nas condições do teorema C.3.1. Assim, fazendo

$$\Gamma_{(2h)} \equiv (\Gamma_{(2h)}^m, m=1, \dots, j_h)' = Z'_{2h}(\theta / \phi_{1h}) = \bar{z}'_{2h} \sigma_{(1h)}, \quad h=1, \dots, \ell_1$$

$$\rho_{(2hm)} \equiv (\rho_{(2hm)}^i, i \in \delta_{hm}^2)' = (\sigma_{(1h)}^i, i \in \delta_{hm}^2)' / \Gamma_{(2h)}^m = \theta_{(\delta_{hm}^2)} / \bar{z}'_{2hm} \theta, \quad m=1, \dots, j_h$$

e sendo d_{hm}^2 a dimensionalidade de δ_{hm}^2 , temos por (68) que para cada $h=1, \dots, \ell_1$

$$\sigma_{(1h)} | \underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, c_{2h} \rightsquigarrow D_{d_h^1}(\underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, -c_{2h}) \text{ sse } \begin{cases} \Gamma_{(2h)} | \underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, c_{2h} \rightsquigarrow D_{j_h}(\bar{z}'_{2h} \underline{b}_{(\delta_h^1)} - c_{2h}) \\ \rho_{(2hm)} | \underline{b}_{(\delta_h^1)} \rightsquigarrow D_{d_{hm}^2}(\underline{b}_{(\delta_{hm}^2)}), m=1, \dots, j_h \\ \Gamma_{(2h)} \parallel \rho_{(2h1)} \parallel \dots \parallel \rho_{(2hj_h)} | \underline{b}_{(\delta_h^1)}, \bar{z}_{2h}, c_{2h} \end{cases} \quad (7)$$

Assim, quando Z é indicadora de duas partições encaixadas, a conjunção dos resultados dos teoremas C.3.2. e C.3.1 exprime que a adopção para θ da distribuição $D_k^{\ell}(\underline{b}, [Z_1 Z_2], (\underline{c}_1, \underline{c}_2))$ é equivalente a considerar que $\phi_1, \{\Gamma_{(2h)}, h=1, \dots, \ell_1\}$ e

$\{\rho_{(2hm)}, m=1, \dots, j_h; h=1, \dots, \ell_1\}$ são independentemente distribuídos segundo as distribuições de Dirichlet acima referidas.

Observe-se ainda que o teorema C.3.1 pode ser visto como um corolário do teorema C.3.2, considerando $\ell_1 = \ell$, $Z_1 = Z$ e $c_2 = 0$. Δ

NOTA 3: Fazendo $c = 0$ obtemos a correspondente caracterização da distribuição de Dirichlet

$$\theta | \underline{b} \rightsquigarrow D_k(\underline{b}) \text{ sse } \left\{ \begin{array}{l} \phi_1 | \underline{b} \rightsquigarrow D_{\ell_1}(Z_1' \underline{b}) \\ \Gamma_{(2h)} | \underline{b} \rightsquigarrow D_{j_h}(\bar{Z}'_{2h} \underline{b} (\delta_h^1)) , h=1, \dots, \ell_1 \\ \rho_{(2hm)} | \underline{b} \rightsquigarrow D_{d_{hm}^2}(\underline{b} (\delta_{hm}^2)) , m=1, \dots, j_h; h=1, \dots, \ell_1 \\ \phi_1 , \{\Gamma_{(2h)}\} , \{\rho_{(2hm)}\} \text{ independentes} \end{array} \right. \quad (71)$$

Δ

A caracterização da distribuição \mathcal{D}_k^ℓ enunciada no teorema C.3.2 e na nota 2 subsequente permite exprimir os respectivos momentos definidos em (65a) em forma fechada. Com efeito, atendendo ao resultado distribucional para $\{\Gamma_{(2h)}, h=1, \dots, \ell_1\}$ segue-se que

$$\begin{aligned} R_{-c}(\underline{b}, Z, c) &= \frac{B(Z_1' \underline{b} - \underline{c}_1 - \bar{c}_2)}{B(Z_1' \underline{b})} \prod_{h=1}^{\ell_1} R_{-c_{2h}}(\underline{b} (\delta_h^1), \bar{Z}_{2h}, c_{2h}) \\ &= \frac{B(Z_1' \underline{b} - \underline{c}_1 - \bar{c}_2)}{B(Z_1' \underline{b})} \prod_{h=1}^{\ell_1} \frac{B(\bar{Z}'_{2h} \underline{b} (\delta_h^1) - c_{2h})}{B(\bar{Z}'_{2h} \underline{b} (\delta_h^1))} \end{aligned}$$

Sendo \underline{s} particionado como \underline{c} , isto é, $\underline{s} = (s_1' s_2')'$, com $s_1' = (s_{1h}, h=1, \dots, \ell_1)$, $s_2' = (s_{21} \dots s_{2h} \dots s_{2j_h} \dots s_{2\ell_1})$ e $s_{2h}' = (s_{2hm}, m=1, \dots, j_h)$, $h=1, \dots, \ell_1$, e definindo $\bar{s}_2 = (s_{2h}, h=1, \dots, \ell_1)'$ onde $s_{2h} = \sum_{m=1}^{j_h} s_{2hm}$, $\underline{b}^* = \underline{b} + \underline{r}$ e $\underline{b}^* (\delta_h^1) = (\underline{b}_i + r_i, i \in \delta_h^1)'$, obtemos por raciocínio análogo

$$R_{-c + s}(\underline{b}^*, Z, c - \underline{s}) = \frac{B(Z_1' \underline{b}^* - \underline{c}_1 - \bar{c}_2 + s_1 + \bar{s}_2)}{B(Z_1' \underline{b}^*)} \prod_{h=1}^{\ell_1} \frac{B(\bar{Z}'_{2h} \underline{b}^* (\delta_h^1) - c_{2h} + s_{2h})}{B(\bar{Z}'_{2h} \underline{b}^* (\delta_h^1))}$$

Resumindo, quando $Z = (Z_1' Z_2)$ é indicadora de partições encaixadas em dois níveis temos

$$\begin{aligned} E_{\theta} | \underline{b}, Z, c \left[\prod_{i=1}^k \theta_i^{r_i} \prod_{h=1}^{\ell_1} (z_{1h}' \theta)^{s_{1h}} \prod_{h=1}^{\ell_1} \prod_{m=1}^{j_h} (z_{2hm}' \theta)^{s_{2hm}} \right] &= \frac{B(\underline{b}^*)}{B(\underline{b})} \times \frac{B(Z_1' \underline{b})}{B(Z_1' \underline{b} - \underline{c}_1 - \bar{c}_2)} \times \\ \times \frac{B(Z_1' \underline{b}^* - \underline{c}_1 - \bar{c}_2 + s_1 + \bar{s}_2)}{B(Z_1' \underline{b}^*)} \prod_{h=1}^{\ell_1} \left\{ \frac{B(\bar{Z}'_{2h} \underline{b} (\delta_h^1))}{B(\bar{Z}'_{2h} \underline{b} (\delta_h^1) - c_{2h})} \frac{B(\bar{Z}'_{2h} \underline{b}^* (\delta_h^1) - c_{2h} + s_{2h})}{B(\bar{Z}'_{2h} \underline{b}^* (\delta_h^1))} \right\} & \quad (72) \end{aligned}$$

A expressão explícita deste momento generalizado quando $\underline{\theta} | \underline{b} \rightsquigarrow \mathcal{D}_k(\underline{b})$ obtém-se de (72) fazendo $\underline{c} = \underline{0}$.

Por iteração dos teoremas C.3.2. e C.3.1. podemos obter uma caracterização similar da distribuição Dirichlet generalizada quando \underline{Z} é indicadora de partições encaixadas em T níveis ($T > 2$) e a decorrente expressão em forma fechada para os seus momentos.

NOTA 4: os resultados distribucionais supra-mencionados não exigem que a matriz \underline{Z} se particione em sub-matrizes indicando, cada uma delas, uma partição de $\{1, \dots, k\}$. É apenas necessário que exista uma nessas condições para a formação do vector $\underline{\phi}_1$ (se houver várias, toma-se obviamente a partição mais grossa). Por exemplo, no caso da estrutura encaixada em 2 níveis, a sub-matriz \underline{Z}_2 não precisa ser indicadora de uma partição, como se considerou acima ao defini-la em ℓ_1 sub-matrizes \underline{Z}_{2h} . Se \underline{Z}_2 revelar que alguns δ_h^1 não estão particionados, podemos definir uma partição arbitrária para eles, a ser indicada pelas sub-matrizes \underline{Z}_{2h}^* , e com estas ampliar \underline{Z}_2 de modo a obter a matriz \underline{Z}_2^* . Definindo correspondentemente \underline{c}_2^* como o vector \underline{c}_2 ampliado com os sub-vectores $\underline{c}_{2h}^* = \underline{0}$, o facto de

$$\mathcal{D}_k(\underline{b}, [\underline{Z}_1 \underline{Z}_2], \begin{pmatrix} -\underline{c}_1 \\ -\underline{c}_2 \end{pmatrix}) \sim \mathcal{D}_k(\underline{b}, [\underline{Z}_1 \underline{Z}_2^*], \begin{pmatrix} -\underline{c}_1 \\ -\underline{c}_2^* \end{pmatrix})$$

permite aplicar os resultados (69) e (70) a este caso de uma estrutura em classes encaixadas em 2 níveis. Δ

Dickey (1983) define ainda novas famílias, com base na função \mathcal{R} e em funções hipergeométricas com argumentos associados a agregados n -dimensionais, generalizando a família \mathcal{D}_k^ℓ , e mostra a sua aplicabilidade em análises Bayesianas de tabelas de contingência multidimensionais sob o modelo de independência.

Referências Bibliográficas

- AITCHISON, J. and SILVEY, S.D. (1960). Maximum likelihood estimation procedures and tests of significance. *J. Royal Statist. Soc.*, B, 31, 154-171.
- ALBERT, J.H. (1985). Bayesian estimation methods for incomplete two-way contingency tables using prior belief of association. In *Bayesian Statistics 2*, eds J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith, North-Holland, Amsterdam, p. 589-602.
- ANTELMAN, G.R. (1972). Interrelated Bernoulli processes. *J. Amer. Statist. Assoc.*, 67, 831-841.
- ASH, R.B. (1972). *Real Analysis and Probability*. Academic Press, New York.
- BAHADUR, R.R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.*, 25, 423-462.
- BAHADUR, R.R. (1955). Statistics and subfields. *Ann. Math. Statist.*, 26, 490-497.
- BARANKIN, E.W. (1960). Sufficient parameters: solution of the minimal dimensionality problem. *Ann. Inst. Statist. Math.*, 12, 91-118.
- BASU, A.P. (1983). Identifiability. In *Encyclopedia of Statistical Sciences 4*, eds N. L. Johnson and S. Kotz, p. 2-6, Wiley, New York.
- BASU, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.*, 72, 355-366.
- BASU, D. and PEREIRA, C.A. de B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *J. Statist. Plann. and Inference*, 6, 345-362.
- BASU, D. and PEREIRA, C.A. de B. (1983). Conditional independence in Statistics. *Sankhyā A*, 45, 324-337.
- BEMIS, K.G. and BHAPKAR, V.P. (1982). On the equivalence of some test criteria based on BAN estimators for the multivariate exponential family. *J. Statist. Plann. Inf.*, 6, 277-286.
- BEMIS, K.G. and BHAPKAR, V.P. (1983a). On BAN estimators for chi-square test criteria. *Ann. Statist.*, 11, 183-196.

- BEMIS, K.G. and BHAPKAR, V.P. (1983b). BAN estimation for chi-square test criteria in categorical data. *Comm. Statist. - Th. Meth.*, 12, 1211-1223.
- BERKSON, J. (1972). Minimum discrimination information, the 'no-interaction' problem, and the logistic function. *Biometrics*, 28, 443-468.
- BHAPKAR, V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Assoc.*, 61, 228-235.
- BIRCH, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Royal Statist. Soc., B*, 25, 220-233.
- BISHOP, Y.M.M., FIENBERG, S.E. and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.
- BLIGHT, B.J.N. (1970). Estimation from a censored sample for the exponential family. *Biometrika*, 57, 389-395.
- BLUMENTHAL, S. (1968). Multinomial sampling with partially categorized data. *J. Amer. Statist. Assoc.*, 63, 542-551.
- BONETT, D.G., WOODWARD, J.A. and BENTLER, P.M. (1985). Some extensions of a linear model for categorical variables. *Biometrics*, 41, 745-750.
- BOWDEN, R. (1973). The theory of parametric identification. *Econometrica*, 41, 1069-1074.
- BUNKE, H. and BUNKE, O. (1974). Identifiability and estimability. *Math. Operationsforsch. und Statist.*, 5, 223-233.
- CARLSON, B.C. (1971). Appell functions and multiple averages. *SIAM J. Math. Anal.*, 2, 420-430.
- CARLSON, B.C. (1977). *Special Functions of Applied Mathematics*. Academic Press, New York.
- CARTER, W.H. Jr. and MYERS, R.H. (1973). Maximum likelihood estimation from linear combinations of discrete probability functions. *J. Amer. Statist. Assoc.*, 68, 203-206.
- CHALONER, K. and DUNCAN, G.T. (1987). Some properties of the Dirichlet-Multinomial distribution and its use in prior elicitation. *Comm. Statist. - Th. Meth.*, 16, 511-523.

- CHEN, T. and FIENBERG, S.E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*, 30, 629-642.
- CHEN, T. and FIENBERG, S.E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics*, 32, 133-144.
- CLICKNER, R.P. (1976). On least squares estimation for categorical data. *Comm. Statist. - Th. Meth.*, A5, 1059-1064.
- COHEN, J.E. (1971). Estimation and interaction in a censored $2 \times 2 \times 2$ contingency table. *Biometrics*, 27, 379-386.
- CRAIG, C.C. (1953). Combination of neighboring cells in contingency tables. *J. Amer. Statist. Assoc.*, 48, 104-112.
- CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton, N.J., Princeton Univ. Press.
- DARROCH, J.N. and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. *Ann. Math. Statist.*, 43, 1470-1480.
- DAWID, A.P. and DICKEY, J.M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.*, 72, 845-850.
- DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc.*, B, 39, 1-38.
- DEISTLER, M. and SEIFERT, H.G. (1978). Identifiability and consistent estimability in econometric models. *Econometrica*, 46, 969-980.
- DICKEY, J.M. (1983). Multiple hypergeometric functions: Probabilistic interpretations and statistical uses. *J. Amer. Statist. Assoc.*, 78, 628-637.
- DICKEY, J.M., JIANG, J.M. and KADANE, J.B. (1983). Bayesian methods for multinomial sampling with noninformatively missing data. *Research Report*, 15, Department of Mathematics and Statistics, State University of New York at Albany, Albany, N.Y., 12222.
- DICKEY, J.M., JIANG, J.M. and KADANE, J.B. (1987). Bayesian methods for censored categorical data. *J. Amer. Statist. Assoc.*, 82, 773-781.
- DRÈZE, J. (1972). Econometrics and decision theory. *Econometrica*, 40, 1-17.

- DRÈZE, J. (1974). Bayesian theory of identification in simultaneous equations models. In *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, North-Holland, Amsterdam, p. 159-174.
- FAY, R.E. (1986). Causal models for patterns of nonresponse. *J. Amer. Statist. Assoc.*, 81, 354-365.
- FLORENS, J.P., MOUCHART, M. and ROLIN, J.M. (1983). *Elements of Bayesian Statistics*. Unpublished monograph.
- FUCHS, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *J. Amer. Statist. Assoc.*, 77, 270-278.
- GABRIELSEN, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8, 261-263.
- GOKHALE, D.V. (1973). Iterative maximum likelihood estimation for discrete distributions. *Sankhyā*, Ser. B 35, 293-298.
- GOKHALE, D.V. and KULLBACK, S. (1978). *The Information in Contingency Tables*. Marcel Dekker, Inc., New York.
- GOODMAN, L.A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing cells. *J. Amer. Statist. Assoc.*, 63, 1091-1131.
- GOODMAN, L.A. (1974). Exploratory latent-structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- GRIZZLE, J.E., STARMER, C.F. and KOCH, G.G. (1969). Analysis of categorical data by linear models. *Biometrics*, 25, 489-504.
- GRIZZLE, J.E. and WILLIAMS, O.D. (1972). Log-linear models and tests of independence for contingency tables. *Biometrics*, 28, 137-156.
- GUNEL, E. (1984). A Bayesian analysis of the multinomial model for a dichotomous response with nonrespondents. *Comm. Statist.-Th Meth.*, 13, 737-751.
- HABER, M. (1984). Fitting a general log-linear model (Algorithm AS 207). *Appl. Statist.*, 33, 358-362.

- HABER, M. (1986). Testing for pairwise independence. *Biometrics*, 42, 429-435.
- HABER, M. and BROWN, M.B. (1986). Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. *J. Amer. Statist. Assoc.*, 81, 477-482.
- HABERMAN, S.J. (1974). Log-linear models for frequency tables derived by indirect observation: maximum likelihood equations. *Ann. Statist.*, 5, 911-924.
- HABERMAN, S.J. (1976). Iterative scaling procedures for log-linear models for frequency tables derived by indirect observation. 1975 Statistical Computing Section, *Proc. Amer. Stat. Assoc.*, 45-50.
- HABERMAN, S.J. (1977). Product models for frequency tables involving indirect observation. *Ann. Statist.*, 5, 1124-1147.
- HALMOS, P.R. (1958). *Finite-dimensional Vector Spaces*, 2nd ed., D. Van Nostrand Co., Inc., Princeton, N.J.
- HALMOS, P.R. and SAVAGE, L.J. (1949). Application of the Radon- Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.*, 20, 225-241.
- HARTLEY, H.O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14, 174-194.
- HARTLEY, H.O. and HOCKING, R.R. (1971). The analysis of incomplete data. *Biometrics*, 27, 783-808.
- HOCKING, R.R. and OXSPRING, H.H. (1971). Maximum likelihood estimation with incomplete multinomial data. *J. Amer. Statist. Assoc.*, 66, 65-70.
- HOCKING, R.R. and OXSPRING, H.H. (1974). The analysis of partially categorized contingency data. *Biometrics*, 30, 469-483.
- IMREY, P.B., KOCH, G.G. and STOKES, M.E. et. al. (1981,1982). Categorical data analysis: Some reflections on the log linear model and logistic regression; Part I: Historical and methodological overview. *Int. Statist. Rev.*, 49, 265-283; Part II: Data analysis. *Int. Statist. Rev.*, 50, 35-64.

- IRELAND, C.T. and KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika*, 55, 179-188.
- KADANE, J.B. (1972). Identification of functions of parameters. *Unpublished Technical Report*, Carnegie - Mellon University.
- KADANE, J.B. (1974). The role of identification in Bayesian theory. In *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, North-Holland, Amsterdam, p. 175-191.
- KADANE, J.B. (1985). Is victimization chronic? A Bayesian analysis of multinomial missing data. *Journal of Econometrics*, 29, 47-67.
- KARSON, M.J. and WROBLESKI, W.J. (1970). A Bayesian analysis of binomial data with a partially informative category. In *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, p. 532-534.
- KAUFMAN, G.M. and KING, B. (1973). A Bayesian analysis of non-response in dichotomous processes. *J. Amer. Statist. Assoc.*, 68, 670-678.
- KOCH, G.G. (1969). A useful lemma for proving the equality of two matrices with applications to least squares type quadratic forms. *J. Amer. Statist. Assoc.*, 64, 969-970.
- KOCH, G.G. and BHAPKAR, V.P. (1982). Chi-square tests. In *Encyclopedia of Statistical Sciences* 1, N.L. Johnson and S. Kotz (eds.), 442-457, Wiley, New York.
- KOCH, G.G., IMREY, P.B. and REINFURT, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. *Biometrics*, 28, 663-692.
- KOCH, G.G. and REINFURT, D.W. (1970). The analysis of complex contingency table data from general experimental designs and sample surveys. *Proceedings of the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing*, Fort Lee, Va.
- KOOPMANS, T.C. and REIERSOL, O. (1950). The identification of structural characteristics. *Ann. Math. Statist.*, 21, 165-181.
- KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.

- LANDIS, J.R., STANISH, W.M., FREEMAN, J.L. and KOCH, G.G. (1976). A computer program for the generalized chi-square analysis of categorical data using weighted least squares (GENCAT). *Computer Programs in Biomedicine*, 6, 196-231.
- LEAMER, E.E. (1978). *Specification Searches*. John Wiley & Sons, New York.
- LECAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, vol.1, p. 128-164, University of California Press, Berkeley, CA.
- LEHMANN, E. (1986). *Testing Statistical Hypotheses*, 2nd ed., John Wiley & Sons, New York.
- LINDLEY, D.V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, 35, 1622-1643.
- LINDLEY, D.V. (1971). *Bayesian Statistics: A Review*. SIAM, Philadelphia.
- LITTLE, R.J.A. (1980). "Superpopulation models for non-response. I: The ignorable case. II: The non-ignorable case". In *Non-response in Sample Surveys: The Theory of Current Practice*; Part V (1983), Panel on incomplete Data, National Academy of Sciences, Washington, D.C.
- LITTLE, R.J.A. (1982). Models for nonresponse in sample surveys. *J. Amer. statist. Assoc.*, 77, 237-250.
- LITTLE, R.J.A. and RUBIN, D.B. (1983a). Incomplete data. In *Encyclopedia of Statistical Sciences*, 4, N.L. Johnson and Kotz (eds.), 46-53, Wiley, New York.
- LITTLE, R.J.A. and RUBIN, D.B. (1983b). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *Amer. Statistician*, 37, 218-220.
- LITTLE, R.J.A. and SCHLUCHTER, M.D. (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, 72, 497-512.
- LOUIS, T.A. (1982). Finding the observed information when using the EM algorithm. *J. Royal Statist. Soc.*, B, 44, 226-233.

- MURRAY, L.W. and SMITH, D.W. (1985). Estimability, testability and connectedness in the cell means model. *Comm. Statist. - Th. Meth.*, 14, 1889-1917.
- NERING, E.D. (1964). *Linear Algebra and Matrix Theory*. John Wiley & Sons, Inc., New York.
- NEYMAN, J. (1949). Contributions to the theory of the χ^2 -test. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (ed.), 239-273. University of California Press, Berkeley.
- NORDHEIM, E.V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *J. Amer. Statist. Assoc.*, 79, 772-780.
- ORCHARD, T. and WOODBURY, M.A. (1972). A missing information principle: theory and applications. In *Proceedings 6th Berkeley Symposium on Math. Statist. and Prob.*, 1, 697-715.
- PAULINO, C.D.M. (1985). Um ensaio sobre identificabilidade de modelos estatísticos paramétricos. *Relatório Técnico IME-USP*, São Paulo, Brasil.
- PICCI, G. (1974). Theory of sufficient statistics and structure analysis of transition probabilities with applications to identifiability. *Rep. CNR-LADSEB 74/06*, Padova, Italy.
- PICCI, G. (1977). Some connections between the theory of sufficient statistics and the identifiability problem. *SIAM J. Appl. Math.*, 33, 383-398.
- RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Massachusetts Institute of Technology Press.
- RAO, C.R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philos. Soc.*, 44, 50-57.
- REIERSOL, O. (1963). Identifiability, estimability, pheno-restricting specifications and zero Lagrange multipliers in the analysis of variance. *Skandinavisk Aktuarietidskrift*, 131-142.
- ROTHENBERG, T. (1971). Identification in parametric models. *Econometrica*, 39, 576-591.

- RUBIN, D.B. (1974). Characterizing the estimation of parameters in incomplete-data problems. *J. Amer. Statist. Assoc.*, 69, 467-474.
- RUBIN, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- RUBIN, D.B. (1978). A note on Bayesian, likelihood and sampling distribution inferences. *J. Educat. Statist.*, 3, 189-201.
- SAS (1985). SAS User's Guide: Statistics, SAS Institute Inc. Cary, North-Carolina.
- SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.
- SHEFRIN, H.M. (1981). On the combinatorial structure of Bayesian learning with imperfect information. *Discrete Mathematics*, 37, 245-254.
- SHEFRIN, H.M. (1983). Markov chains, imperfect state information and Bayesian learning. *Mathematical Modelling*, 4, 1-7.
- SILVEY, S.D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.*, 30, 389-407.
- SILVEY, S.D. (1975). *Statistical Inference*. Chapman and Hall, London.
- SMITH, P.J. and GUNEL, E. (1984). Practical Bayesian approaches to the analysis of a 2×2 contingency table with incompletely categorized data. *Comm. Statist. - Th. Meth.*, 13, 1941-1963.
- STANISH, W.M., GILLINGS, D.B. and KOCH, G.G. (1978). An application of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data. *Biometrics*, 34, 305-317.
- SOC (1988). Software Científico, Núcleo Tecnológico para Informática Agropecuária, EMBRAPA, C. Postal 5010, 13081 - Campinas, SP.
- SUNDBERG, R. (1974). Maximum likelihood theory for solution of the incomplete data from an exponential family. *Scand. J. Statist.*, 1, 49-58.
- SUNDBERG, R. (1976). An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Comm. Statist. - Simul. Comput.*, B5, 55-64.
- TIERNEY, L. and KADANE, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, 81, 82-86.

- TITTERINGTON, D. M. (1984). Recursive parameter estimation using incomplete data. *J. Royal Statist. Soc., B*, 46, 257-267.
- TITTERINGTON, D. M. and JIANG, J. M. (1983). Recursive estimation procedures for missing-data problems. *Biometrika*, 70, 613-624.
- TUNCER, Y. (1985). Equivalence of parametric identifiability and estimability. *Comm. Statist. - Th. Meth.*, 14, 393-400.
- VARDI, Y., SHEPP, L.A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography (with discussion). *J. Amer. Statist. Assoc.*, 80, 8-37.
- WATSON, G.S. (1956). Missing and 'mixed-up' frequencies in contingency tables. *Biometrics*, 12, 47-50.
- WILKS, S.S. (1962). *Mathematical Statistics*. John Wiley & Sons, Inc., New York.
- WOODBURY, M.A. (1971). Discussion of the paper by Hartley and Hocking. *Biometrics*, 27, 808-817.
- WOOLSON, R.F. and BRIER, S.S. (1981). Equivalence of certain chi-squared test statistics. *Amer. Statistician*, 35, 250-253.
- WOOLSON, R.F. and CLARKE, W.R. (1984). Analysis of categorical incomplete longitudinal data. *J. Royal Statist. Soc., A*, 147, 87-99.
- WU, J.C.F. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, 11, 95-103.
- ZACKS, S. (1971). *The Theory of Statistical Inference*. John Wiley, New York.
- ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.