Imprevistos e suas Conseqüências

Rosangela Helena Loschi

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU
DE
DOUTOR EM ESTATÍSTICA

Área de Concentração: **Estatística**

Orientador: **Profa. Dra. Pilar Loreto Iglesias Zuazola**

Co - Orientador: **Prof. Dr. Reinaldo B. Arellano Valle**

-São Paulo, Julho de 1998-

Apoio Financeiro: CAPES, FONDECYT e FAPEMIG

Imprevistos e suas Conseqüências

Este exemplar corresponde à redação
final da tese, devidamente corrigida,
apresentada por Rosangela Helena Loschi e
aprovada pela Comissão Julgadora.

São Paulo, 16 de Setembro de 1998.

Banca Examinadora

-Prof. Dr. Reinaldo Boris Arellano Valle (Presidente)	PUC-CHILE
-Prof. Dr. Heleno Bolfarine	IME-USP
-Prof. Dr. Josemar Rodrigues	ICMC-USP
-Prof. Dr. Dani Gamerman	IM-UFRJ
-Prof. Dr. Hélio dos Santos Migon	IM-UFRJ

Abstract

The influence of information not previously supposed to become available in the process of construction of posterior distributions is discussed. Firstly, we will consider how should probabilities be updated when unexpected information becomes available immediately before elicitation of posterior opinions. Some updating procedures will be introduced and their fragility exposed. Jeffrey's Rule will be seen in detail, with its predictivistic character, and the concept of conjugation relative to this rule will be introduced. Bayesian Conditioning will also be presented and its failure in incorporating unexpected information will be discussed. A predictivistic characterization of the t-linear model arises from judgments about observables only. This procedure give us a completely operational way of modeling and yield more examples of Jeffrey-conjugation. We will finally present a procedure for inference about changepoints, which sometimes arises because unpredictables ocurr, using the Partition Product Model. This procedure will be specified for inferences about changepoints of means and variances when data are supposed to be conditionally normally distributed. Procedures obtained here will be applied to financial data.

Agradecimentos

Depois de tanto tempo(eles já devem estar de saco cheio), finalmente, posso tornar pública a minha gradidão à Pilar, por ter aceitado ser minha orientadora, por me incentivar e discutir a maior parte deste trabalho, por seu entusiasmo, por ter me obrigado a abrir os olhos para outros horizontes dentro da estatística, pelas suas valiosas sugestões e idéias e por ter convidado ao Reinaldo para ser meu co-orientador. Ao Reinaldo gostaria de agradecer pelo carinho e dedicação com que leu e corrigiu todas as minhas contas e por me ter permitido continuar um trabalho que certamente ele faria com os pés nas costas. Aos dois gostaria de agradecer, sinceramente, por todo o carinho com que me receberam no Chile, por ter acreditado sempre que eu era capaz de fazer quaisquer das maluquices que me propuseram, por todas as cervejas e por muito mais.

Ao Sergio gostaria de agradecer por todas as pulgas que vem colocando atrás da minha orelha desde o mestrado (inclusive por uma das molas mestras deste trabalho). Para algumas de suas perguntas eu tenho boas respostas, para outras não sei se algum dia chegarei a ter. Me desculpe, mas eu vou continuar tentando.

Aos professores Carlos Pereira e Nelson Tanaka, por seus comentários e críticas na fase inicial deste trabalho, os quais foram muito enriquecedores para a minha formação, ao professor Heleno Bolfarine por sua leitura minusciosa do capítulo 3 e seus bons comentários a respeito deste trabalho e aos professores Guido del Pino e Héctor Varela por seus comentários e sugestões a respeito do capítulo 1 e 2.

Ao Michel, pela implementação do programa exibido no apêndice 2, ao Marcel pelas gírias em espanhol e pela implementação do programa apresentado no apêndice 3 e ao Marcelo por sua paciência e ajuda na correção destes programas.

À minha família por ter- me ensinado a nunca a desistir, antes de tentar de verdade, a

realizar meus sonhos. Não me canso de falar que sem o carinho de vocês isso tudo não teria nexo. Em particular, muito obrigada Dona Helena, por todas as orações, Hilda, Wagner, Carlos e Cláudio, pela torcida sempre organizada que fazem por mim, Fê, por seu espiritísmo de porco; Ana Lo,por aquele mundo rosinha para onde você sempre nos remete e Ana K, pelo seu grande talento de driblar as coisas ruins da vida.

À Silvia (às vezes, Pfaifer, outras Popo), Zó, Marcelo e Fernando pela incrível resistência ao meu bom humor matinal e lágrimas noturnas. Apesar disso, espero que as boas risadas tenham podido compensar as lamúrias. Se não, que tal uma cervejinha ou um chocolatinho prá amenizar??? Afinal, vocês são minha segunda família!

À meus grandes incentivadores e grandes amigos (Lu, Ching, Ju, Lucho, Nani, Paulinha, No) quem sabe depois desta possamos nos encontrar de quando em vez? Já estou cansada de só saber de vocês e de levar puxão de orelha por via eletrônicas(desta vez, juro que não reclamo de nada).

À todos os meus amigos e colegas de doutorado, pelos cafés, pelas boas risadas, pelos bons papos,etc.

À PUC-Santiago, em particular, à Facultad de Matemáticas, por ter permitido e favorecido a realização de parte deste trabalho em suas instalações. Mais que isso gostaria de agradecer a todas as pessoas que conheci no Chile, especialmente, à TODOS professores do depto de estatística, à Rosita, Mônica, Lorena, Lucia, Flor, Suzi e ao Job. Muito obrigada pelo carinho, por todo esforço que fizeram para que eu me sentisse menos turista, pelos piscos e pela salsa.

Ao depto de Estatística da UFMG por ter me permitido estar onde precisava sempre que lhe foi possível. E à Mercedes, Aldy, Enrico e Fred pelo carinho, incentivo e amizade.

Faltou alguém... Eu sei, mas só vou lembrar depois de entregar a tese. Lamento!!!

Ao meu pai

Índice

Introdução 4

Algumas Notações 8

1 Imprevistos e o Condicionamento Bayesiano 10

1.1 Introdução 10

1.2 O Problema 12

1.3 O Condicionamento Bayesiano 17

1.4 Supercondicionamento e Condicionamento Abstrato 23

1.5 Imprevistos e o Paradoxo dos Três Prisioneiros 33

2 A Regra de Jeffrey 37

2.1 Introdução 37

2.2 Apresentando a Regra de Jeffrey 39

2.2.1 Definição e Conexões com Bayes e Suficiência 39

2.2.2 Comutatividade e um Pouco Mais de Fundamentos 50

2.3 A Regra Preditivista de Jeffrey 65

2.4 A Jeffrey Conjugação 80

2.5 J-Conjugação Natural para a Família Exponencial 88

3 Caracterização Preditivista de Modelos Lineares t 100

3.1 Introdução 100

3.2 Simetrias Esféricas: Definições e Interpretações 103

3.3 Preditivismo e o Modelo Linear t 123

4. Imprevistos e Pontos de Mudança 135

4.1 Introdução 135

4.2 Modelos Partição Produto - MPP 137

4.3 O Cálculo da Relevância 146

4.4 Aplicações a Modelos t 147

4.5 Gibbs e MPP: A Distribuição *a posteriori* para ρ 152

5. Aplicações 156

5.1 Introdução 156

5.2 Mercado Eficiente e o Modelo Adotado 158

5.3 Os Dados 160

5.4 A Regra de Jeffrey e o Mercado Chileno 178

5.5 O MPP e o Mercado Chileno 194

Considerações Finais 218

Apêndice Um 220

Apêndice Dois 224

Apêndice Três 233

Apêndice Quatro 252

Referências Bibliográficas 258

Introdução

Como **Você** (pessoa subjetivista) deve atualizar suas opiniões iniciais após receber alguma informação relevante mas cuja ocorrência não havia sido previamente antecipada?

Este é um problema ainda não satisfatoriamente resolvido por Bayesianos. Segundo Goldstein(1985), não existe um procedimento cujo uso seja imperativo neste caso, por que não existe uma definição de coerência que justifique a existências de qualquer relação matemática entre suas opiniões *a priori* e *a posteriori*. De fato, a inexistência desta definição de coerência também impede a construção de um mecanismo normativo para atualizar probabilidades mesmo naquelas situações onde tudo o que é possível ocorrer já havia sido previamente antecipado. Por este motivo, Você pode atualizar suas probabilidades usando qualquer procedimento que julgar razoável.

Um mecanismo de atualização de probabilidades cujo uso é bastante difundido é o Condicionamento Bayesiano. Quando seu uso é possível, o Condicionamento Bayesiano permite a construção de sua opinião *a posteriori*, partindo de sua distribuição *a priori*, utilizando a fórmula de Bayes. Uma característica importante deste procedimento é que a distribuição *a posteriori* gerada por este mecanismo também é uma distribuição de probabilidade desde que sua opinião inicial seja coerente - no sentido estático proposto por de Finetti(1931). Apesar desta qualidade, o Condicionamento Bayesiano só responde a questão primeira em situações muito particulares.

Por essa razão, iniciamos este trabalho apresentando formalmente o Condicionamento Bayesiano, discutindo a diferença essencial entre este mecanismo de atualização de probabi-

lidades e a fórmula de Bayes e explorando algumas de suas limitações enquanto mecanismo de atualização de probabilidades. A sua pouca flexibilidade para incorporar informações imprevistas na construção da distribuição *a posteriori* será vista em detalhes. Também no capítulo 1 apresentaremos duas outras formas de condicionamento as quais permitem atualizar probabilidades em situações um pouco mais gerais do que o Condicionamento Bayesiano: o Supercondicionamento e o Condicionamento Abstrato. Tanto o Supercondicionamento, definido por Diaconis e Zabell(1982), quanto o Condicionamento Abstrato, não encontrado por nós na literatura, permitem uma reavaliação arbitrária da função responsável pela incorporação das informações na atualização da distribuição *a priori* e por isto podem ser úteis na incorporação de notícias imprevistas. Apresentaremos também resultados particulares para a validade destas formas de atualização de probabilidades. Para finalizar este capítulo, apresentaremos uma nova explicação para o Paradoxo dos Três Prisioneiros dentro do contexto estudado em parte deste trabalho - a saber, como atualizar probabilidades diante do conhecimento de algo completamente novo ou não previamente antecipado.

Informações imprevista podem também agir sobre seu estado de conhecimento modificando arbitrariamente sua opinião inicial sobre alguns eventos definidos sobre o espaço amostral de interesse. Quando isto ocorre, o Condicionamento Bayesiano não permite atualizar a distribuição *a priori*, salvo nas situações muito particulares em que a informação recebida o conduza à certeza *a posteriori* sobre o evento condicionante. Para muitas destas situações, ainda não foi dada uma resposta. Quando a informação recebida mantém inalterado o espaço amostral inicial e faz com que Você modifique arbitrariamente suas opiniões iniciais sobre todos os eventos de uma partição do espaço amostral, sem necessariamente conduzi-lo à certeza da ocorrência de qualquer deles, Jeffrey(1965) propõe um mecanismo para construir sua probabilidade *a posteriori* para qualquer outro evento deste espaço amostral a qual é conhecida como Regra de Jeffrey. Esta regra, além do mais, requer que todas as probabilidades condicionais nos eventos desta partição se mantenham inalteradas *a posteriori*. No capítulo 2, apresentaremos a Regra de Jeffrey detalhadamente, discutindo algumas de suas limitações e suas conexões com o Condicionamento Bayesiano. Alguns aspectos fundamentais, como a comutatividade e as conseqüências da violação da Lei de Cromwell à luz

de tal regra, serão discutidos. Neste capítulo também poderá ser encontrada uma versão preditivista para a Regra de Jeffrey e estabeleceremos as definições de Conjugação e Conjugação Natural segundo a Regra de Jeffrey - a Jeffrey-conjugação e a Jeffrey-conjugação natural. Resultados particulares para a Jeffrey-conjugação natural para membros da família exponencial serão exibidos. Tanto a Regra Preditivista de Jeffrey quanto os resultados sobre conjugação segundo a Regra de Jeffrey não foram encontrados na literatura.

No capítulo 3, forneceremos mais alguns resultados sobre Jeffrey- Conjugação natural, só que agora dentro de um contexto completamente preditivista. Nesta parte do trabalho mostramos a inevitabilidade da adoção de um particular modelo t - e, conseqüentemente, de uma particular distribuição *a priori* Normal-Gama-Inversa - se uma certa condição de invariância e particulares condições de linearidade sobre a esperança condicional de observáveis forem percebidas na seqüência de variáveis aleatórias. Neste caso, a Jeffrey-conjugação ocorrerá se a informação recebida mantiver a condição de invariância inicialmente considerada e modificar apenas os coeficientes que indexam as condições de linearidade para a esperança condicional. Mais que simples exemplos de famílias Jeffrey-conjugadas naturais, os resultados obtidos neste capítulo destinam-se a exibir uma caracterização completamente preditivista para modelos lineares t (particularmente, para modelos de regressão, modelos de análise de variância, etc) partindo apenas de suposições sobre objetos observáveis. Ao fazer isto, fornecemos um procedimento completamente operacional para modelagem que permite evitar a adoção conveniente e arbitrária de particulares distribuições *a priori* para descrever sua incerteza sobre parâmetros de caráter não operacional eventualmente envolvidos no problema. Também será fornecido neste capítulo uma prova alternativa (e mais simples do que a encontrada na literatura) para o teorema tipo de Finetti que representa seqüências de observações $O(\mathcal{M})$-invariantes. Os principais resultados deste capítulo parece-nos inéditos.

Muitas vezes os dados já foram observados e Você percebe que ocorrem mudanças em seu comportamento ao longo do tempo (por exemplo). Estas mudanças muitas vezes foram provocadas pelo recebimento de algumas informações não previamente consideradas e conhecer os instantes precisos destas mudanças podem ser esclarecedores sobre o fenômeno que está sendo estudado. O capítulo 4 deste trabalho visa apresentar um procedimento para fazer

inferências sobre estes pontos de mudança utilizando o Modelo Partição Produto definido por Hartigan(1990). Nossa maior contribuição aqui está em fornecer resultados particulares para inferências sobre pontos de mudança na média e na variância de dados julgados condicionalmente distribuídos segundo uma Normal, estendendo os resultados de Correa(1998) e Crowley(1997). O procedimento para inferir sobre o número de pontos de mudança ocorridos também é discutido. Devido a complexidade dos cálculos que envolvem a construção desta distribuição *a posteriori*, a determinação exata de sua forma é praticamente inviável, exigindo o uso de procedimentos computacionais para gerar uma aproximação desta distribuição. Este procedimento computacional é desenvolvido detalhadamente neste capítulo e embora tenha sido sugerido na literatura, não encontramos sua implementação em qualquer referência.

Finalmente, no capítulo 5, apresentaremos uma aplicação da Regra Preditivista de Jeffrey e dos resultados apresentados no capítulo 4 a dados do mercado acionário chileno. A Regra Preditivista de Jeffrey é o núcleo do algoritmo que sugerimos para fazer previsões de retornos futuros incorporando imprevistos. Confrontaremos as previsões feitas utilizando este procedimento com aquelas obtidas utilizando apenas o Condicionamento Bayesiano. O algoritmo proposto, que foi implementado em FORTRAN, pode ser encontrado no apêndice dois no final do trabalho. O procedimento computacional apresentado no capítulo 4 é implementado em PERL (ver apêndice três), permitindo concluir sobre a existência de pontos de mudança tanto no retorno esperado quanto na volatilidade na série de dados analisada . Justificaremos o modelo e as distribuições *a priori* adotadas aqui baseando-nos em fundamentos da teoria econômica de mercados eficientes.

Esperamos também que este trabalho contribua de alguma forma para o enriquecimento do H de cada leitor. A seguir apresentamos algumas notações que serão utilizadas ao longo deste trabalho.

Algumas Notações

- **v.a.** - variável aleatória.
- **q.a.** - quantidade aleatória.
- **ciid** - condicionalmente independentes e identicamente distribuídas.
- $\mathcal{L}(U)$ - lei da q.a. U.
- $1_{\{A\}}$ - função indicadora do evento A.
- $\mathbf{1}_{a\times b}$ - matriz de 1's formada por a linhas e b colunas.
- I_n - matriz identidade de ordem n.
- P - distribuição ou probabilidade *a priori*.
- P^* - distribuição ou probabilidade *a posteriori* gerada por algum mecanismo de atualização de probabilidades.
- $P_{\mathcal{E}}$ - distribuição ou probabilidade *a posteriori* gerada por algum mecanismo de atualização de probabilidades após o conhecimento da informação que gerou uma partição $\mathcal{E}$ do espaço amostral.
- $P_{\mathcal{E}\mathcal{F}}$ - distribuição ou probabilidade *a posteriori* gerada por algum mecanismo de atualização de probabilidades após o conhecimento sucessivo das informações que geraram as partições $\mathcal{E}$ e $\mathcal{F}$ do espaço amostral (nesta ordem).
- P^k_{β,σ^2} - lei do vetor aleatório com distribuição $\mathcal{N}(\mathbf{X}_k\beta, \sigma^2 I_k)$.

- $P^{\mathcal{M}}$ - Projeção sobre o espaço $\mathcal{M}$

- Q^k_{n,β,σ^2} - lei do vetor aleatório com distribuição Pearson tipo II com vetor de locação $\mathbf{X}_k\beta$, matriz de dispersão $\sigma^2 V_k$ e parâmetro de forma $\frac{n-p-k-2}{2}$.

- $\sigma(A)$ - sigma-álgebra gerada pelo objeto aleatório A.

- $\mathcal{B}^a$ - sigma-álgebra de Borel definida sobre $\mathcal{R}^a$.

- $\mathbf{X}_{(n)}$ - vetor aleatório $(X_1, \ldots, X_n)^t$.

- $\mathbf{X}^i_{(n)}$ - vetor aleatório $(X_{(i-1)k+1}, \ldots, X_{ik})^t$.

- X_{ij} - bloco de observações $[X_{i+1}, \ldots, X_j]$.

- $\mathbf{X}$ - matriz de constantes conhecidas de ordem $n \times p$, com posto completo p.

- $\| \ \|$ - norma euclidiana.

- $\| \ \|_V$ - distância de variação total.

- $\mathbf{supp}(P) = \{w \in \Omega : P(w) \neq 0\}$ - suporte da distribuição P.

- $=^d$ - igual em distribuição.

Capítulo 1

Imprevistos e o Condicionamento Bayesiano

1.1 Introdução

Este capítulo apresenta o Princípio de Condicionalização Bayesiana e suas limitações como mecanismo de atualização de probabilidades em situações onde "imprevistos" ocorrem.

Na seção 1.2, apresentaremos um problema a ser abordado em grande parte deste trabalho - a saber, a incorporação de informações imprevistas no processo de atualização de probabilidades - ressaltando a importância deste estudo através de alguns exemplos que ilustram, entre outras, situações onde o uso do Condicionamento Bayesiano é impossível para tal fim.

Atualizar probabilidades significa obter através de algum procedimento (fórmulas matemáticas, retaxações arbitrárias, etc) a distribuição de probabilidade procedente do conhecimento de novas informações - cuja disponibilidade ou existência tenham sido ou não previamente antecipadas - partindo-se da distribuição de probabilidade inicial- isto é, da distribuição de probabilidade que descrevia sua incerteza antes da aquisição de novas informações .

Usualmente, a distribuição de probabilidade inicial é chamada de distribuição *a priori* e a distribuição de probabilidade determinada após novas informações é denominada distribuição *a posteriori*, sendo esta última obtida da primeira através de um Condicionamento Bayesiano (menos rigorosamente falando, via Fórmula de Bayes - ver seção 1.3), sempre que este seja possível, ou através de qualquer outro mecanismo de atualização de probabilidades que Você julgar adequado.

Salientamos que a distribuição *a posteriori*, mesmo que obtida da distribuição *a priori* via Condicionamento Bayesiano, é declarada em um instante de tempo e geralmente em um estado de informação distintos daqueles em que foi declarada a distribuição *a priori*, como seus próprios nomes insinuam. (Ver mais sobre isto na seção 1.3).

Perseguir uma fórmula matemática que atualize probabilidades é buscar uma economia nas tarefas mentais que constróem sua distribuição *a posteriori* - pois evitaremos retaxações sobre todos os eventos de interesse - e além disto, é procurar uma maneira de verificar que suas novas atribuições de probabilidade são em algum sentido coerentes. Segundo Goldstein(1985) a relação entre sua antiga opinião e sua nova opinião ocorre apenas no "reino das crenças" e por isto não se pode obter uma fórmula matemática cujo uso seja inevitável na atualização das antigas para novas opiniões. Isto é, para Goldstein(1985), não há qualquer definição de coerência que garanta a existência de uma única fórmula matemática a qual Você está obrigado a usar, se não quiser perder certamente, para determinar suas probabilidades *a posteriori*, partindo-se de suas probabilidades *a priori*. Sendo assim, Você está livre para escolher qualquer mecanismo de atualização de probabilidades que julgar conveniente pois a inexistência de tal definição de coerência não permite a construção de uma teoria normativa para atualização de probabilidades.

Um mecanismo de atualização de probabilidades bastante utilizado é o **Condicionamento Bayesiano** que será visto em detalhes na seção 1.3. Nesta seção também discutiremos a quase impossibilidade de se incorporar de informações imprevistas na obtenção da distribuição *a posteriori* quando o mecanismo de atualização adotado é o Condicionamento Bayesiano e também abordaremos a diferença essencial deste para a fórmula de Bayes que geralmente o titula. Na seção 1.4, apresentaremos versões mais gerais de condicionamento

explorando o conceito de **Supercondicionamento** apresentado por Diaconis e Zabell(1982) e definindo a **Regra Geral de Condicionamento** a qual não encontramos na literatura. Forneceremos resultados para a possibilidade de uso destas versões de condicionamento. Finalmente, na seção 1.5 forneceremos uma nova explicação para o Paradoxos dos Três Prisioneiros dentro do contexto de atualização de probabilidades diante de imprevistos.

1.2 O Problema

Nos últimos anos, o uso da fórmula de Bayes - ou, mais corretamente, do Condicionamento Bayesiano - como mecanismo de atualização de probabilidades tem sido amplamente difundido graças à crescente adoção de uma postura mais subjetivista em Estatística.[1]

O Condicionamento Bayesiano mostra-se razoável como mecanismo de incorporação de informações em sua opinião inicial sempre que Você está em situações muito bem comportadas, onde **todo** recebimento de informação é previamente considerado e nenhum fato distinto daqueles anteriormente supostos lhe é revelado antes de Você tornar pública sua distribuição de probabilidade atualizada - isto é, antes da declaração de suas probabilidades *a posteriori*. (Ver detalhes na seção 1.3).

Por exemplo, suponha que Você queira realizar dois lançamentos de uma moeda, seqüencialmente e utilizar o resultado obtido no primeiro lançamento para atualizar sua opinião *a priori* para o evento $A =$ ocorre cara no segundo lançamento da moeda. Represente por X_i o evento "ocorre cara no i-ésimo lançamento da moeda" e suponha que X_i assume o valor um ou zero se ocorre ou não uma cara no i-ésimo lançamento da moeda, respectivamente.

[1] Pouco se sabe sobre os primórdios da adoção de tal postura em Estatística mas é bem aceito dentre os estatísticos que um de seus precursores foi o Rv. Thomas Bayes, a quem também se atribui o estabelecimento do teorema de Bayes que é uma ferramenta para a Estatística Subjetivista ou Bayesiana. Mais que emprestar seu nome à Escola Subjetivista de Estatística, Bayes parece ter sido de fato um Bayesiano (um Subjetivista) como pode ser percebido em sua obra *An Essay Towards Solving a Problem in the Doctrine of Chances* publicado postumamente em 1763 e 1958. Argumentos defendendo o subjetivismo de Bayes, bem como a sua postura preditivista e operacionalista, são apresentados detalhadamente em Stigler(1982).

Usando esta notação o espaço amostral adequado para representar este experimento é

$$\Omega = \{0,1\} \times \{0,1\} = \{0,1\}^2.$$

Inicialmente Você julga que a moeda é honesta e portanto cada ponto do espaço amostral descrito anteriormente tem chance 1/4 de ocorrer. Sendo assim, *a priori* a probabilidade de A ocorrer é

$$P(A) = P(X_2 = 1) = 1/2.$$

Se de fato ocorre uma cara no primeiro lançamento e esta é toda a informação da qual Você toma conhecimento antes de declarar sua opinião *a posteriori*, Você pode atualizar sua opinião sobre a ocorrência de cara no segundo lançamento usando o Condicionamento Bayesiano e com isto obtém-se

$$P^*(A) = P(X_2 = 1|X_1 = 1) = 1/2$$

onde P^* denota a sua opinião *a posteriori*.

Note que neste exemplo toda a informação aprendida por Você havia sido previamente considerada no espaço amostral inicialmente determinado e, portanto, o uso do Condicionamento Bayesiano para atualizar as probabilidades iniciais é - em princípio - perfeitamente justificado (ver seção 1.3).

Em muitas situações, a consideração antecipada de toda informação relevante para o problema não é possível, ou por desconhecimento completo do fato, ou simplesmente por que Você esqueceu-se de considerá-lo em sua análise prévia. O conhecimento deste novo fato ou a mera recordação de sua existência pode provocar a modificação do espaço amostral inicialmente considerado ou pode, apenas, levar Você a uma alteração de sua opinião *a priori* sobre os eventos do espaço amostral inicial sem modificá-lo, no entanto. Esta modificação na opinião inicial (ou no próprio espaço amostral) pode muitas vezes ser feita sem que Você saiba descrever precisamente o que provocou tal mudança, como exemplificaremos mais adiante.

Um exemplo fictício mas interessante para ilustrar situações onde uma modificação no espaço amostral ocorre ao se tomar conhecimento de novas informações é o problema de

estimação do número de espécies animais N em uma determinada região do planeta - por exemplo, no deserto do Saara. Suponha que o número de espécies animais distintas que habitam a terra é um número conhecido t e que, por simplicidade, faremos apenas uma observação X que revela a espécie observada na realização do experimento.

Ao descrever o conjunto de todas as possibilidades para N, Você pode considerá-lo o mais amplo possível, incluíndo na sua determinação todas as espécies que habitam atualmente a terra, inclusive aquelas espécies que Você julga não terem chance de viver em regiões de clima desértico - um urso polar, por exemplo. Ou seja, para Você

$$\Omega_N = \{0, 1, \ldots, t\}.$$

Se, ao realizar o experimento, Você observa um camelo - espécie que reconhecidamente habita regiões desérticas e para a qual Você atribuiu probabilidade inicial $P(X = \text{camelo}) > 0$ - Você poderá atualizar sua opinião sobre N utilizando o Condicionamento Bayesiano obtendo a seguinte opinião *a posteriori* para N

$$\begin{aligned} P^*(N = n) &= P(N = n | X = x) \\ &= \frac{P(X = x | N = n)}{P(X = x)} P(N = n), \quad \forall n \in \Omega_N, \end{aligned} \tag{1.1}$$

onde $P^*(N = n)$ e $P(N = n)$ são, respectivamente, suas opiniões *a posteriori* e *a priori* para o evento $\{N = n\}$ e $x = \text{camelo}$.

Por outro lado, se a espécie observada X é um urso polar - para o qual Você atribuiu probabilidade inicial zero, isto é, $P(X = \text{urso polar}) = 0$ - embora Você fique surpreso, esta informação não altera o espaço amostral inicial Ω_N nem leva Você a questionar a opinião *a priori* atribuída à quantidade aleatória número de espécies animais no deserto do Saara - $P(N = n)$, $n = 1, \ldots, t$ - pois, embora por Você julgada impossível, a ocorrência de um urso polar já havia sido admitida ao se tomar Ω_N o mais amplo possível. Esta informação, mesmo que intuitivamente impossível, já havia sido considerada tanto na declaração de sua opinião *a priori* sobre o número de espécies da região quanto na determinação de Ω_N e por tanto não provoca mudanças exógenas em nenhuma destas atribuições. Nesta situação, como

$P(X = \text{urso polar}) = 0$ o Condicionamento Bayesiano não pode ser utilizado para atualizar sua opinião sobre N, pois neste caso a razão

$$\frac{P(X = \text{urso polar}|N = n)}{P(X = \text{urso polar})}$$

a qual é necessária na obtenção de P^* através do Condicionamento Bayesiano (ver expressão (1.1)) não pode ser determinada. Ou seja, o Condicionamento Bayesiano é fundamentalmente impossível.

A ocorrência de informações julgada improváveis neste caso, provoca apenas um questionamento sobre o modelo assumido e não uma modificação em sua opinião original sobre o número de espécies animais no deserto do Saara.

Suponha que ao invés de um urso polar, na observação feita surge uma espécie completamente nova, sobre a qual nada se sabia e que não constava nos registros de espécies observadas na terra em qualquer época, nem em qualquer outro planeta já visitado pelo Homem - o *Bayesianus Genuinus*. Ao tomar conhecimento da ocorrência desta espécie completamente rara, Você amplia o espaço amostral original Ω_N para um novo espaço $\Omega_N^* = \{0, 1, \ldots, t, t+1\}$ e além disto, esta informação - que se revela surpreendente pois já se julgava conhecer todas as espécies sobre a terra - provoca um questionamento sobre a distribuição *a priori* atribuída a N.

Remetendo-nos novamente à expressão (1.1) verificamos aqui também a impossibilidade do uso do Condicionamento Bayesiano para a obtenção da distribuição *a posteriori* para N. Primeiramente, notamos que a razão

$$\frac{P(X = \textit{Bayesianus Genuinus}|N = n)}{P(X = \textit{Bayesianus Genuinus})}$$

não está bem determinada pois não foram atribuídas probabilidades *a priori* para eventos que envolviam a ocorrência da espécie *Bayesianus Genuinus*. Isto impede a determinação das probabilidades *a posteriori* para eventos $\{N = n\}$ mesmo quando estes estavam previamente definidos no espaço amostral. Além deste fato, percebemos que a impossibilidade de atualizar-se as probabilidades iniciais através do Condicionamento Bayesiano surge também por desconhecermos a distribuição *a priori* para o evento $\{N = t+1\}$, o que sugere uma reavaliação completa das distribuições de probabilidades inicialmente declaradas.

Uma sugestão usualmente dada para solucionar o problema de atualização de probabilidades no último caso abordado acima é considerar desde o início um espaço amostral suficientemente grande de tal forma que a ocorrência de imprevistos (do tipo ocorrências de espécies desconhecidas) fosse evitada. Com este tratamento, as atualizações de probabilidades poderiam ser feitas através do Condicionamento Bayesiano, pois não haveriam modificações em Ω e nem um questionamento sobre a distribuição *a priori* adotada para $P\{N = n\}$. Esta proposta de solução é apresentada por Zabell(1992).

Para o caso do urso polar, como o Condicionamento Bayesiano não é fundamentalmente possível, poderíamos pensar, por exemplo, em uma retaxação completa de probabilidades para solucionar o problema ou em alguma maneira alternativa para atualizar probabilidades (ver seção 1.4). Note que esta situação reforça a sugestão geralmente feita a um subjetivista de não violar a Lei de Cromwell (ver seção 2.2.2-cap.2 e O'Hagan(1994)).

Uma outra situação em que o uso do Condicionamento Bayesiano é desprovido de sentido foi descrita por Jeffrey(1965) através do seguinte exemplo.

Exemplo 1.1: *(observação na penumbra)*

Você esqueceu a cor de seu pijama, mas lembra-se que pode ser apenas Verde, Azul ou Lilás com as seguintes probabilidades:

$$\begin{aligned} P(\text{Verde}) &= 0,3 \\ P(\text{Azul}) &= 0,4 \\ P(\text{Lilás}) &= 0,3. \end{aligned}$$

Um carro passa com o farol ligado iluminando parcialmente o quarto onde Você está. Depois de observar seu pijama na penumbra gerada pela luz do carro, Você continua incerto sobre a cor real do pijama, mas tem sua opinião modificada para a seguinte distribuição de probabilidade

$$\begin{aligned} P^*(\text{Verde}) &= 7/35 \\ P^*(\text{Azul}) &= 16/35 \\ P^*(\text{Lilás}) &= 12/35. \end{aligned}$$

Ao observar o pijama na penumbra algo acontece em sua retina levando Você a acreditar um pouco mais na cor Azul do que nas demais. Este "algo" que provoca sua mudança de opinião sobre os eventos é uma experiência que Você muito dificilmente conseguirá explicar ou exprimir linguisticamente de maneira precisa (como num evento) de tal forma que seja possível justificar o Condicionamento Bayesiano como o mecanismo que gerou a distribuição *a posteriori* P^*.

Contrariamente ao que ocorria no problema de estimação do número de espécies, no exemplo anterior o espaço amostral permanece inalterado após a observação ser feita e não existem eventos inicialmente julgados improváveis. Apesar disto, sua opinião inicial é completamente modificada.

Geralmente, quando a informação é recebida através de alguma experiência sensorial - paladar, visão, audição, etc - dificilmente Você conseguirá definir o que exatamente deve ser observado para que Você passe a acreditar mais em uma das possibilidades que em outras. A grosso modo, nestas situações Você não consegue deixar pronta uma receita de atualização de probabilidades para que uma outra pessoa possa escrever posteriormente em um papel o que seria a sua distribuição *a posteriori*, como permite o Condicionamento Bayesiano.

Um dos aspectos que abordaremos neste trabalho é o como incorporar em sua distribuição de probabilidade *a priori* aquelas informações (obtidas sensorialmente ou não) que modificam o sua opinião inicial, mantendo inalterado o espaço amostral previamente determinado, como foi ilustrado no exemplo anterior.

Na próxima seção apresentaremos o Condicionamento Bayesiano discutindo suas limitações como mecanismo de atualização de probabilidades em situações similares à descrita acima.

1.3 O Condicionamento Bayesiano

Bruno de Finetti(1931,1937) partindo da definição de coerência (por ele introduzida com o intuito de traduzir matematicamente a noção de Racionalidade) estabelece um teorema

garantindo que opiniões sobre quantidades aleatórias, declaradas à luz de coerência, são probabilidades. Isto é, admitindo que Você expressa sua opinião sobre uma coleção de objetos desconhecidos numericamente, esta coleção de números - se coerente - satisfaz os Axiomas de Probabilidades e a definição de probabilidade condicional estabelecidos por Kolmogorov (ver também Loschi e Wechsler(1997)). Por esta razão, a fórmula de Bayes, que é conseqüência direta destas propriedades, também é uma conseqüência de coerência.

Também pode ser visto que, ao longo de sua vasta obra (por exemplo, de Finetti(1974, 1975, 1989)), de Finetti defendia que **toda** probabilidade é condicional. É condicional em todo o conhecimento acumulado por Você até o momento em que se torna inevitável Você declarar suas probabilidades. Vista assim, a definição de coerência que sustenta todo o cálculo de probabilidades possui apenas valor instantâneo, levando-nos a pensar que suas declarações de probabilidades - inclusive aquelas fornecidas pela fórmula de Bayes - são válidas naquele instante de tempo em que foram declaradas. De fato, da construção de probabilidades apresentada em de Finetti(1931) podemos concluir apenas que a fórmula de Bayes traduz sua incerteza sobre eventos condicionais da forma $A|E$ revelando simplesmente a sua opinião (coerente) para a ocorrência de A **supondo** conhecida a verdade sobre o evento E, levada em conta a totalidade de informação conhecida por Você até aquele instante de declarar suas opiniões. Citando de Finetti(1972- pg.193):

" If we speak of a conditional probability $P(A|E)$ ... the expression has meaning if and only if A and E are events. Also E must express the assumed information in its totality. ... "assumed" is suggestive of $P(A|E)$ as the probability that we would regard as fair for a bet on A to be immediately, but to become operative only if E occurs ..."

Por outro lado, ao falarmos de distribuição *a posteriori* referimo-nos àquelas declarações de probabilidades feitas **após** serem conhecidas as informações que influenciaram em seu estado de conhecimento - ou seja, as declarações *a priori* e *a posteriori* são feitas em estágios diferentes de conhecimento. Parece claro então que a fórmula de Bayes *per se* NÃO é um mecanismo de atualização de probabilidades, apesar de seu nome ser utilizado indevidamente como sinônimo de atualização de probabilidades feitas através do Princípio de Condicionalização Bayesiana (brevemente, Condicionamento Bayesiano) definido em Jeffrey(1965), por

exemplo, e que será apresentado a seguir no teorema 1.3.1. Hacking(1967) também discute esta questão, alegando haver diferenças entre probabilidade dado fatos e probabilidade condicional. Em outras palavras, entre supor e saber.

Sejam P e P^* medidas de probabilidade definidas sobre um espaço mensurável$(\Omega, \mathcal{A})$ onde Ω é um conjunto enumerável e $\mathcal{A}$ é a σ-álgebra de interesse. Suponha que P^* e P representam suas opiniões *a posteriori* e *a priori* sobre eventos de Ω em $\mathcal{A}$, respectivamente.

Definição 1.3.1 *(Condicionamento Bayesiano)*

Dizemos que uma atualização de probabilidades é feita através de Condicionamento Bayesiano sempre que a distribuição de probabilidades a posteriori P^* for obtida partindo-se da distribuição de probabilidade a priori P da seguinte forma

$$P^*(.) = P(.|E) \tag{1.2}$$

onde E é algum evento $\mathcal{A}$-mensurável tal que $P(E) > 0$.

Esta definição, que pode ser encontrada em Jeffrey(1965 - pg 80) entre outros, diz o que é a atualização de uma medida de probabilidade por Condicionamento Bayesiano mas não estabelece as condições sob as quais o uso deste procedimento de atualização de probabilidades faz sentido. Estas condições poderão ser vistas mais adiante no teorema 1.3.1.

Algumas pessoas advogam para o Condicionamento Bayesiano o adjetivo de único procedimento coerente para atualizar probabilidades. Argumentos favoráveis a isto podem ser encontrados em Skyrms(1987), Lewis(1975) e em Teller(1973). No entanto, percebemos da definição acima que uma das exigências feitas para que o Condicionamento Bayesiano seja julgado como um procedimento possível para atualizar-se probabilidades é a existência, no espaço amostral, de um evento condicionante, bem definido e cuja ocorrência é inicialmente julgada possível (Howson (1996) e Howson e Urbach(1993)). Além disto, via de regra, a este evento condicionante deve ser atribuída uma probabilidade inicial menor que 1. Como ilustra o exemplo da observação na penumbra visto na seção anterior, nem sempre existe tal evento condicionante. Muitas vezes recebemos informações que provocam mudanças em

nossas opiniões sem que com isto saibamos definir exatamente o evento que provocou tal mudança. Ou seja, Você pode atualizar suas probabilidades por procedimentos outros distintos de condicionamento.

Jeffrey(1992), por exemplo, estabelece o seguinte resultado fornecendo condições sob as quais o Condicionamento Bayesiano pode ser aceitável como mecanismo de obtenção da distribuição *a posteriori*.

Considere P e P^* como anteriormente definidas. Aqui e ao longo de todo este trabalho, denote por $\bar{E}$ o evento complementar de E com respeito a Ω e por (A, E) o evento interseção $A \cap E$.

Teorema 1.3.1 *Seja $E \in \mathcal{A}$ um evento tal que $P(E) > 0$. Então, para todo $A \in \mathcal{A}$*

$$\text{(i)} \quad P^*(E) = 1 \text{ e} \tag{1.3}$$

$$\text{(ii)} \quad P^*(A|E) = P(A|E) \tag{1.4}$$

se, e somente se,

$$P^*(A) = P(A|E) \tag{1.5}$$

Prova :Suponha válidas as condições (i) e (ii) e seja A um evento qualquer de Ω. Então, usando propriedades de cálculo de probabilidades, obtemos:

$$\begin{aligned} P^*(A) &= P^*(A|E)P^*(E) + P^*(A|\bar{E})P^*(\bar{E}) \\ &= P(A|E)P^*(E) + P(A|\bar{E})(1 - P^*(E)) \\ &= P(A|E) \end{aligned}$$

Por outro lado, se $P^*(A) = P(A|E)$, $\forall\ A \in \mathcal{A}$ e se $P(E) > 0$, em particular, temos que para $A = E$,

$$P^*(E) = P(E|E) = 1 \tag{1.6}$$

e além disso, para todo $A \in \mathcal{A}$

$$P^*(A|E) = \frac{P^*(A, E)}{P^*(E)}$$

$$
\begin{aligned}
&= \frac{P(A, E|E)}{P(E|E)} \\
&= P(A|E),
\end{aligned}
$$

Concluindo a prova do teorema.

□

Perceba que o Condicionamento Bayesiano é um mecanismo que autoriza Você a atualizar suas opiniões iniciais usando a fórmula de Bayes, desde que agregados a sua vontade de utilizá-lo como mecanismo de atualização de probabilidades estejam as condições (i) e (ii) do teorema anterior.

O resultado apresentado no teorema 1.3.1 é chamado por Howson e Urbach(1993) de **Princípio de Condicionalização Bayesiana** e as condições (i) e (ii) são também conhecidas por **certeza** e **suficiência**, respectivamente. Usando esta terminologia podemos dizer que o Condicionamento Bayesiano é um procedimento aceitável para atualizar probabilidades se a informação recebida levar Você a mover-se de um estado inicial de dúvida sobre E para a certeza *a posteriori* de sua ocorrência e se além disso a partição $\{E, \bar{E}\}$ contiver toda a informação relevante para a construção de sua distribuição *a posteriori* - isto é, $\{E, \bar{E}\}$ é uma partição suficiente para a família de distribuições de probabilidade $\{P, P^*\}$.

Em uma linguagem mais corrente, podemos dizer que uma das exigências feitas para que seja possível obter distribuições *a posteriori* via Condicionamento Bayesiano é que toda a informação relevante aprendida por Você antes da declaração de sua opinião *a posteriori* é a verdade de E (além da manutenção das probabilidades condicionais). Isto sugere que o uso do Condicionamento Bayesiano como mecanismo de atualização de probabilidades é razoável nas situações em que toda a informação a ser recebida é desde o início considerada no espaço amostral, como algumas vezes temos em experimentos planejados onde toda a informação a ser utilizada em suas inferências vem do experimento realizado e nenhuma surpresa ocorre.

Sendo assim, o uso do Condicionamento Bayesiano é inadequado para incorporar informações imprevistas, e mesmo aquelas informações não traduzíveis como eventos, a menos que estas conduzam à certeza sobre o evento condicionante.

Particularmente, denotanto por supp(.) o suporte da da distribuição desejada e considerando Ω é enumerável temos os seguintes resultados.

Proposição 1.3.1 *Se* $\text{supp}(P^*) = \text{supp}(P)$, *então* P^* *é obtida de* P *por Condicionamento Bayesiano se, e somente se,* $P^* = P$.

Prova: Suponha que P^* é obtida de P por Condicionamento Bayesiano. Então, existe um evento $E \in \mathcal{A}$ tal que $P(E) > 0$, e $P^*(A) = P(A|E)$ para todo $A \in \mathcal{A}$.

Como $\text{supp}(P^*) = \text{supp}(P)$, o evento E é da forma $\text{supp}(P) \cup B$, onde B é um evento $\mathcal{A}$ - mensurável com $P(B) = 0$. Então, para todo $w \in \Omega$ temos

$$P^*(\{w\}) = P(\{w\}|E) = \frac{P(\{w\})}{P(E)} = P(\{w\}).$$

Por outro lado, suponha que $P^* = P$. Considere $E = \text{supp}(P^*)$. Neste caso, temos que $P^*(E) = 1$ e para todo $A \in \mathcal{A}$, $P^*(A|E) = P(A|E)$. Sendo assim, como Ω é enumerável, pelo teorema 1.3.1, concluímos que P^* é obtida de P por Condicionamento Bayesiano.

□

Outro resultado interessante que obtivemos é o seguinte:

Proposição 1.3.2 *Se* $\text{supp}(P) \subset \text{supp}(P^*)$ *e* Ω *é enumerável, então* P^* *não é obtida de* P *por Condicionamento Bayesiano.*

Prova: Como, por hipótese, $\text{supp}(P) \subset \text{supp}(P^*)$ existe $w_o \in \Omega$ tal que $P(\{w_o\}) = 0$ e $P^*(\{w_o\}) = p \neq 0$. Suponha, por absurdo, que P^* é obtida de P por Condicionamento Bayesiano. Neste caso, para todo $w \in \Omega$ teríamos

$$P^*(\{w\}) = P(\{w\}|E)$$

para algum $E \in \mathcal{A}$ tal que $P^*(E) = 1$ e $P(E) > 0$.

Considere $w = w_o$. Então,

$$p = P^*(\{w_o\}) = P(\{w_o\}|E) \leq \frac{P(\{w_o\})}{P(E)} = 0,$$

o que uma contradição, pois $p \neq 0$. Logo, P^* não é obtida de P por Condicionamento Bayesiano.

□

Conjectura-se que a estensão dos resultados anteriores para Ω não enumerável não seja muito dificil.

Perceba, das proposições anteriores, que no exemplo 1.1 o Condicionamento Bayesiano somente poderia ter sido o mecanismo utilizado para a construção da distribuição *a posteriori* se, ao observar a peça de roupa na penumbra, Você mantivesse sua opinião inicial, ou se ficasse convencido de que uma das cores era impossível. (No último caso, é necessário ainda impor a condição de suficiência sobre a partição gerada pelo evento condicionante.)

A seguir apresentaremos algumas definições mais gerais de condicionamento, discutindo o Supercondicionamento ou Condicionamento Bayesiano Generalizado definido por Diaconis e Zabell(1982) e sugerimos uma outra maneira de atualizar probabilidades - a Regra Geral de Condicionamento. Também apresentaremos alguns resultados aparentemente ainda não examinados na literatura para a validade de cada uma destas formas de condicionamento.

1.4 Supercondicionamento e Regra Geral de Condicionamento

Considerando, como no contexto Bayesiano paramétrico, $\Omega = \mathcal{X} \times \Theta$ onde $\mathcal{X}$ é o espaço amostral e Θ é o espaço paramétrico, ambos supostamente enumeráveis, percebemos que o Condicionamento Bayesiano nos fornece a seguinte relação entre a distribuição *a posteriori* P^* e a distribuição *a priori* P

$$P^*(\theta) \ \alpha \ P(x|\theta)P(\theta) = l(x,\theta)P(\theta).$$

onde l é usualmente chamada de função de verossimilhança gerada por x para θ.

Ou seja, ao considerarmos o Condicionamento Bayesiano usual visto na seção anterior estamos supondo que as atualizações de probabilidade são feitas pelo produto entre a distribuição *a priori* e uma função de θ e x ambas definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A})$ onde $\mathcal{A}$ representa uma σ-álgebra de eventos de Ω.

Estendendo esta idéia para situações mais gerais, Diaconis e Zabell(1982) definem o Condicionamento Bayesiano Generalizado ou Supercondicionamento (nome dado em Jeffrey(1992)) estabelecendo que a distribuição *a posteriori* é obtida através do produto da distribuição *a priori* (definida sobre $(\Omega, \mathcal{A})$) por uma função de verossimilhança definida sobre um espaço diferente do original.

Formalmente, considerando Ω enumerável, $\mathcal{A}$ a σ-álgebra de interesse e P e P^* duas medidas de probabilidades definidas sobre $(\Omega, \mathcal{A})$ representando suas opiniões *a priori* e *a posteriori*, respectivamente, Diaconis e Zabell(1982) definem:

Definição 1.4.1 *(Condicionamento Bayesiano Generalizado ou Supercondicionamento)*
Dizemos que P^* é obtida de P por Supercondicionamento ou Condicionamento Bayesiano Generalizado se existem um espaço de probabilidades $(\bar{\Omega}, \bar{\mathcal{A}}, Q)$ e uma classe $\mathcal{D} = \{E_w\bar{\mathcal{A}},\ w \in \Omega\}$, satisfazendo:

(i) $E_w = 1$ se, e somente se $\{w\} = 1$,

(ii) $Q(E_w) = P(\{w\})$, $\forall w \in \Omega$ e

(iii) $P^*(\{w\}) = Q(E_w|E)$ para todo $w \in \Omega$ e para um evento $E \in \bar{\mathcal{A}}$ tal que $Q(E) > 0$.

Apesar de $\bar{\Omega}$ ser um espaço diferente de Ω, $\bar{\Omega}$ não está completamente desvinculado de Ω uma vez que a ocorrência de um evento E_w de $\bar{\Omega}$ é determinada pela ocorrência de um átomo $\{w\} \in \Omega$. Além disto, considerando (ii) e (iii) da definição acima e denotando por w o evento atômico $\{w\}$ temos que para qualquer $w \in \Omega$

$$P^*(w) \;=\; Q(E_w|E) = \frac{Q(E|E_w)Q(E_w)}{Q(E)}$$

$$= \frac{Q(E|E_w)P(w)}{Q(E)} \ \alpha \ Q(E|E_w)P(w), \tag{1.7}$$

onde $Q(E|E_w)$ faz o papel de uma função de verossimilhança para w gerada pelo conhecimento de E, definida em $(\bar{\Omega}, \bar{\mathcal{A}}, Q)$ (pois é uma função de w e é a probabilidade condicional em E_w de E).

Da expressão (1.7) podemos observar que esse procedimento permite atualizar a distribuição *a priori* nas situações onde o evento observado tem probabilidade inicial zero, como é o caso da ocorrência de um urso polar no exemplo da estimação do número de espécies visto na seção 2. Isto é possível pois neste tipo de condicionamento é permitido utilizar uma nova função de verossimilhança definida sobre um novo espaço de probabilidade onde ao evento condicionante tenha sido atribuída uma probabilidade inicialmente não-nula, ou seja, por permitir uma reavaliação arbitrária da função de verossimilhança inicialmente eleita por Você para descrever o comportamento dos dados.

Considerando a definição 1.4.1 Diaconis e Zabell(1982) apresentam o seguinte resultado estabelecendo quando é matematicamente possível atualizar probabilidades via Supercondicionamento.

Teorema 1.4.1 *Sejam P e P^* medidas de probabilidades definidas sobre $(\Omega, \mathcal{A})$ onde Ω é um conjunto enumerável. P^* é obtida de P através de um supercondicionamento se, e somente se, existe uma constante $B \geq 1$ tal que*

$$P^*(w) \leq BP(w), \quad \forall \ w \in \Omega. \tag{1.8}$$

Prova : Inicialmente, suponha que P^* é obtida de P através de um Supercondicionamento. Considere $w \in \Omega$ qualquer. Então pela definição anterior

$$\begin{aligned} P^*(w) &= Q(E_w|E) \leq \frac{Q(E_w)}{Q(E)} \\ &= \frac{1}{Q(E)} P(w), \end{aligned} \tag{1.9}$$

para alguma medida Q definida em $(\bar{\Omega}, \bar{\mathcal{A}})$ satisfazendo as condições da definição 1.4.1.

Tome $B = \frac{1}{Q(E)}$. Como $Q(E) \in (0,1]$, $B \geq 1$.

Por outro lado, considere primeiramente $B = 1$ e suponha que exista um único w_o tal que

$$P^*(w_o) < P(w_o). \tag{1.10}$$

Então, como $P^*(w) = P(w)$ para todo $w \neq w_o$ temos

$$\sum_{w\in\Omega} P^*(w) < \sum_{w\in\Omega} P(w) \Rightarrow 1 < 1, \tag{1.11}$$

chegando a uma contradição. Neste caso, temos $P^* = P$.

Considere agora $B > 1$ e defina a medida de probabilidade

$$P^{**}(w) = \frac{B}{(B-1)}P(w) - \frac{1}{(B-1)}P^*(w), \quad \forall \ w \in \Omega. \tag{1.12}$$

Denote por (x, y) o produto cartesiano dos eventos atômicos $\{x\}$ e $\{y\}$. Então, considerando $\bar{\Omega} = \Omega \times \{a, b\}$, $E_w = (w, a) \cup (w, b)$, $E = \cup_w (w, a)$ e uma medida de probabilidade Q definida sobre a σ - álgebra de eventos de $\bar{\Omega}$ tal que

$$\begin{aligned} Q(w,a) &= \frac{1}{B}P^*(w) \\ Q(w,b) &= \left(1 - \frac{1}{B}\right) P^*(w), \end{aligned}$$

temos que

$$\begin{aligned} Q(E_w) &= Q\left((w,a) \cup (w,b)\right) = Q(w,a) + Q(w,b) \\ &= \frac{1}{B}P^*(w) + \frac{1}{(B-1)}P^{**}(w) = P(w) \end{aligned}$$

e, para todo $w_o \in \Omega$

$$\begin{aligned} Q(E_{w_o}|E) &= \frac{Q(E_{w_o}, E)}{Q(E)} = \frac{Q(w_o, a)}{\sum_{w\in\Omega} Q(w,a)} \\ &= \frac{BP^*(w_o)}{B\sum_{w\in\Omega} P^*(w)} = P^*(w_o). \end{aligned}$$

Portanto, P^* é obtida de P por Supercondicionamento.

□

No teorema anterior são estabelecidas apenas condições matemáticas para sugerir que uma atualização seja feita via Supercondicionamento.

Diaconis e Zabell(1982) comentam que se Ω é um conjunto não - enumerável, a condição para que P^* seja obtida de P por Supercondicionamento é que P^* seja dominada por P - isto é, P^* deve ser nula para todo valor w tal que $P(w) = 0$ - e $\frac{dP^*}{dP} \in L_\infty$.

Como corolário do teorema anterior surge naturalmente que:

Corolário 1.4.1 *Se P^* é obtida de P por Condicionamento Bayesiano, então P^* é obtida de P através de Supercondicionamento.*

Prova : Basta considerar $B = \frac{1}{P(E)} + 1$ onde E é o evento condicionante.

□

Perceba do teorema 1.4.1 que se Ω é um conjunto finito com $\text{supp}(P^*) \subseteq \text{supp}(P)$, P^* sempre será obtida de P por Supercondicionamento, pois neste caso sempre existirá uma cota finita $B > 1$ para a razão $P^*(.)/P(.)$. Este fato permite-nos afirmar que a recíproca do corolário anterior não é verdadeira, como pode ser visto no seguinte exemplo.

Exemplo 1.2: Considere novamente o exemplo da penumbra, visto na seção 1.2.

Como Ω é finito, o teorema 1.4.1 garante que a distribuição *a posteriori* fornecida no exemplo 1.1 é sempre obtida da distribuição inicial por Supercondicionamento.

Agora, pela proposição 1.3.1 verificamos que a distribuição P^* fornecida no exemplo 1.1 não poderia ter sido obtida de P por Condicionamento Bayesiano pois neste caso, Ω é o único evento tal que a probabilidade *a posteriori* de ocorrência é um, o que conduz a

$$P^*(\text{azul}|\Omega) = \frac{16}{35} \neq 0.4 = P(\text{azul}|\Omega).$$

Particularmente, obtemos condições para a validade do Supercondicionamento quando P e P^* são ambas medidas discretas e membros da família exponencial uni-paramétrica, as quais serão exibidas na proposição 1.4.1 a seguir.

Dizemos que uma medida de probabilidade P é membro da família exponencial uni-paramétrica se sua função de probabilidade (ou densidade) p assume a seguinte forma:

$$p(w) = a(w)b(\theta)\exp\{f(w)c(\theta)\} \tag{1.13}$$

para todo $w \in \Omega$ e para escolhas convenientes das funções reais a, b, c e f, onde a e b são não-negativas (Casella e Berger(1990)).

Sejam P e P^* medidas de probabilidades definidas sobre $\Omega = \{0, 1, 2, \ldots\}$ pertencentes à família exponencial uni-paramétrica e sejam

$$p(w) = a(w)b(\theta)\exp\{f(w)c(\theta)\} \quad \text{e} \quad p^*(w) = a^*(w)b^*(\theta^*)\exp\{f(w)c^*(\theta^*)\} \tag{1.14}$$

suas respectivas funções de probabilidades.

Proposição 1.4.1 *Sejam $\Omega = \{0, 1, \ldots\}$, $\mathcal{A} = 2^{\Omega}$, P eP^* medidas de probabilidades definidas sobre $(\Omega, \mathcal{A})$, com função de probabilidade p e p^*, respectivamente, dadas na expressão (1.14). Isto é, P e P^* pertencem à família exponencial uni-paramétrica com mesma estatística f. P^* é obtida de P por Supercondicionamento quando*

(i) $\text{supp}(P^*) \subseteq \text{supp}(P)$;

(ii) $f(w) \geq 0, \forall\, w \in \Omega$;

(iii) $c^*(\theta^*) \leq c(\theta) + \inf_{\mathcal{C}} \left\{ \frac{1}{f(w)} \log \frac{a(w)}{a^*(w)} \right\}$, *onde* $\mathcal{C} = \{w \in \text{supp}(P^*) \ : \ f(w) > 0\}$;

(iv) $0 < \frac{a^*(w)}{a(w)} \leq 1$, *para todo* $w \in \text{supp}(P^*)$ *tal que* $f(w) = 0$.

Prova : Suponha válidas as condições sobre p^* e p. Sob a suposição (i) temos que

$$\frac{p^*(w)}{p(w)} = \begin{cases} \frac{b^*(\theta^*)}{b(\theta)} \frac{a^*(w)}{a(w)} \exp\{f(w)[c^*(\theta^*) - c(\theta)]\} & \text{se } w \in \text{supp}(P^*) \\ 0 & \text{caso contrário.} \end{cases}$$

Faça $l^*(w,\theta,\theta^*) = \frac{a^*(w)}{a(w)} \exp\{f(w)[c^*(\theta^*) - c(\theta)]\}$.

Considere inicialmente $w \in f^{-1}(\{0\}) \cap \text{supp}(P^*)$. Da hipótese (iv) temos que,

$$0 < \frac{a^*(w)}{a(w)} \leq 1 \Longleftrightarrow 0 < l^*(w,\theta,\theta^*) \leq 1. \tag{1.15}$$

Então, $\frac{p^*(w)}{p(w)} < \frac{b^*(\theta^*)}{b(\theta)}$ para quaisquer valores $c^*(\theta^*)$ e $c(\theta)$.

Por outro lado, para todo $w \in \text{supp}(P^*)$ tal que $f(w) > 0$ temos pela hipótese (iii) que

$$\begin{aligned}
-\infty < c^*(\theta^*) \quad &\leq \quad c(\theta) + \inf_{\mathcal{C}} \left\{ \frac{1}{f(w)} \log \frac{a(w)}{a^*(w)} \right\} \\
&\Longleftrightarrow \quad -\infty < c^*(\theta^*) - c(\theta) \leq \frac{1}{f(w)} \log \frac{a(w)}{a^*(w)}, \quad \forall w \in \mathcal{C} \\
&\Longleftrightarrow \quad -\infty < \log\{\exp\{f(w)[c^*(\theta^*) - c(\theta)]\}\} \leq \log \frac{a(w)}{a^*(w)}, \ \forall w \in \mathcal{C} \\
&\Longleftrightarrow \quad 0 < \frac{a^*(w)}{a(w)} \exp\{f(w)[c^*(\theta^*) - c(\theta)]\} \leq 1, \ \forall w \in \mathcal{C} \\
&\Longleftrightarrow \quad 0 < l^*(w,\theta,\theta^*) \leq 1 \\
&\Rightarrow \quad \frac{p^*(w)}{p(w)} < \frac{b^*(\theta^*)}{b(\theta)}, \ \forall w \in \mathcal{C}.
\end{aligned}$$

Portanto, sob as hipóteses, existe $B = b^*(\theta^*)/b(\theta) + 1$ tal que

$$p^*(w) \leq Bp(w). \tag{1.16}$$

Logo, pelo teorema 1.4.1, P^* é obtida de P por Supercondicionamento.

□

Um resultado análogo pode ser obtido quando $f(w) \leq 0$ para todo $w \in \Omega$. Neste caso, basta substituir a condição (iii) da proposição acima pela seguinte:

$$c^*(\theta^*) \leq c(\theta) + \sup_{\mathcal{C}} \left\{ \frac{1}{f(w)} \log \frac{a(w)}{a^*(w)} \right\},$$

onde $\mathcal{C} = \{w \in \text{supp}P^* : f(w) < 0\}$.

Exemplo 1.3: Seja $\Omega = \{0, 1, 2, \ldots\}$. Suponha que sua opinião *a priori* sobre eventos de Ω seja Geométrica com parâmetro $\lambda > 0$ e que sua opinião *a posteriori* seja Geométrica com parâmetro $\alpha > 0$. Neste caso, verificamos que $\frac{a^*(w)}{a(w)} = 1$ para todo $w \in \Omega$. Então uma vez que $f(w) = w \geq 0$, para todo $w \geq 0$, P^* é obtida de P por Supercondicionamento se

$$\log(1-\alpha) \leq \log(1-\lambda) + \inf_{w \in \Omega - \{0\}} \left\{ \frac{1}{w} \right\}.$$

Isto é, se $\alpha > \lambda$ vale o Supercondicionamento.

Observe que o Supercondicionamento não poderá ser justificado como mecanismo de atualização de probabilidades em várias situações. Por exemplo, jamais conseguiremos justificar a passagem de uma Binomial para uma Poisson por Supercondicionamento. Neste caso, a σ-álgebra sobre a qual está definida a distribuição *a posteriori* não está contida na σ- álgebra sobre a qual está definida a distribuição *a priori*. Também não é possível justificar matematicamente a passagem de uma Geométrica com parâmetro $1/2$ para uma Geométrica com parâmetro $1/3$ por Supercondicionamento, apesar de ambas possuirem o mesmo suporte.

Problemas como o da geométrica apontado acima poderão ser solucionados se permitirmos uma definição um pouco mais ampla de condicionamento que chamaremos de Regra Geral de Condicionamento.

Definição 1.4.2 *(Regra Geral de Condicionamento)*

Sejam P e P^* medidas de probabilidade definidas sobre $(\Omega, \mathcal{A})$ e Ω é um conjunto enumerável. Diz-se que P^* é obtida da P através da Regra Geral de Condicionamento se existirem

(i) um espaço mensurável $(\bar{\Omega}, \bar{\mathcal{A}})$ e

(ii) um conjunto $\mathcal{D} = \{E_w \in \bar{\mathcal{A}} \ : \ w \in \Omega\}$, com $E_w = 1$ se, e somente se $\{w\} = 1$ e se existir

(iii) uma função real e não-negativa g com domínio $\mathcal{D}$ satisfazendo a condição

$$\sum_{w \in \Omega} g(E_w)P(w) < \infty,$$

de tal forma que

$$P^*(w) \;\; \alpha \;\; g(E_w)P(w), \quad \forall w \in \Omega.$$

A seguir, fornecemos a condição para que a Regra Geral de Condicionamento possa ser utilizado como procedimento de atualização de probabilidade.

Proposição 1.4.2 *Considere P e P^* medidas de probabilidade definidas sobre $(\Omega, \mathcal{A})$, onde Ω é um conjunto enumerável. P^* será obtida de P através da Regra Geral de Condicionamento, se $supp(P^*) \subseteq supp(P)$.*

Prova: Suponha $\bar{\Omega} = \Omega$ e considere $\mathcal{D} = \{\{w\} : w \in \Omega\}$. Tome

$$g(E_w) = \begin{cases} \frac{p^*(w)}{p(w)} & \text{se } w \in \text{supp}(P^*) \\ 0 & \text{caso contrário.} \end{cases}$$

Como P e P^* são medidas de probabilidade, g é uma função real e não-negativa e pode-se facilmente verificar que

$$p^*(w) \;\; \alpha \;\; g(E_w)p(w), \quad \forall w \in \Omega.$$

Além disso, observa-se que

$$\begin{aligned} \sum_{w \in \Omega} g(E_w)p(w) &= \sum_{w \in \text{supp}(P^*)} p^*(w) \\ &= 1 < \infty, \end{aligned}$$

concluindo a prova.

□

Exemplo 1.4: Suponha que sua opinião *a priori* P e *a posteriori* P^* sobre eventos de $\Omega = \{0, 1, 2, \ldots\}$ sejam respectivamente Geométrica com parâmetros λ e α, onde $\alpha < \lambda$. Então, pelo resultado anterior, P^* é obtida de P pela Regra Geral de Condicionamento.

Na situação abordada no exemplo anterior o Supercondicionamento não permitia a atualização de probabilidades (ver exemplo 1.3). De fato, a Regra Geral de Condicionamento é mais abrangente que o Supercondicionamento, como pode ser visto no seguinte resultado.

Proposição 1.4.3 *Se P^* é obtida de P por Supercondicionamento, então P^* é sempre obtida de P pela Regra Geral de Condicionamento.*

Prova: Tome $g(E_w) = Q(E|E_w)$, para todo $w \in \Omega$. Lembrando que $P(w) = Q(E_w)$ temos que

$$\begin{aligned} \sum_{w \in \Omega} g(E_w)P(w) &= \sum_{w \in \Omega} Q(E, E_w) \\ &= Q(\cup_w (E, E_w)) < \infty. \end{aligned}$$

É imediato verificar que

$$P^*(w) \ \alpha \ g(E_w)P(w), \quad \forall w \in \Omega.$$

□

Apesar de mais flexível que o Supercondicionamento, a Regra Geral de Condicionamento não permite atualizar probabilidades quando $\text{supp}(P) \subset \text{supp}(P^*)$. Para estes casos, ainda não temos uma solução. Nunca haverá uma solução geral (Dawid(1985)).

Na próxima seção veremos uma nova explicação para o Paradoxo dos Três Prisioneiros dentro do contexto explorado neste trabalho, fazendo o uso do Condicionamento Bayesiano. Este problema, da forma que será considerada aqui, também servirá como ilustração da diferença entre os usos da fórmula de Bayes e do Condicionamento Bayesiano.

1.5 Imprevistos e o Paradoxo dos Três Prisioneiros

Nesta seção apresentaremos uma explicação alternativa às usualmente dadas ao Paradoxo dos Três Prisioneiros supondo, como ocorre na literatura, que o prisioneiro A opta por atualizar sua opinião inicial utilizando o Condicionamento Bayesiano.

A idéia é mostrar que não é paradoxal o problema resolvido pelo prisioneiro A se este for considerado dentro do contexto de recebimento de informações imprevistas.

Paradoxo dos Três Prisioneiros: *Três prisioneiros A, B e C estão no cárcere à espera do anúncio de suas sentenças, que será feito amanhã. É sabido por todos, inclusive pelos prisioneiros, que dois deles morrerão e apenas um será libertado mas não se sabe qual deles. O prisioneiro A pede ao carcereiro que lhe diga o nome de um dos outros dois prisioneiros que será condenado, explicando que, como pelo menos um dos outros dois morrerá, o carcereiro não fornecerá qualquer informação relevante sobre sua condição. O carcereiro concorda e diz a ele que B morrerá. A fica feliz por que após esta informação, A verifica que sua probabilidade de ser ele o sobrevivente aumenta de* $1/3$ *para* $1/2$. (Jeffrey(1992)-pg.122).

Como A argumenta, a informação dada pelo carcereiro nada revela a A sobre sua própria condição que não houvesse sido previamente considerado por A no processo de construção de suas probabilidades iniciais. Por este motivo, e supondo que a informação fornecida pelo carcereiro é toda a informação relevante aprendida por A antes de tornar pública sua opinião *a posteriori*, espera-se que a opinião inicial de A sobre o evento " A viverá" não sofra alterações. Suponhamos que esta seja a situação e que A não tenha qualquer evidência que permita atribuir probabilidades distintas de sobrevivência a qualquer um dos prisioneiros.

Qual é o equívoco cometido por A? A explicação usualmente dada a este problema é que A comete um erro na determinação do espaço amostral que considera para suas inferências e por isto, ao atualizar suas probabilidades usando o Condicionamento Bayesiano, condiciona em um evento que não reflete a informação verdadeiramente recebida por A. Isto é, A parte do espaço amostral $\Omega = \{A, B, C\}$ e considerando este espaço, atualiza sua opinião inicial condicionando no evento $E = \{A, C\}$ que traduz que "B morrerá" e não o que é de fato

observado por A que é "o carcereiro informa que B morrerá". (Ver Jeffrey(1992)).

Como desde o início A planeja perguntar ao carcereiro sobre a situação dos outros dois prisioneiros, ele deve considerar para sua análise o espaço amostral

$$\Omega_1 = \{Ab, Ac, Bc, Cb\} \tag{1.17}$$

onde Xy denota que o prisioneiro X viverá e a informação a ser dada pelo carcereiro é que o prisioneiro Y morrerá.

Se parte deste espaço amostral, A pode obter sua opinião *a posteriori* condicionando na informação realmente recebida por ele - lembrando, "o carcereiro informa que B morrerá" - que se traduz no evento $E_1 = \{Ab, Cb\}$ [2] do espaço Ω_1.

Considerando o espaço amostral descrito em (1.17) e supondo que o carcereiro informa que B morrerá, A calcula a sua chance de ser o sobrevivente condicional no evento E_1 da seguinte forma:

$$\begin{aligned} P(Ab \cup Ac | Ab \cup Cb) &= \frac{P(Ab)}{P(Ab \cup Cb)} \\ &= \frac{1/2 \;\; 1/3}{1/2 \;\; 1/3 + 1 \;\; 1/3} = 1/3. \end{aligned} \tag{1.18}$$

Se após consultar o carcereiro, A toma conhecimento de que efetivamente B será condenado (e esta é toda a informação relevante aprendida por A desde a declaração de sua distribuição *a priori* até a eliciação de sua distribuição *a posteriori*) e se A julga que as probabilidades condicionais em E_1 inicialmente declaradas se mantêm *a posteriori*, o Condicionamento Bayesiano visto anteriormente na seção 1.3 autoriza A a explicitar sua opinião *a posteriori* para o evento "A viverá" como sendo a probabilidade obtida na expressão (1.18), concordando com nossa expectativa inicial sobre o que deveria ter sido a distribuição *a posteriori* declarada pelo prisioneiro A.

Note que aqui o evento condicionante "o carcereiro informa que B morrerá" é um evento ao qual inicialmente atribuiu-se probabilidade *a priori* positiva e que após a informação dada pelo carcereiro tornou-se um evento certo, permitindo o uso do Condicionamento Bayesiano.

[2] Perceba que neste novo espaço, o evento E considerado por A torna-se $\{Ab,\ Ac,\ Cb\}$.

Suponha que, contrário ao que descrevemos anteriormente, A não tenha planejado perguntar ao carcereiro sobre a condição de seus companheiros de infortúnio. Neste caso, o espaço amostral adequado para descrever o experimento realizado por A é

$$\Omega = \{A, B, C\} \tag{1.19}$$

onde X representa o prisioneiro X viverá.

Como antes, consideremos que não existem evidências que favoreçam a qualquer dos prisioneiros de tal forma que para A, inicialmente, cada prisioneiro tem chance 1/3 de ser o prisioneiro que viverá. Neste mesmo instante, A pode também declarar a probabilidade de ser ele o sobrevivente supondo que B morrerá, por exemplo, a qual é dada da seguinte forma

$$\begin{aligned} P^*(A) &= P(A|A \cup C) \\ &= \frac{P(A)}{P(A \cup C)} \\ &= 1/2 \end{aligned} \tag{1.20}$$

Após declarar sua chance de ser o sobrevivente, A fica desolado e começa a pensar no seu último desejo antes de morrer quando ouve o juiz responsável pelo caso dizer ao carcereiro que B morrerá. Ao ouvir isto, A resolve reavaliar suas probabilidades iniciais utilizando a informação recebida que é traduzida no evento $E = \{A \cup C\}$. *A posteriori*, A julga que E é um evento certo e que para ele todas as probabilidades condicionais neste evento não se alteram. Sendo assim, pelo teorema 1.3.1 a probabilidade *a posteriori* para o evento " A sobreviverá" é dada pela expressão 1.20, confirmando as afirmações iniciais do prisioneiro.

Nas duas situações a informação recebida é a mesma, no entanto as probabilidades *a posteriori* se modificam por que no segundo caso a informação foi recebida inesperadamente, justificando a mudança de expectativa de A com respeito a sua chance de sobreviver.

Observe que a maneira como a informação é recebida influencia no cálculo da distribuição *a posteriori* do evento " A viverá". No caso em que a informação recebida por A não é previamente programada fica perfeitamente justificável, pelos argumentos anteriormente dados neste capítulo, uma mudança no valor da probabilidade *a posteriori* declarada por A

sobre o evento " A viverá", não sendo contraditório o julgamento feito pelo prisioneiro A . (Ver também Pereira(1995)).

No próximo capítulo apresentaremos outro procedimento de atualização de probabilidade, o qual é conhecido na literatura por Regra de Jeffrey. Além de aspectos fundamentais que envolvem esta regra, também apresentaremos alguns de seus aspectos matemáticos.

Capítulo 2

A Regra de Jeffrey

2.1 Introdução

Como vimos no capítulo anterior, o Condicionamento Bayesiano usual permite mudanças exógenas de probabilidades apenas em casos extremos (como visto, por exemplo, na seção 1.5) e por isto muitas vezes não responde a questão de atualização de probabilidades em situações como a considerada no exemplo 1.1 e geralmente nem em situações onde informações não previamente consideradas ocorrem.

Por **mudanças exógenas** de probabilidades entenderemos as mudanças feitas sem uso de fórmulas matemáticas adequadas para este fim, como são por exemplo as retaxações ou reavaliações arbitrárias de probabilidade.

Informações obtidas através de alguma percepção sensorial, em geral, não podem ser incorporadas pela fórmula de Bayes no cálculo da distribuição *a posteriori* por serem dificilmente traduzíveis como eventos e por conduzirem à certeza sobre o evento condicionante apenas em situações muito particulares. Richard Jeffrey (1965) em seu livro *The Logic of Decision* lança uma proposta para atualizar probabilidades que permite incorporar este tipo de informações em situações mais gerais do que faz o Condicionamento Bayesiano, mas também exigindo que tais informações não modifiquem o espaço amostral inicialmente considerado. A

regra proposta por Jeffrey permite mudanças exógenas de probabilidades para alguns eventos do espaço amostral de interesse, sem exigir certeza *a posteriori* sobre qualquer deles e a partir destas novas probabilidades estende as modificações para os demais elementos deste espaço amostral. Esta proposta de atualização de probabilidades é conhecida na literatura filosófica como **Probabilidade Cinética** ou **Regra de Condicionamento de Jeffrey**.

Por permitir mudança exógenas, para valores menores que um, nas probabilidade atribuídas a alguns eventos do espaço amostral, a Regra de Jeffrey também poderá ser útil em incorporar informações cujo recebimento não havia sido previamente antecipado em um leque mais amplo de problemas do que aquele permitido pelo Condicionamento Bayesiano (onde isto só é permitido em situações muito particulares, como vimos). Para que isso seja possível esta regra exige que a informação recebida atue sobre seu estado de conhecimento modificando apenas sua opinião sobre alguns eventos de Ω mas não a estrutura inicialmente concebida por Você para Ω.

Não nos preocuparemos em estudar definições de coerência que justifiquem o uso da Regra de Jeffrey como mecanismo de atualização de probabilidades. Discussões sobre definições de coerência justificando o uso de procedimentos de atualização de probabilidades podem ser encontrados em Goldstein(1985, 1983), Teller(1973), Lewis(1975) e Arment(1980). Alguns outros autores também destinam seus estudos à justificativas filosóficas para o uso de probabilidades Cinéticas (muitas vezes chamadas Dinâmicas) como representantes das opiniões modificadas de uma pessoa. Estas questões também não serão consideradas neste trabalho e além dos autores citados acima podemos sugerir Van Frassen(1980), Hacking(1967), Domotor(1980), dentre outros.

A seguir apresentaremos a definição da Regra de Jeffrey para espaços enumeráveis, exploraremos algumas de suas propriedades matemáticas e algumas de suas limitações enquanto mecanismo de atualização de probabilidades. Sua ligação com o Condicionamento Bayesiano também será considerada. Os resultados apresentados na seção 2.2 foram em sua maioria estabelecidos por Diaconis e Zabell(1982), Jeffrey(1992) e Howson e Urback(1993). Na seção 2.3 apresentaremos uma versão Preditivista para a Regra de Jeffrey e as seções 2.4 e 2.5 visam o estudo da relação de conjugação definida pela Regra de Jeffrey. Alguns resultados

particulares de conjugação para a família de distribuições exponencial serão fornecidos na seção 2.5. Os resultados sobre a Regra de Jeffrey exibidos nas três últimas seções deste capítulo não foram encontrados na literatura.

2.2 Apresentando a Regra de Jeffrey

Ao longo desta seção faremos basicamente uma revisão dos trabalhos encontrados sobre a Regra de Jeffrey, apresentando a sua formulação para espaços enumeráveis. Algumas de suas limitações e algumas justificativas para o seu uso serão exibidas considerando principalmente os trabalhos de Jeffrey(1992), Howson e Urback(1993) e Diaconis e Zabell(1982). Esta última referência preocupa-se basicamente com aspectos matemáticos da Regra de Jeffrey dando pouca importância ao seu aspecto fundamental.

Nossas contribuições nesta seção consistem em explicitar provas de alguns teoremas estabelecidos pelos autores anteriormente citados e em explorar as conseqüências da Violação da Lei de Cromwell (O'Hagan(1994)) à luz da Regra de Jeffrey.

Para uma apresentação formal da Regra de Jeffrey e para os resultados que serão apresentados ao longo desta seção vamos considerar um conjunto enumerável Ω, uma σ-álgebra $\mathcal{A}$, duas medidas P e $P_{\mathcal{E}}$ definidas sobre $(\Omega, \mathcal{A})$ e uma partição enumerável $\mathcal{E} = \{E_i, i \in I\}$ de Ω, a qual poderá ser finita ou infinita, onde I é um conjunto apropriado de índices.

A seguir apresentaremos a Regra de Jeffrey e suas conexões com o Condicionamento Bayesiano usual e com o conceito de Suficiência.

2.2.1 Definição e Conexões com Bayes e com Suficiência

Percebendo as limitações do Condicionamento Bayesiano usual Jeffrey(1965), ao abordar a questão sobre atualização de probabilidades, apresenta a seguinte proposta para este fim, a qual ao longo deste trabalho será denominada simplesmente por Regra de Jeffrey.

Definição 2.2.1 *(Regra de Jeffrey)*
Sejam P e $P_{\mathcal{E}}$ medidas de probabilidade definidas sobre $(\Omega, \mathcal{A})$. Dizemos que uma atualização de probabilidades é feita através da Regra de Jeffrey sempre que a distribuição a posteriori $P_{\mathcal{E}}$ for obtida partindo-se da distribuição a priori P, utilizando a seguinte fórmula:

$$P_{\mathcal{E}}(.) = \sum_i P(.|E_i)P_{\mathcal{E}}(E_i), \quad E_i \in \mathcal{E} \tag{2.1}$$

onde $\mathcal{E} = \{E_i, i \in I\}$ é uma partição de Ω para a qual $P_{\mathcal{E}}(E_i) \geq 0$ para todo $i \in I$, $\sum_{i \in I} P_{\mathcal{E}}(E_i) = 1$ e I é um conjunto de índices qualquer.

Interpretando P como sua opinião *a priori* e $P_{\mathcal{E}}$ como sua opinião após o recebimento imprevisto de alguma informação (ambas definidas sobre $(\Omega, \mathcal{A})$) percebemos da expressão (2.1) que a Regra de Jeffrey permite incorporar estas informações na atualização de opiniões iniciais por permitir mudanças exógenas sobre todos os elementos da partição de Ω gerada por tal informação. A atualização de opiniões feita pela Regra de Jeffrey nem sempre é, *a posteriori*, estaticamente coerente no sentido definido por de Finetti(1931), isto é, o uso da Regra de Jeffrey só nos fornece uma medida de probabilidade *a posteriori* sob certas condições (ver teorema 2.2.1, a seguir).

Nos casos em que o uso da Regra de Jeffrey é julgado adequado, a informação imprevista pode ser incorporada em sua opinião com uma economia de retaxações de probabilidades - basta declararmos novas probabilidades para os elementos da partição. De fato, a Regra de Jeffrey sob certas condições é um mecanismo de extensão de $P_{\mathcal{E}}$ - antes definida apenas sobre elementos da partição - para todos os demais eventos de $\mathcal{A}$.

Voltemos ao exemplo 1.1. da observação na penumbra para ilustrar a Regra de Jeffrey.

Exemplo 2.1: Suponha como antes que sua opinião inicial sobre as cores da roupa a ser posteriormente observada seja

$$\begin{aligned} P\{\text{Verde}\} &= 0,3 \\ P\{\text{Azul}\} &= 0,4 \\ P\{\text{ Lilás}\} &= 0,3 \end{aligned}$$

Ao observar a peça de roupa em pouca luz, Você passa a achar que existe maior chance de tal roupa ser não-Verde, mas ainda não consegue descartar completamente a hipótese de que a roupa possa também ser Verde. Se Você atribui probabilidade *a posteriori* igual a 0.8 para o evento " a roupa é não-Verde", a Regra de Jeffrey nos fornece a seguinte opinião *a posteriori* para os eventos atômicos inicialmente considerados:

$$
\begin{aligned}
P_{\mathcal{E}}\{\text{Azul}\} &= P_{\mathcal{E}}\{\text{Verde}\}P\{\text{Azul}|\text{Verde}\} + P_{\mathcal{E}}\{\text{não-Verde}\}P\{\text{Azul}|\text{não-Verde}\} = 16/35 \\
P_{\mathcal{E}}\{\text{Verde}\} &= 7/35 \\
P_{\mathcal{E}}\{\text{ Lilás}\} &= 12/35
\end{aligned}
$$

Como antes Você não sabe descrever em palavras o acontecimento que provocou a mudança de sua opinião sobre os elementos da partição formada pelos conjuntos $E = \{\text{Verde}\}$ e $\bar{E} = \{\text{ não-Verde}\}$. Apesar disto, a Regra de Jeffrey permite a atualização de suas probabilidades iniciais para uma nova medida $P_{\mathcal{E}}$ exigindo apenas que Você declare sua opinião *a posteriori* sobre os eventos E e $\bar{E}$, poupando o esforço mental que Você teria para eliciar sua opinião sobre todo evento de interesse.

Um dos inconvenientes da Regra de Jeffrey é que ela não fornece mecanismos para a obtenção da opinião *a posteriori* para os elementos da partição. A tarefa de eliciar novas opiniões sobre estes elementos pode ser tão psicologicamente complexa quanto a de declarar a distribuição de probabilidade *a priori* - o que depõe contra o uso da Regra de Jeffrey em algumas situações.

Outro inconveniente da Regra de Jeffrey é que, como está definida, esta não nos fornece uma opinião *a posteriori* estaticamente coerente isto é, não fornece uma medida de probabilidade *a posteriori*. Perceba que particularmente no exemplo acima Você é coerente *a posteriori* e nesse exemplo foi considerado que todas as probabilidades condicionais nos eventos E e $\bar{E}$ permanecem inalteradas *a posteriori*. Desta forma Jeffrey(1992) fornece o seguinte resultado estipulando condições para a coerência (estática) das atualizações feitas através da Regra de Jeffrey.

Teorema 2.2.1 *Seja (Ω, A) um espaço de probabilidade qualquer com Ω (enumerável). Considere uma evento $\mathcal{A}$-mensurável A e suponha que para todo $E_i \in \mathcal{E}$, $P_{\mathcal{E}}(E_i) > 0$ são conhecidos. Então,*

$$P_{\mathcal{E}}(A) = \sum_i P(A|E_i)P_{\mathcal{E}}(E_i) \tag{2.2}$$

se, e somente se, para todo $A \in \mathcal{A}$ e para todo $E_i \in \mathcal{E}$ segue que

$$P(A|E_i) = P_{\mathcal{E}}(A|E_i).^1 \tag{2.3}$$

Prova: Para suficiência basta notar que para cada $A \in \mathcal{A}$ e para qualquer medida $P_{\mathcal{E}}$ definida sobre $(\Omega, \mathcal{A})$

$$\begin{aligned}
\sum_{i\in I} P(A|E_i)P_{\mathcal{E}}(E_i) &= \sum_{i\in I} P_{\mathcal{E}}(A|E_i)P_{\mathcal{E}}(E_i) \\
&= \sum_{i\in I} P_{\mathcal{E}}(A, E_i) \\
&= P_{\mathcal{E}}(A, \cup_{i\in I} E_i) = P_{\mathcal{E}}(A).
\end{aligned}$$

Para mostrar a condição necessária basta provar que a expressão (2.3) vale para os átomos $\omega \in \Omega$.

Se $\omega \in E_i$ então usando a hipótese obtém-se:

$$\begin{aligned}
P_{\mathcal{E}}(\{\omega\}|E_i) &= \frac{P_{\mathcal{E}}(\{\omega\}, E_i)}{P_{\mathcal{E}}(E_i)} = \frac{P_{\mathcal{E}}(\{\omega\})}{P_{\mathcal{E}}(E_i)} \\
&= \frac{P(\{\omega\}|E_i)P_{\mathcal{E}}(E_i)}{P_{\mathcal{E}}(E_i)} \\
&= P(\{\omega\}|E_i),
\end{aligned}$$

[1]Se, ao invés de eventos condicionais em Ω, tivéssemos considerado eventos condicionais da forma $A|B$, com A e B pertencentes a $\mathcal{A}$, a Regra de Jeffrey tornar-se-ia

$$P_{\mathcal{E}}(A|B) = \sum_i P(A|B, E_i)P_{\mathcal{E}}(E_i|B)$$

e a J-condição a ser verificada assumiria a seguinte forma:

$$P(A|B, E_i) = P_{\mathcal{E}}(A|B, E_i)$$

caso contrário, temos $P_{\mathcal{E}}(\{\omega\}|E_i) = 0 = P(\{\omega\}|E_i)$. Ou seja, se vale a hipótese, temos que para todo $\omega \in \Omega$ e para todo $E_i \in \mathcal{E}$ tal que $P_{\mathcal{E}}(E_i) > 0$,

$$P_{\mathcal{E}}(\{\omega\}|E_i) = P(\{\omega\}|E_i).$$

Considerando A como a união de alguns átomos de Ω, os Axiomas de Kolmogorov concluem a prova do teorema.

□

Este teorema nos fornece uma condição para a validade da Regra de Jeffrey. Se a informação recebida de maneira imprevista atuar sobre seu conhecimento de tal forma que a condição (2.3) possa ser julgada válida (por Você) e se provocar mudanças exógenas para valores estritamente positivos apenas nas probabilidades de eventos pertencentes a uma partição do espaço amostral, a Regra de Jeffrey diz como Você pode estender a medida $P_{\mathcal{E}}$ para os outros eventos de $\mathcal{A}$, coerentemente.

A condição (2.3) é chamada de **J- Condição** e é conseqüência dos julgamentos que Você faz *a priori* e *a posteriori* sobre os eventos condicionais de interesse - isto é, a J-condição não pode ser matematicamente verificada. Em muitas situações, julgar válida a J-condição pode não ser razoável como ilustra o seguinte exemplo.

Exemplo 2.2: Suponha que num hospital serão realizadas três transplantes de fígado utilizando um novo procedimento cirúrgico. Considere a variável aleatória

$$X_i = \begin{cases} 1 & \text{se a } i\text{- ésima cirurgia é sucesso} \\ 0 & \text{caso contrário} \end{cases}$$

para $i = 1, 2, 3$. Considere os eventos $E = \{X_1 = 1\}$ denotando a primeira cirurgia é sucesso e a variável aleatória $N = \sum_{i=1}^{3} X_i$ que conta o número de sucessos nas três cirurgias. Inicialmente, Você julga que

$$\begin{aligned} P(N = 0|E) &= 0 \\ P(N = 1|E) &= 1/4 \end{aligned}$$

$$
\begin{aligned}
P(N = 2|E) &= 2/4 \\
P(N = 3|E) &= 1/4.
\end{aligned}
$$

Suponha que Você receba a informação de que este mesmo procedimento cirúrgico foi realizado no Hospital das Clínicas e dentre os 100 pacientes submetidos à cirurgia, 94 foram bem sucedidos. Neste caso, não parece muito razoável continuar considerando $P_{\mathcal{E}}(N = 3|E) = 1/4$. Intuitivamente, esta informação nos leva a pensar que $P_{\mathcal{E}}(N = 3|E)$ é muito maior que $1/4$.

Julgar válida a J-condição aqui não nos parece razoável e, por conseqüência, a atualização de probabilidades através da Regra de Jeffrey não é viável.

O exemplo anterior também ilustra que não é possível fazermos uma verificação algébrica da J-condição. A avaliação da J-condição é completamente subjetiva e pessoal como são os julgamentos de independência e permutabilidade e a avaliação de probabilidades - todos frutos de sua percepção sobre o objeto desconhecido.

No entanto, é possível estudar características da partição de Ω, gerada pelo recebimento imprevisto de informação, que permitam garantir a atualização de probabilidades via Regra de Jeffrey. Diaconis e Zabell(1982) notam que se a informação recebida faz com que Você particione o espaço amostral de tal forma que esta partição seja suficiente para a família $\{P, P_{\mathcal{E}}\}$ temos como conseqüência da definição de partição suficiente $\mathcal{E} = \{E_i, i \in I\}$ que para todo $\omega \in \Omega$ e para todo $Q \in \{P, P_{\mathcal{E}}\}$ existe uma função real g definida em Ω tal que para todo $E_i \in \mathcal{E}$

$$Q(\{\omega\}|E_i) = g(\omega). \tag{2.4}$$

(Ver Pereira *et alii* (1993 - cap 6) para resultados de suficiência.) Então sob suficiência temos que $P(\{\omega\}|E_i) = P_{\mathcal{E}}(\{\omega\}|E_i)$ para todo $E_i \in \mathcal{E}$ e para todo $\omega \in \Omega$. Conseqüentemente, se $\mathcal{E}$ é uma partição suficiente para $\mathcal{P} = \{P, P_{\mathcal{E}}\}$ a J-condição é verificada viabilizando a atualização de P através da Regra de Jeffrey.

Encontrar uma partição suficiente para $\mathcal{P}$ é sempre possível uma vez que a partição trivial $\mathcal{E} = \{\omega : \omega \in \Omega\}$ é sempre uma partição suficiente para $\mathcal{P}$. Nestes casos, o uso da Regra

de Jeffrey para atualizar probabilidades é viável pois a J-condição é verificada mas para qualquer evento $A \in \mathcal{A}$ temos

$$P_{\mathcal{E}}(A|\{\omega\}) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{caso contrário} \end{cases}$$

e por conseguinte temos

$$\begin{aligned} P_{\mathcal{E}}(A) &= \sum_{w} P(A|\{\omega\})P_{\mathcal{E}}(\{\omega\}) \\ &= \sum_{\omega \in A} P_{\mathcal{E}}(\{\omega\}) = P_{\mathcal{E}}(A), \end{aligned}$$

ou seja, o uso da Regra de Jeffrey equivale a uma retaxação completa de probabilidades.

Interessa-nos termos situações onde seja exigido um número pequeno de declarações de novas probabilidades. Por este motivo, a partição gerada pela nova informação deveria conter o menor número de elementos possível. Diaconis e Zabell(1982) estabelecem o seguinte resultado o qual fornece uma versão alternativa para a Regra de Jeffrey e que garante a existência de uma partição suficiente mínima para a qual a Regra de Jeffrey pode ser considerada válida. Esta partição é a mais grossa pois cada um de seus elementos pode ser escrito como uma união disjunta de elementos de alguma outra partição suficiente para $\mathcal{P}$ e por isto em muitos casos pode diminuir o número de retaxações arbitrárias de probabilidades a serem feitas ao usarmos a Regra de Jeffrey, salvo naquelas situações em que a partição suficiente mínima coincida com a partição trivial, como vimos.

Teorema 2.2.2 *Sejam P e $P_{\mathcal{E}}$ medidas de probabilidades com suporte comum sobre um conjunto enumerável Ω. Se $\mathcal{E} = \{E_i, i \in I\}$ é uma partição de Ω tal que $P(E_i) > 0$ e $P(A|E_i) = P_{\mathcal{E}}(A|E_i)$ para todo $A \subseteq \Omega$ e para todo $E_i \in \mathcal{E}$, então para cada $\omega \in \Omega$,*

$$P_{\mathcal{E}}(\{\omega\}) = \frac{P_{\mathcal{E}}(E_i)}{P(E_i)} P(\{\omega\}), \quad \text{se } \omega \in E_i. \tag{2.5}$$

Além disso, se $R = \{x : \frac{P_{\mathcal{E}}(\{\omega\})}{P(\{\omega\})} = x, \omega \in \Omega\}$ e $E_x = \{\omega : \frac{P_{\mathcal{E}}(\{\omega\})}{P(\{\omega\})} = x, \omega \in \Omega\}$, então $\mathcal{E}^* = \{E_x, x \in R\}$ é uma partição suficiente mínima para $\mathcal{P} = \{P, P_{\mathcal{E}}\}$.

Prova : Primeiramente considere $\omega \in E_i$. Como por hipótese $P(A|E_i) = P_{\mathcal{E}}(A|E_i)$ para todo $E_i \in \mathcal{E}$ e para todo A, podemos usar a Regra de Jeffrey para o evento $A = \{\omega\}$ e obtemos

$$\begin{aligned} P_{\mathcal{E}}(\{\omega\}) &= \sum_j P(\{\omega\}|E_j)P_{\mathcal{E}}(E_j) = P(\{\omega\}|E_i)P_{\mathcal{E}}(E_i) \\ &= \frac{P(\{\omega\}, E_i)}{P(E_i)} P_{\mathcal{E}}(E_i) \\ &= \frac{P_{\mathcal{E}}(E_i)}{P(E_i)} P(\{\omega\}), \end{aligned}$$

verificando a primeira parte do teorema.

Para a segunda parte, note da definição de E_x que $\mathcal{E}^*$ é uma partição de Ω. Para verificar que $\mathcal{E}^*$ é suficiente considere um elemento qualquer $E_x \in \mathcal{E}^*$.

Para todo $\omega \in E_x$ temos, também pela definição de E_x, que

$$P_{\mathcal{E}}(\{\omega\}) = xP(\{\omega\}). \tag{2.6}$$

Seja $\omega \in \Omega$. Então $\omega \in E_x$ para algum x. Para este x a expressão (2.6) fornece

$$\begin{aligned} P_{\mathcal{E}}(\{\omega\}|E_x) &= \frac{P_{\mathcal{E}}(\{\omega\})}{P_{\mathcal{E}}(E_x)} \\ &= \frac{xP(\{\omega\})}{\sum_{y \in E_x} P_{\mathcal{E}}(\{y\})} \\ &= \frac{P(\{\omega\})}{\sum_{y \in E_x} P(\{y\})} \\ &= \frac{P(\{\omega\})}{P(E_x)} = P(\{\omega\}|E_x). \end{aligned}$$

Concluíndo que $\mathcal{E}^*$ é uma partição suficiente para $\mathcal{P}$.

Considere uma partição qualquer $\mathcal{F} = \{F_j, j \in J\}$ suficiente para $\mathcal{P}$ onde J é um conjunto qualquer de índices. Vamos mostrar que para cada x, E_x é uma união disjunta de elementos de $\mathcal{F}$.

Considere $\omega \in E_x$. Como $\omega \in \Omega$ e como $\mathcal{F}$ é uma partição suficiente para $\mathcal{P}$ existe um único F_j tal que $\omega \in F_j$. Daí, pela parte um do teorema e considerando que $\frac{P_{\mathcal{E}}(\omega)}{P(\omega)} = x$, temos que para cada $\omega' \in F_j$

$$\frac{P_{\mathcal{E}}(\{\omega'\})}{P(\{\omega'\})} = \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} = \frac{P_{\mathcal{E}}(\{\omega\})}{P(\{\omega\})} = x.$$

Portanto, $F_j \subseteq E_x$. Desta forma,

$$\bigcup \left[F_j : \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} = x \right] \subseteq E_x.$$

Suponha, por absurdo, que exista $\omega \in E_x$ tal que $\omega \notin \bigcup \left[F_j : \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} = x \right]$. Então existe um $F_{j_0} \in \mathcal{F}$ com $F_{j_0} \not\subseteq \bigcup \left[F_j : \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} = x \right]$ tal que $\omega \in F_{j_0}$. Como $\mathcal{F}$ é suficiente para $\mathcal{P}$

$$\frac{P_{\mathcal{E}}(\{\omega\})}{P(\{\omega\})} = \frac{P_{\mathcal{E}}(F_{j_0})}{P(F_{j_0})} \neq x \;\Rightarrow\; \omega \notin E_x.$$

Portanto, para todo $\omega \in E_x$, temos que

$$\omega \in \bigcup \left[F_j : \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} = x \right]$$

e com isto, verificamos que $\mathcal{E}^* = \{E_x : x \in R\}$ é uma partição suficiente mínima. Concluíndo a prova.

□

Uma prova alternativa para este teorema pode ser obtida utilizando-se o teorema da Fatoração de Fisher-Neyman e resultados de Blackwell e Girshick (1954 - pg. 221).

O seguinte exemplo ilustra o uso da Regra de Jeffrey na versão dada no teorema anterior. Perceba, da parte um do teorema 2.2.2, que dentro de um mesmo elemento E_i da partição as probabilidades de cada átomo são atualizadas multiplicando-se a probabilidade *a priori* por um fator constante próprio daquele elemento da partição - ou seja, a informação recebida de forma imprevista não discrimina átomos de um mesmo conjunto da partição.

Exemplo 2.3: Sejam A, B e C corredores que disputam a prova dos cem metros livres. A tem tido mal desempenho nas últimas corridas em que participou e por isso Você acredita

que a chance dele vencer esta prova é pequena. Inicialmente Você atribui as seguintes chances para a vitória de cada atleta:

$$\begin{aligned} P(A) &= 2/11 \\ P(B) &= 4/11 \\ P(C) &= 5/11. \end{aligned}$$

Chega inesperadamente aos seus ouvidos a informação de que os atletas B e C tiveram alguns problemas de saúde na última semana antes da prova e por isso o treinamento deles não pode ser realizado de maneira satisfatória. Esta informação (que favorece A) não era previamente suspeitada por Você e faz com que Você passe (arbitrariamente) a acreditar que a chance de A vencer aumenta para $1/2$.

Considerando a seguinte partição

$$E_1 = \{A\} \text{ e } E_2 = \{\bar{A}\} = \{B, C\}$$

verificamos que, com a nova informação, sua opinião atualizada sobre E_2 é $P_{\mathcal{E}}(E_2) = 1/2$ e com isto,

$$\frac{P_{\mathcal{E}}(E_2)}{P(E_2)} = \frac{1/2}{9/11} = \frac{11}{18}.$$

Agora a nova informação não permite distinguir entre B e C. Então usando o teorema 2.2.2, verificamos que a influência na nova informação sobre sua opinião faz com que as chances de vitória de B e C passem para

$$\begin{aligned} P_{\mathcal{E}}(B) &= \frac{4}{11}.\frac{11}{18} = \frac{4}{18} \\ P_{\mathcal{E}}(C) &= \frac{5}{11}.\frac{11}{18} = \frac{5}{18}. \end{aligned}$$

O teorema 2.2.2 constrói uma partição suficiente mínima para $\mathcal{P}$ a partir das probabilidades retaxadas, mas não garante que a partição construída seja não - trivial. Se no exemplo anterior a informação recebida permitisse também uma distinção entre B e C, a partição

suficiente mínima seria a trivial e portanto a atualização de probabilidades consistiria em uma reavaliação (arbitrária) completa das probabilidades iniciais.

Perceba também que o uso do Condicionamento Bayesiano visto no capítulo anterior seria aqui inapropriado, pois a informação recebida inesperadamente não o leva à certeza sobre vitória de A.

Partindo da definição 2.2.1, percebemos que ao atualizar sua opinião sobre um evento qualquer A pela Regra de Jeffrey, quando $P(E_i) \neq 0, \forall i \in I$, esta atualização é feita através do produto de uma função influenciada pela informação não - programada pela distribuição *a priori* atribuída a A, como pode ser visto na seguinte relação

$$\begin{aligned} P_{\mathcal{E}}(A) &= \sum_i P(A|E_i)P_{\mathcal{E}}(E_i) \\ &= \sum_i \frac{P(E_i|A)}{P(E_i)} P_{\mathcal{E}}(E_i)P(A). \end{aligned}$$

Fora casos particulares como este, atualizações de probabilidades feitas usando a Regra de Jeffrey nem sempre são alguma forma de condicionamento no sentido de existir alguma função que propicie a atualização da distribuição de probabilidades *a priori*.

Quando $P(E_i) = 0$, para algum $i \in I$, a escolha a ser feita para o valor de $P(A|E_i)$ é arbitrária. Sendo assim, supondo $P(E_{i_o}) = 0$ e considerando $P_{\mathcal{E}}(E_{i_0}) > 0$, uma atualização de probabilidades através da Regra de Jeffrey torna-se

$$P^*(A) = \sum_{i \neq i_o} \frac{P(E_i|A)P^*(E_i)}{P(E_i)} P(A) + P(A|E_{i_o})P^*(E_{i_o}), \tag{2.7}$$

o que não necessariamente é da forma $f(\cdot)P(A)$, onde f é uma função real não-negativa qualquer.

Apesar disso, observamos que o Condicionamento Bayesiano usual é um caso particular da Regra de Jeffrey. Este fato pode ser apreciado no seguinte resultado encontrado em Jeffrey(1992).

Teorema 2.2.3 *Sejam P e $P_{\mathcal{E}}$ medidas de probabilidades definidas sobre $(\Omega, \mathcal{A})$, com $P_{\mathcal{E}}$ obtida de P através da Regra de Jeffrey. Considere $\mathcal{E} = \{E, \bar{E}\}$ uma partição de Ω com*

$0 < P(E) < 1$. *Se* $P_{\mathcal{E}}(E) = 1$, *para qualquer evento* A *de* Ω, *temos*

$$P_{\mathcal{E}}(A) = P(A|E) \tag{2.8}$$

Prova: É conseqüência imediata das hipóteses e da definição 2.2.1.

Este teorema ressalta mais uma vez que o Condicionamento Bayesiano usual também permite mudanças exógenas de probabilidades sobre elementos de uma partição de Ω. Para que o uso do Condicionamento Bayesiano seja justificado a informação não-programada ou obtida através de alguma percepção sensorial deve conduzir Você à certeza sobre algum dos eventos da partição.

Na próxima seção discutiremos mais algumas questões fundamentais sobre a Regra de Jeffrey, apresentando resultados estabelecidos por Diaconis e Zabell(1982) sobre comutatividade desta regra de atualização de probabilidades. Também abordaremos as implicações da violação da Lei de Cromwell à luz da Regra de Jeffrey, o que não encontramos na literatura.

2.2.2 Comutatividade e um Pouco Mais de Fundamentos

Um aspecto importante das atualizações de probabilidade feitas através do Condicionamento Bayesiano é que este procedimento não permite atualizar probabilidades que eram inicialmente zero - quer dizer, o Condicionamento Bayesiano usual não fornece probabilidades *a posteriori* não-nulas partindo-se de probabilidades iniciais nulas. Por este motivo costuma-se dizer que um Bayesiano não deve violar a **Lei de Cromwell** (O'Hagan(1994)) - isto é, não deve atribuir probabilidades zero a eventos que julga serem pouco prováveis de ocorrer.

Com a Regra de Jeffrey a situação é um pouco diferente. Por permitir retaxações arbitrárias de probabilidades sobre eventos de uma partição de Ω, nada impede que Você atribua posteriormente probabilidade não-nula para algum evento desta partição que Você *a priori* considerava improvável.

O mesmo não ocorre, no entanto, se o evento A não for um evento da partição. Neste caso, a probabilidade *a posteriori* para A dada pela Regra de Jeffrey é também nula, pois

$$P_{\mathcal{E}}(A) = \sum_{i \in I} \frac{P(E_i|A)P(A)}{P(E_i)} P_{\mathcal{E}}(E_i) = 0. \tag{2.9}$$

Ou seja, desde que a todos os elementos da partição sejam atribuídas probabilidades *a priori* não-nulas, a Regra de Jeffrey não permite atualizar probabilidades de eventos julgados inicialmente improváveis. Desta forma, a Lei de Cromwell poderá ser violada sem conseqüências indesejáveis em atualizações feitas pela Regra de Jeffrey apenas quando Você estiver certo que todo evento para o qual Você atribuiu probabilidade inicial igual a zero pertencerá à partição gerada pela informação que Você receberá - o que, se puder ser antecipado, depõe contra o caráter imprevisto do recebimento desta informação.

Outro aspecto importante a ser explorado é a comutatividade de atualizações feitas usando a Regra de Jeffrey quando mais de uma informação não previamente antecipada é recebida. Quando o Condicionamento Bayesiano é utilizado para atualizar probabilidades, a ordem em que as informações são incorporadas em sua opinião não afeta a distribuição *a posteriori*, ou seja, o Condicionamento Bayesiano é comutativo como pode ser visto no teorema que segue.

Considere $\mathcal{E} = \{E, \bar{E}\}$ a partição de Ω gerada pelo conhecimento de uma dada informação e $\mathcal{F} = \{F, \bar{F}\}$ uma outra partição de Ω gerada ao tomar-se ciência de uma outra informação. Defina $P_{\mathcal{E}}$ (e, analogamente,$P_{\mathcal{F}}$) como sendo a nova medida de probabilidade atualizada partindo-se da distribuição inicial P após o conhecimento da informação que gerou $\mathcal{E}$ ($\mathcal{F}$) e considere $P_{\mathcal{E}\mathcal{F}}$ ($P_{\mathcal{F}\mathcal{E}}$) como sendo a medida de probabilidade obtida partindo da distribuição *a priori* $P_{\mathcal{E}}$ ($P_{\mathcal{F}}$) após o conhecimento da informação que gerou $\mathcal{F}$ ($\mathcal{E}$).

Teorema 2.2.4 *Se as atualizações são feitas através de Condicionamento Bayesiano e se* $P(E, F) \neq 0$ *então* $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$.

Prova: Seja A um evento qualquer de Ω. Como $P(E, F) \neq 0$, verificamos que

$$P_{\mathcal{E}\mathcal{F}}(A) = P_{\mathcal{E}}(A|F)$$

$$
\begin{aligned}
&= \frac{P(A,F|E)}{P(F|E)} \\
&= \frac{P(A,E|F)}{P(E|F)} \\
&= \frac{P_{\mathcal{F}}(A,E)}{P_{\mathcal{F}}(E)} \\
&= P_{\mathcal{F}\mathcal{E}}(A).
\end{aligned}
$$

Portanto, o Condicionamento Bayesiano é comutativo.

□

Ao fazermos usos sucessivos da Regra de Jeffrey para atualizarmos probabilidades, além de realizar testes subjetivos para a J-condição em cada estágio, nos deparamos com a não - comutatividade da Regra de Jeffrey como é ilustrado no exemplo que segue.

Exemplo 2.4 : Um crime é cometido e existem quatro suspeitos sob a mira da polícia identificados por A, B, C e D. Inicialmente, não existem provas concretas que permitem incriminar qualquer destes suspeitos o que leva o detetive X a pensar que todos eles têm a mesma chance de ser o culpado, isto é, para X

$$P(A) = P(B) = P(C) = P(D) = 1/4,$$

onde o evento M representa que o suspeito M é culpado.

Após algumas investigações o detetive X encontra evidências de que o criminoso é canhoto e na opinião dele a chance do criminoso ser canhoto é $0,8$. Esta informação gera a seguinte partição:

$$E_1 = \{A, B\} \text{ e } E_2 = \{C, D\},$$

com $P_{\mathcal{E}}(E_1) = 0,8$.

Para X as probabilidades condicionais em E_1 e em E_2 *a posteriori* são iguais às probabilidades *a priori* condicionais nestes mesmos eventos e com isso utilizando o teorema 2.2.2,

obtém-se:

$$P_{\mathcal{E}}(A) = P_{\mathcal{E}}(B) = 0,4$$
$$P_{\mathcal{E}}(C) = P_{\mathcal{E}}(D) = 0,1.$$

Novas investigações são feitas e X descobre que o criminoso poderia ter sido uma mulher com probabilidade (a seu ver) $0,7$. Esta informação induz a seguinte partição:

$$F_1 = \{A, B\} \text{ e } F_2 = \{C, D\},$$

com $P_{\mathcal{E}\mathcal{F}}(F_1) = 0,7$.

Para X as novas probabilidades condicionais em elementos desta partição são idênticas às probabilidades condicionais nestes mesmos eventos determinadas apenas com base na primeira informação. Sendo assim que ele usa a Regra de Jeffrey para atualizar suas probabilidades obtendo

$$P_{\mathcal{E}\mathcal{F}}(A) = P_{\mathcal{E}\mathcal{F}}(B) = 0,35 \tag{2.10}$$
$$P_{\mathcal{E}\mathcal{F}}(C) = P_{\mathcal{E}\mathcal{F}}(D) = 0,15. \tag{2.11}$$

Suponha agora que as informações tenham aparecido em ordem inversa, isto é, a primeira evidência da qual X tomou conhecimento foi a de que o criminoso poderia ter sido uma mulher com probabilidade $0,7$ e depois X tomou ciência que o criminoso poderia ter sido um canhoto com probabilidade $0,8$. Supondo válida a J-condição em cada passo e fazendo cálculos análogos determinamos que

$$P_{\mathcal{F}\mathcal{E}}(A) = P_{\mathcal{F}\mathcal{E}}(B) = 0,40 \tag{2.12}$$
$$P_{\mathcal{F}\mathcal{E}}(C) = P_{\mathcal{F}\mathcal{E}}(D) = 0,10. \tag{2.13}$$

Comparando os resultados obtidos em (2.10) e (2.12) por exemplo, verificamos que

$$P_{\mathcal{E}\mathcal{F}} \neq P_{\mathcal{F}\mathcal{E}} \tag{2.14}$$

e, portanto, o cálculo de distribuições *a posteriori* através da Regra de Jeffrey é alterado pela ordem de chegada das informações.

Perceba que para obter-se $P_{\mathcal{E}\mathcal{F}}$ (segue analogamente para $P_{\mathcal{F}\mathcal{E}}$) foi necessário verificar a J-condição em cada passo, isto é Você teve que julgar válidas as seguintes igualdades

$$\begin{aligned} P_{\mathcal{E}}(.|E_i) &= P(.|E_i), \quad i \in I \\ P_{\mathcal{E}\mathcal{F}}(.|F_j) &= P_{\mathcal{E}}(.|F_j), \quad j \in J. \end{aligned}$$

Sob que condições a comutatividade também é uma propriedade da Regra de Jeffrey? Diaconis e Zabell(1982) estabelecem que se há comutatividade nas avaliações de probabilidades feitas para os elementos de ambas partições, então esta propriedade estende-se ao espaço todo, permitindo a comutatividade das atualizações realizadas pela Regra de Jeffrey. Formalmente, se é razoável considerar que

$$P_{\mathcal{E}}(E_i) = P_{\mathcal{F}\mathcal{E}}(E_i) = p_i \quad i \in I \tag{2.15}$$

$$P_{\mathcal{F}}(F_j) = P_{\mathcal{E}\mathcal{F}}(F_j) = q_j \quad j \in J, \tag{2.16}$$

Diaconis e Zabell(1982) estabelecem o seguinte resultado

Teorema 2.2.5 *Suponha válidas as condições (2.15) e (2.16). Se $P_{\mathcal{F}\mathcal{E}}(F_j) = q_j$ e $P_{\mathcal{E}\mathcal{F}}(E_i) = p_i$ para todo $i \in I$ e para todo $j \in J$, então*

$$P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}. \tag{2.17}$$

Prova: Segue do teorema 3.2 de Csiszár(1975) que garante que $P_{\mathcal{F}\mathcal{E}}$ e $P_{\mathcal{E}\mathcal{F}}$ são as I-projeções de P sobre o conjunto de medidas $\mathcal{C} = \{M : \Omega \to [0,1] : M(E_i) = p_i$ e $M(F_j) = q_j,\ i \in I,\ j \in J\}$. Como $\mathcal{C}$ é um conjunto convexo, a unicidade da I-projeção é garantida (teorema 2.1, Csiszár(1975)) e portanto,

$$P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}.$$

□

OBS: Csiszár(1975) define I-projeção de uma medida P sobre um conjunto de medidas $\mathcal{C}$ como a medida $Q \in \mathcal{C}$ tal que

$$I(Q, P) = \min_{R \in \mathcal{C}} I(R, P)$$

onde

$$I(Q,P) = \begin{cases} \int q \log q dP & \text{se } Q << P \\ +\infty & \text{se } P << Q \end{cases}$$

e q é a densidade de Q com respeito a P.

Para ilustrar a importância da hipótese do teorema anterior na constatação da comutatividade das atualizações feitas através da Regra de Jeffrey vamos considerar o seguinte exemplo:

Exemplo 2.5: No exemplo 2.4, verificamos que as condições (2.15) e (2.16) são aceitas por constatar que

$$\begin{aligned} P_{\mathcal{E}}(E_1) &= P_{\mathcal{F}\mathcal{E}}(E_1) = 0,8 \\ P_{\mathcal{E}}(E_2) &= P_{\mathcal{F}\mathcal{E}}(E_2) = 0,2, \end{aligned}$$

e

$$\begin{aligned} P_{\mathcal{F}}(F_1) &= P_{\mathcal{E}\mathcal{F}}(F_1) = 0,7 \\ P_{\mathcal{F}}(F_2) &= P_{\mathcal{E}\mathcal{F}}(F_2) = 0,3. \end{aligned}$$

Apesar disso, determinamos que não existe comutatividade dentro das partições pois

$$P_{\mathcal{E}\mathcal{F}}(E_1) = 0,7 \neq 0,8 = P_{\mathcal{F}\mathcal{E}}(E_1),$$

seguindo, pelo teorema anterior, a não comutatividade das atualizações feitas pela Regra de Jeffrey.

Perceba que para a verificação de comutatividade da Regra de Jeffrey é inicialmente suposto que as probabilidades atribuídas a elementos das partições, modificadas de maneira exógena, são mantidas fixas (isto é, consideram-se as condições (2.15) e (2.16)). Por exemplo, não importa se sua opinião *a priori* seja P ou a já atualizada pelo recebimento da informação que gera $\mathcal{E}$, $P_{\mathcal{E}}$, ao tomar conhecimento da informação que gera $\mathcal{F}$ suas atribuições de probabilidades sobre elementos de $\mathcal{F}$ é tal que $P_{\mathcal{F}}(F_j) = P_{\mathcal{E}\mathcal{F}}(F_j), \forall j$. Esta suposição insinua um certo tipo de independência entre as informações para que a comutatividade possa ser

constatada. De fato, o conceito de Jeffrey comutatividade esta vinculado a um conceito de independência mais fraco que o habitual - a Jeffrey independência. Este conceito de Jeffrey independência foi estabelecido por Diaconis e Zabell(1982) e revela as características que sua opinião deve ter para que duas partições sejam independentes de acordo com a Regra de Jeffrey.

Considere P como sua distribuição *a priori* e $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ como definidos nas expressões (2.15) e (2.16).

Definição 2.2.2 *(Jeffrey Independência)*
As partições $\mathcal{E} = \{E_i, i \in I\}$ e $\mathcal{F} = \{F_j, j \in J\}$ de Ω são Jeffrey-independentes com respeito a P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ se

$$P_{\mathcal{E}}(F_j) = P(F_j) \quad j \in J \tag{2.18}$$

$$P_{\mathcal{F}}(E_i) = P(E_i) \quad i \in I. \tag{2.19}$$

Considere $P_{\mathcal{E}\mathcal{F}}$ e $P_{\mathcal{F}\mathcal{E}}$ como as probabilidades obtidas através de mudanças sucessivas pela Regra de Jeffrey. O conceito de Jeffrey independência é conectado à comutatividade pelo seguinte teorema.

Teorema 2.2.6 *Considere P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ dadas. Então, $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$ se, e somente se, $\mathcal{E}$ e $\mathcal{F}$ são Jeffrey independentes com respeito a P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$.*

Prova: Seja A um evento qualquer de Ω e suponha válidas as J-condições referentes às mudanças sucessivas via Regra de Jeffrey que seguem. Neste caso, obtemos:

$$\begin{aligned} P_{\mathcal{E}\mathcal{F}}(A) &= \sum_{ij} P_{\mathcal{E}}(E_i) P_{\mathcal{E}\mathcal{F}}(F_j) \frac{P(A, F_j, E_i)}{P(E_i) P_{\mathcal{E}}(F_j)} \\ &= \sum_{ij} p_i q_j \frac{P(A, F_j, E_i)}{P(E_i) P_{\mathcal{E}}(F_j)}, \end{aligned} \tag{2.20}$$

e, também,

$$P_{\mathcal{F}\mathcal{E}}(A) = \sum_{ij} P_{\mathcal{F}}(F_j) P_{\mathcal{F}\mathcal{E}}(E_i) \frac{P(A, F_j, E_i)}{P(F_j) P_{\mathcal{F}}(E_i)}$$

$$= \sum_{ij} p_i q_j \frac{P(A, F_j, E_i)}{P(F_j)P_{\mathcal{F}}(E_i)}. \tag{2.21}$$

Como por hipótese, $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$ temos que

$$\sum_{ij} p_i q_j \frac{P(A, F_j, E_i)}{P(E_i)P_{\mathcal{E}}(F_j)} = \sum_{ij} p_i q_j \frac{P(A, F_j, E_i)}{P(F_j)P_{\mathcal{F}}(E_i)}, \tag{2.22}$$

para todo A. Em particular, vale para $A = \{E_{i_0}, F_{j_0}\}$. Substituindo este valor na expressão (2.22) acima obtemos

$$P(E_{i_0})P_{\mathcal{E}}(F_{j_0}) = P(F_{j_0})P_{\mathcal{F}}(E_{i_0}), \quad \forall\ i_0,\ j_0. \tag{2.23}$$

Então somando-se os dois lados da igualdade em (2.23) em j_0 obtemos

$$P(E_{i_0}) = P_{\mathcal{F}}(E_{i_0}), \quad \forall\, i_0. \tag{2.24}$$

Analogamente, somando (2.23) em i_0 obtemos

$$P(F_{j_0}) = P_{\mathcal{E}}(F_{j_0}), \quad \forall\, j_0. \tag{2.25}$$

Portanto, $\mathcal{E}$ e $\mathcal{F}$ são Jeffrey independentes.

Por outro lado, se $P(E_i) = P_{\mathcal{F}}(E_i)$ para todo $i \in I$ e $P(F_j) = P_{\mathcal{E}}(F_j)$ para todo $j \in J$ temos que

$$P(E_i)P_{\mathcal{E}}(F_j) = P(F_j)P_{\mathcal{F}}(E_i) \quad \forall i, j. \tag{2.26}$$

Então voltando a (2.20) concluímos que

$$P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}.$$

□

Para verificar a comutatividade, além da suposição de Jeffrey independência é necessário manter as considerações iniciais sobre as probabilidades para os elementos das partições atualizadas pelas informações que as geraram. A saber, são consideradas fixas

$$\begin{aligned} P_{\mathcal{E}}(E_i) &= P_{\mathcal{F}\mathcal{E}}(E_i) = p_i \quad \forall i \\ P_{\mathcal{F}}(F_j) &= P_{\mathcal{E}\mathcal{F}}(F_j) = q_j \quad \forall j. \end{aligned}$$

O próximo passo é verificar que a Jeffrey independência é mesmo um julgamento mais fraco que independência. No sentido usual, duas partições $\mathcal{E} = \{E_i, i \in I\}$ e $\mathcal{F} = \{F_j, j \in J\}$ de Ω são independentes com respeito a P se

$$P(E_i|F_j) = P(E_i) \text{ e} \tag{2.27}$$

$$P(F_j|E_i) = P(F_j) \tag{2.28}$$

para todo i e para todo j. Perceba que as condições (2.27) e (2.28) são equivalentes se $P(E_i) \neq 0$ e $P(F_j) \neq 0$.

Teorema 2.2.7 *Duas partições $\mathcal{E}$ e $\mathcal{F}$ são independentes com respeito a P se, e somente se, $\mathcal{E}$ e $\mathcal{F}$ são Jeffrey independentes com respeito a P e a* **todas** *as possíveis escolhas de $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$.*

Prova: Suponha $\mathcal{E}$ e $\mathcal{F}$ sejam independentes com respeito a P. Para todo $j \in J$ e pela condição anterior temos:

$$\begin{aligned} P_{\mathcal{E}}(F_j) &= \sum_{i \in I} P_{\mathcal{E}}(E_i) P(F_j|E_i) \\ &= \sum_{i \in I} p_i P(F_j) = P(F_j). \end{aligned}$$

Analogamente, temos para todo $i \in I$ que

$$P_{\mathcal{F}}(E_i) = P(E_i).$$

Portanto, $\mathcal{E}$ e $\mathcal{F}$ são Jeffrey independentes.

Por outro lado, suponha por absurdo que $\mathcal{E}$ e $\mathcal{F}$ não são independentes. Então existem i_0 e j_0 tais que

$$P(F_{j_0}|E_{i_0}) \neq P(F_{j_0}),$$

isto é, sem perda de generalidade, para i_0 considere

$$P(F_{j_0}|E_{i_0}) = P(F_{j_0}) + \delta,$$

onde δ é um valor não - nulo apropriado e suponha que para todo $i \neq i_o$

$$P(F_{j_0}|E_i) = P(F_{j_0}).$$

Considere que $p_{i_0} = 1 - \epsilon$ para $\epsilon > 0$ suficientemente pequeno onde $p_{i_o} = P_{\mathcal{E}}(E_{i_o}) = P_{\mathcal{E}\mathcal{F}}(E_{i_o})$. Desta forma,

$$\begin{aligned}
P_{\mathcal{E}}(F_{j_0}) &= \sum_i P_{\mathcal{E}}(E_i)P(F_{j_0}|E_i) \\
&= P_{\mathcal{E}}(E_{i_0})P(F_{j_0}|E_{i_0}) + \sum_{i \neq i_0} P_{\mathcal{E}}(E_i)P(F_{j_0}|E_i) \\
&= (P(F_{j_0}) + \delta)(1-\epsilon) + P(F_{j_0}) \sum_{i \neq i_0} p_i \\
&= \delta(1-\epsilon) + (1-\epsilon)P(F_{j_0}) + \epsilon P(F_{j_0}) \\
&= \delta(1-\epsilon) + P(F_{j_0}) \neq P(F_{j_0}),
\end{aligned}$$

contradizendo a hipótese.

Portanto, $\mathcal{E}$ e $\mathcal{F}$ são independentes.

□

Este teorema mostra que se Você verifica a Jeffrey independência de $\mathcal{E}$ e $\mathcal{F}$ com respeito as suas particulares declarações de probabilidades P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ esta constatação não garantirá independência de $\mathcal{E}$ e $\mathcal{F}$, no sentido usual.

Antes de exibirmos um exemplo que ilustre esse fato vamos considerar o seguinte algoritmo que permite verificar mais rapidamente a Jeffrey independência. Este algoritmo bem como os resultados apresentados nos teoremas 2.2.6 e 2.2.7 podem ser encontrados em Diaconis e Zabell(1982).

Considerando o Coeficiente de Dependência de Johnson, a saber

$$r_{ij} = \frac{P(E_i, F_j)}{P(E_i)P(F_j)}, \tag{2.29}$$

Diaconis e Zabell(1982) constatam que

$$\sum_i r_{ij}p_i = \frac{P_{\mathcal{E}}(F_j)}{P(F_j)} \text{ e} \sum_j r_{ij}q_j = \frac{P_{\mathcal{F}}(E_i)}{P(E_i)}. \tag{2.30}$$

Então para que $\mathcal{E}$ e $\mathcal{F}$ sejam Jeffrey independentes com respeito a P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ basta verificarmos que

$$\sum_i r_{ij} p_i = 1 = \sum_j r_{ij} q_j. \tag{2.31}$$

Usando o resultado acima facilmente verifica-se no seguinte exemplo que Jeffrey independência para particulares p_i's e q_j's não implicam em independência.

Exemplo 2.6: Considere a seguinte tabela:

	F_1	F_2	F_3	
E_1	1/4	1/8	1/8	1/2
E_2	1/8	0	1/8	1/4
E_3	1/8	1/8	0	1/4
	1/2	1/4	1/4	1

Perceba que $\mathcal{E}$ e $\mathcal{F}$ não são independentes pois

$$P(E_3, F_3) = 0 \neq \frac{1}{16} = P(E_3)P(F_3)$$

mas, considerando $P_{\mathcal{E}}(E_i) = p_i$ e $P_{\mathcal{F}}(F_j) = q_j$ constatamos que para $j = 1, 2, 3$

$$\sum_i r_{ij} p_i = 1 \Longleftrightarrow \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{2.32}$$

implicando que

$$\begin{aligned} p_1 &= p \\ p_2 &= p_3 = (1-p)/2, \end{aligned}$$

para algum $p \in [0, 1]$.

Similarmente, para $i = 1, 2, 3$, obtemos

$$\sum_j r_{ij} q_j = 1 \Longleftrightarrow q_1 = q \quad \text{e} \quad q_2 = q_3 = (1-q)/2. \tag{2.33}$$

para algum $q \in [0,1]$.

Então tomando por exemplo $p = 1/2$ e $q = 1/3$ temos que $\mathcal{E}$ e $\mathcal{F}$ são Jeffrey independentes.

Em particular, se pelo menos uma das partições possuem apenas dois elementos, Jeffrey independência para particulares p_i's e q_j's é sinônimo de independência, como pode ser visto no teorema a seguir.

Teorema 2.2.8 *Se pelo menos uma das partições $\mathcal{E}$ e $\mathcal{F}$ possuem apenas dois elementos, Jeffrey independência para P, $\{p_i, i \in I\}$ e $\{q_j, j \in J\}$ dadas é equivalente a independência com respeito a P.*

Prova: A condição suficiente é conseqüência imediata do teorema 2.2.7.

Para a condição necessária, considere $\mathcal{E} = \{E, \bar{E}\}$ com $P_{\mathcal{E}}(E) = p \neq P(E)$ e uma partição qualquer $\mathcal{F}$.

Suponha, por absurdo, que $\mathcal{E}$ e $\mathcal{F}$ não são independentes. Então existe pelo menos um j_0 tal que

$$P(E|F_{j_0}) \neq P(E).$$

Considere $k = \frac{P(E,F_{j_0})}{P(E)P(F_{j_0})}$, $k \neq 1$. Conseqüentemente, temos que

$$P(\bar{E}|F_{j_0}) = 1 - kP(E).$$

Como $\mathcal{E}$ e $\mathcal{F}$ são J-independentes, usando (2.31) obtemos para todo j

$$p\frac{P(E|F_j)}{P(E)P(F_j)} + (1-p)\frac{P(\bar{E}|F_j)}{P(\bar{E})P(F_j)} = 1.$$

Em particular, temos para $j = j_0$

$$pk + (1-p)\frac{1-kP(E)}{1-P(E)} = 1$$

$$\Rightarrow \quad p(k-1) = (k-1)P(E)$$

$$\Rightarrow \quad p = P(E)$$

$$\Rightarrow \quad P_{\mathcal{E}}(E) = P(E),$$

o que contradiz a hipótese.

Logo, $\mathcal{E}$ e $\mathcal{F}$ são independentes.

□

Esses últimos resultados nos permitem verificar a possibilidade de comutatividade com base na distribuição a priori P - se as partições $\mathcal{E}$ e $\mathcal{F}$ são independentes, no sentido usual, com respeito a P, atualizações sucessivas feitas via Regra de Jeffrey são comutativas.

Um outro aspecto interessante de atualizações sucessivas via Regra de Jeffrey é que, sobre certas hipóteses, atualizar sucessivamente considerando $\mathcal{E}$ e depois $\mathcal{F}$ fornece o mesmo resultado que uma atualização direta considerando uma partição $\mathcal{G} = \mathcal{E} \times \mathcal{F} = \{\{E_i, F_j\} : E_i \in \mathcal{E}, F_j \in \mathcal{F}\}$.

Considere P como sendo uma medida de probabilidade definida sobre $(\Omega, \mathcal{A})$. Sejam $P_{\mathcal{E}\mathcal{F}}$ a medida de probabilidade obtida atualizando P após receber $\mathcal{E}$ e sucessivamente $\mathcal{F}$ e $P_{\mathcal{G}}$ a medida de probabilidade obtida de P após o conhecimento de $\mathcal{G}$, ambas definidas sobre $(\Omega, \mathcal{A})$. Considere $P(E_i, F_j) \neq 0$ para todo $\{E_i, F_j\} \in \mathcal{G}$.

Teorema 2.2.9 *Suponha $P_{\mathcal{E}}(.|E_i) = P(.|E_i)$, $P_{\mathcal{E}\mathcal{F}}(.|F_j) = P_{\mathcal{E}}(.|F_j)$ e $P_{\mathcal{G}}(.|E_i, F_j) = P(.|E_i, F_j)$ para todo $E_i \in \mathcal{E}$ e para todo $F_j \in \mathcal{F}$. Se $P_{\mathcal{G}}(E_i, F_j) = P_{\mathcal{E}\mathcal{F}}(E_i, F_j)$ para todo $\{E_i, F_j\} \in \mathcal{G}$ então temos que $P_{\mathcal{G}} = P_{\mathcal{E}\mathcal{F}}$.*

Prova: Seja A um evento qualquer, $\mathcal{A}$- mensurável. Como a J-condição é suposta válida, sabemos que $P_{\mathcal{G}}$ é obtida de P via Regra de Jeffrey. Desta forma e usando as outras hipóteses obtemos:

$$P_{\mathcal{G}}(A) \quad = \quad \sum_{ij} P(A|E_i, F_j) P_{\mathcal{G}}(E_i, F_j)$$

$$= \sum_{ij} \frac{P(A, E_i, F_j)}{P(E_i, F_j)} P_{\mathcal{E}\mathcal{F}}(E_i, F_j)$$

$$= \sum_{ij} P(A, E_i, F_j) \frac{P_{\mathcal{E}}(E_i) P_{\mathcal{E}\mathcal{F}}(F_j)}{P(E_i) P_{\mathcal{E}}(F_j)} = P_{\mathcal{E}\mathcal{F}}(A).$$

Logo, $P_{\mathcal{G}} = P_{\mathcal{E}\mathcal{F}}$

□

Fazendo a conexão com suficiência e supondo que $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{G}}$, Jeffrey(1992) propõe

Teorema 2.2.10 *Se $\mathcal{E}$ é uma partição suficiente para $\{P, P_{\mathcal{E}}\}$ e se $\mathcal{F}$ é uma partição suficiente para $\{P_{\mathcal{G}}, P_{\mathcal{E}}\}$, então $\mathcal{G}$ é uma partição suficiente para $\{P_{\mathcal{G}}, P\}$.*

Prova: Seja A um evento qualquer pertencente a $\mathcal{A}$ e seja $\{E_i, F_j\} \in \mathcal{G}$. Usando as hipóteses verificamos que

$$\begin{aligned} P_{\mathcal{G}}(A|E_i, F_j) &= \frac{P_{\mathcal{G}}(A, E_i|F_j)}{P_{\mathcal{G}}(E_i|F_j)} \\ &= \frac{P_{\mathcal{E}}(A, E_i|F_j)}{P_{\mathcal{E}}(E_i|F_j)} \\ &= \frac{P(A, F_j|E_i)}{P(F_j|E_i)} = P(A|E_i, F_j). \end{aligned}$$

Portanto, $\mathcal{G}$ é suficiente para $\{P_{\mathcal{G}}, P\}$.

□

Perceba que sob as hipóteses acima ,

$$\frac{P_{\mathcal{G}}(E_i, F_j)}{P(E_i, F_j)} = \frac{P_{\mathcal{E}}(E_i|F_j) P_{\mathcal{G}}(F_j)}{P_{\mathcal{E}}(F_j|E_i) P(E_i)} = \frac{P_{\mathcal{G}}(F_j) P_{\mathcal{E}}(E_i)}{P_{\mathcal{E}}(F_j) P(E_i)}.$$

Ou seja, se para Você $\mathcal{E}$ é suficiente para $\{P, P_{\mathcal{E}}\}$ e $\mathcal{F}$ é suficiente para $\{P_{\mathcal{G}}, P_{\mathcal{E}}\}$, atualizações sucessivas utilizando a Regra de Jeffrey primeiro em $\mathcal{E}$ e depois em $\mathcal{F}$, fornecem o

mesmo resultado que uma única atualização com base na partição $\mathcal{G} = \mathcal{E} \times \mathcal{F}$. Note ainda que o julgamento de suficiência é tão subjetivo quanto a verificação da J-condição. Mais alguns outros comentários e discussões fundamentais sobre a Regra de Jeffrey podem ser encontrados em Jeffrey(1992) e Howson e Urback(1993).

Diaconis e Zabell(1982) exploram algumas outras propriedades matemáticas da Regra de Jeffrey. Uma proposta para atualização mecânica de probabilidade via Regra de Jeffrey - isto é, sem a necessidade de um teste subjetivo para a J-condição - surge ao estabelecerem que a medida de probabilidade $P_{\mathcal{E}}$ obtida de P através da Regra de Jeffrey é a medida mais próxima de P (em algum sentido), dentre todas as medidas Q que atribuem mesmo valor aos elementos da partição, isto é, dentre todas as medidas Q tais que $Q(E_i) = P_{\mathcal{E}}(E_i) = p_i$ para todo $E_i \in \mathcal{E}$. Diaconis e Zabell(1982) provam que esta medida $P_{\mathcal{E}}$ (dada pela Regra de Jeffrey) é a única medida satisfazendo a condição acima que minimiza a distância de Hellinger e o número de Kullback-Leibler, quando estes são calculados com relação a P. Além disto, $P_{\mathcal{E}}$ é uma das medidas minimizadoras da distância de Variação Total com respeito a P, quando consideramos todas as medidas Q tais que $Q(E_i) = P_{\mathcal{E}}(E_i)$ para todo $E_i \in \mathcal{E}$.

Baseado em resultados sobre I-projeção apresentados por Csiszár(1975), Diaconis e Zabell (1982) também apresentam um procedimento mecânico para atualizações sucessivas via Regra de Jeffrey com base no IPFP (Iterated Proportional Fitting Procedure). Esta técnica é muito usada para ajustar tabelas de contingência cujas marginais são fixas, partindo-se de uma tabela inicial conhecida. Csiszár(1975) prova que este procedimento de ajuste converge para a I-projeção da tabela inicial sobre o conjunto de todas as tabelas cujas marginais coincidem com as marginais consideradas. Desta forma, imaginando que as colunas formam a partição $\mathcal{E}$ e as linhas a partição $\mathcal{F}$ cujos elementos terão suas probabilidades retaxadas, obteremos mecanicamente as probabilidades obtidas via Regra de Jeffrey a qual, como vimos antes, é a I-projeção da distribuição *a priori* sobre o conjunto $\{Q : Q(E_i) = p_i, Q(F_j) = q_j, \forall i, j\}$.

Leitores interessados em maiores detalhes sobre atualizações mecânicas de probabilidades podem consultar a seção 5 de Diaconis e Zabell(1982) e também podem encontrar explicações sobre o IPFP em Fienberg(1970) e Csiszár(1975).

Nas próximas seções apresentaremos alguns resultados inéditos sobre a Regra de Jeffrey, explorando sua versão preditivista e uma definição para conjugação segundo a operação definida pela Regra de Jeffrey. O leitor também poderá encontrar uma aplicação prática da Regra de Jeffrey no capítulo 5 deste trabalho.

2.3 A Regra Preditivista de Jeffrey

Em muitas situações, nosso interesse está em fazer inferências sobre objetos observáveis porém desconhecidos considerando os resultados já observados de objetos julgados (de alguma forma) similares ao nosso objeto de interesse. Por exemplo, um médico ao fazer comparações da eficiência de dois tratamentos levando em conta a informação obtida da reação de seus pacientes a cada tratamento esta de fato interessado em saber qual tratamento aplicar ao próximo paciente que chegar no seu consultório para que este seja curado mais rapidamente.

Em situações como esta (que essencialmente é a mesmas vista nas seções anteriores salvo pela particular estrutura do espaço amostral envolvido - a saber, aqui $\Omega = \Omega_1 \times \Omega_2 \times \ldots$ onde Ω_i pode ou não ser um conjunto enumerável) também é possível ocorrerem informações não previamente consideradas, embora relevantes para a inferência, as quais deverão ser incorporadas na opinião do médico para que tome suas decisões utilizando toda a informação disponível (ver exemplo 2.2 na seção 2.2). Como antes, o uso da fórmula de Bayes para a obtenção da distribuição *a posteriori* neste caso nem sempre é viável o que, juntamente com esta a particular estrutura de Ω, nos motiva a explorar uma versão para a Regra de Jeffrey quando alguma condição de invariância é percebida na seqüência de quantidades aleatórias cujos valores serão usados para inferências. Ou seja, buscamos uma versão preditivista para a Regra de Jeffrey que nos permita obter a distribuição *a posteriori* diante de imprevistos em situações como a descrita acima. A esta regra chamaremos de Regra Preditivista de Jeffrey.

Por **Preditivismo** entende-se a atitude de subordinar a existência de parâmetros a algum julgamento de invariância sobre observáveis (Wechsler(1993)). Esta visão, sugerida por

de Finetti(1937,1972), reforça a idéia de que tanto a função de verossimilhança quanto a distribuição *a priori* têm o mesmo caráter subjetivo - ambas fazem parte de sua opinião preditiva sobre o processo. Considerando que o objetivo primordial de um pesquisador é fazer afirmações sobre objetos desconhecidos (por isso, aleatórios) porém observáveis, parece-nos razoável sugerir a postura preditivista como a mais adequada para fazer-se inferências. Uma discussão sobre estas idéias no contexto de populações finitas pode ser encontrado em Iglesias(1993) e para uma visão mais geral sobre preditivismo, sugerimos Bernardo e Smith(1994) e Wechsler(1993). (Ver também a seção 3.1).

Para tornar mais clara a associação da Regra de Jeffrey apresentada na seção anterior - onde Ω é simplesmente um conjunto enumerável sem assumir qualquer estrutura particular - com a Regra Preditivista de Jeffrey que exibiremos posteriormente na proposição 2.3.1 consideraremos a versão generalizada da Regra de Jeffrey, apresentada em Diaconis e Zabell(1982) e o Teorema de Representação de de Finetti (1937) para variáveis aleatórias permutáveis.

Ao longo desta e das próximas seções deste capítulo denotaremos por P^* à distribuição de probabilidade *a posteriori* obtida de P pela Regra de Jeffrey com respeito à σ - álgebra $\mathcal{A}_0$, onde P e P^* são medidas de probabilidade definidas sobre $(\Omega,\ \mathcal{A})$ as quais interpretaremos, respectivamente, como a sua opinião *a priori* e *a posteriori* sobre eventos $\mathcal{A}$-mensuráveis.

Diremos que P é dominada por P^* - abrevidamente, $P << P^*$ - se para todo $A \in \mathcal{A}$ tal que $P^*(A) = 0$, temos $P(A) = 0$.

Definição 2.3.1 *(Regra de Jeffrey Generalizada)*
Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Considere uma nova medida de probabilidade P^* definida sobre $\mathcal{A}$ e uma sub-σ-álgebra $\mathcal{A}_0$ de $\mathcal{A}$, $\mathcal{A}_0 \subseteq \mathcal{A}$. Seja C um conjunto $\mathcal{A}_0$-mensurável, tal que $P(C) = 0$ e suponha que $\bar{P} << \bar{P}^*$ para subconjuntos $\mathcal{A}_0$-mensuráveis de $\Omega - C$, onde $\bar{P}$ e $\bar{P}^*$ são restrições de P e P^* a $\mathcal{A}_0$, respectivamente. Se

$$\mathcal{A}_0 \text{ é suficiente para } \{P, P^*\}, \qquad \text{(J-condição)}$$

a Regra de Jeffrey torna-se

$$P^*(A) = \int_{\Omega - C} P(A|\mathcal{A}_0)(\omega) dP^*(\omega) + P^*(A, C), \quad \forall A \in \mathcal{A} \tag{2.34}$$

onde $P(A|\mathcal{A}_0)$ é a probabilidade condicional de A dado $\mathcal{A}_0$. Se $P^* << P$, pode-se tomar C como sendo o conjunto vazio.

Essa versão dada para a Regra de Jeffrey surge naturalmente da definição geral de probabilidade condicional, do julgamento de suficiência feito sobre $\mathcal{A}_0$ e por lembrar que C é um conjunto $\mathcal{A}_0$ mensurável. A existência deste conjunto C na estrutura em que foi definido vem ressaltar que a Regra de Jeffrey permite partir de probabilidades inicialmente consideradas nulas e obter, através de uma retaxação arbitrária, probabilidades *a posteriori* diferentes de zero por incluir em sua expressão o termo $P^*(A, C)$ - isto também ocorria quando Ω era apenas um conjunto enumerável.

Note ainda que considerar a dominação de $\bar{P}$ por $\bar{P}^*$ em subconjuntos de $\Omega - C$, $\mathcal{A}_0$-mensuráveis é exigir que uma vez que *a priori* Você considerou um evento como provável mesmo que com probabilidade baixa de ocorrência, *a posteriori* Você também deve atribuir uma probabilidade não nula a este evento.

Como vimos na seção anterior, a definição 2.3.1 estabelece que quando a informação recebida faz com que Você mude sua opinião inicial P sobre alguns elementos de $\mathcal{A}$ através de uma retaxação arbitrária (por exemplo) sem alterar a opinião condicional nestes elementos, a Regra de Jeffrey diz como você deve atualizar sua opinião sobre os demais eventos de $\mathcal{A}$.

Para atingir o objetivo proposto, além da definição 2.3.1, considere a seguinte versão do Teorema de Representação de de Finetti(1937) para variáveis aleatórias permutáveis.

Teorema 2.3.1 *(Teorema Geral de Representação) Seja $X_1, X_2, \ldots$ uma seqüência infinita de variáveis aleatórias permutáveis e com lei P. Então, existe uma medida de probabilidade Q definida sobre $\mathcal{F}$ - o espaço de todas as funções de distribuições em $\mathcal{R}$ - tal que para todo n,*

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \Pi_{i=1}^n F(x_i) dQ(F) \tag{2.35}$$

onde Q é a P-lei da quantidade aleatória $F = \lim_{n\to\infty} F_n$ e $F_n = \{F_n(x); -\infty < x < \infty\}$ onde para cada x

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}. \tag{2.36}$$

A prova desse teorema será omitida e pode ser encontrada, por exemplo, em Chow e Teicher(1988) e Iglesias(1993). Da representação (2.35) podemos interpretar $\Pi_{i=1}^{n} F(x_i)$ como a P-lei de n variáveis aleatórias $X_1, \ldots, X_n$ as quais são independentes e identicamente distribuídas dado o objeto aleatório F e F pode ser vista como um parâmetro infinito-dimensional com P-lei Q.

Denote por $\sigma(A)$ a σ-álgebra gerada pelo objeto aleatório A e por $\mathcal{B}^\infty$ a σ-álgebra de Borel definida sobre $\mathcal{R}^\infty$.

Proposição 2.3.1 *Seja $\mathcal{P}$ uma família de medidas de probabilidades definidas sobre $(\mathcal{R}^\infty, \mathcal{B}^\infty)$, permutáveis. Sejam P e P^* elementos de $\mathcal{P}$. Então, P^* é obtida através da Regra de Jeffrey de P, com respeito a σ-álgebra $\mathcal{A}_0 = \sigma(F)$, onde $F = \lim_{n\to\infty} F_n$ e F_n é como o definido na expressão (2.36).*

Prova: O Teorema da Representação de de Finetti (1937), visto no teorema 2.3.1 estabelece que para cada medida de probabilidade $P' \in \mathcal{P}$ que faz da seqüência infinita de q.a. $X_1, X_2, \ldots$ permutável, existe uma medida de probabilidade Q' no espaço de funções de distribuições $\mathcal{F}$ definidas sobre $\mathcal{R}$ tal que para qualquer $n \geq 1$,

$$P'(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \Pi_{i=1}^{n} F(x_i) dQ'(F)$$

onde Q' é a P'-lei do objeto aleatório F anteriormente definido.

Sendo assim, para cada função de $F \in \mathcal{F}$

$$\begin{aligned} \Pi_{i=1}^{n} F(x_i) &= P'(X_1 \leq x_1, \ldots, X_n \leq x_n | F) \\ &= P'(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(\mathbf{x}), \quad \forall\, P' \in \mathcal{P} \quad \text{e} \quad \forall\, \mathbf{x} \in \mathcal{R}^\infty. \end{aligned} \tag{2.37}$$

Note da expressão (2.37) que F é uma estatística suficiente para a família de medidas de probabilidades que faz da seqüência $X_1, X_2, \ldots$ permutável. Então, F é uma estatística

suficiente para $\{P, P^*\}$ e, conseqüentemente, temos que $\mathcal{A}_0$ é uma σ–álgebra suficiente para $\{P, P^*\}$.

Portanto, sob permutabilidade a J–condição para $\{P, P^*\}$ é verificada e fica garantida a existência de uma medida de probabilidade Q^* definida sobre $\mathcal{F}$ tal que,

$$P^*(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} P(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(\mathbf{x})\, dQ^*(\mathbf{x}), \tag{2.38}$$

para todo $n \geq 1$, onde Q^* é uma distribuição marginal de P^*. Além disto, de (2.38) vê-se que $P^* << P$, concluíndo a prova da proposição.

□

É possível mostrar que $\mathcal{A}_0$ é essencialmente a σ-álgebra de eventos permutáveis (ver Iglesias(1993)).

Note que quando $X_1, X_2, \ldots$ é permutável, tanto sob sua opinião *a priori* P quanto sob sua opinião *a posteriori* P^* a sua opinião atualizada através da Regra de Jeffrey é dada por

$$P^*(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \Pi_{i=1}^n F(x_i) dQ^*(F), \quad \forall n \geq 1 \tag{2.39}$$

onde $\mathcal{F}$, F e Q^* são definidas na prova da proposição anterior e Q^* pode ser interpretada como a sua opinião sobre F reavaliada após o recebimento de alguma informação não previamente considerada. À expressão (2.39) chamaremos de **Regra Preditivista de Jeffrey** para medidas de probabilidade permutáveis.

O teorema anterior nos faz perceber que quando o julgamento de permutabilidade feito *a priori* é mantido *a posteriori*, após o recebimento de alguma informação não previamente antecipada, a Regra de Jeffrey é sempre possível uma vez que a existência de uma σ–álgebra suficiente é sempre garantida. Esta possibilidade estende-se para qualquer condição de invariância que gere um teorema de representação do tipo de Finetti como por exemplo os teoremas estabelecidos por Kingman(1972), Smith(1981) e Iglesias(1993). (Uma revisão sobre

modelagem sob condições de invariância pode ser encontrada em Bernardo e Smith(1993)). Nota-se também que quando algum julgamento de invariância é feito sobre a seqüência de quantidades aleatórias, a Regra de Jeffrey nada mais é que uma mistura da lei de quantidades aleatórias independentes e identicamente distribuídas pela medida misturadora obtida da medida misturadora inicialmente considerada através de uma mudança exógena. Por exemplo, quando uma seqüência infinita de variáveis aleatórias assumindo valores em $\{0,1\}$ é considerada e definindo $\mathcal{P}$ como a família de medidas de probabilidade P que fazem da seqüência $X_1, X_2, \ldots$ permutáveis, o Teorema de Representação de de Finetti estabelece que a função de probabilidade de qualquer sub-seqüência $X_1, \ldots, X_n$, $n \geq 1$, pode ser representada como

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_{\Theta} \Pi_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta) \tag{2.40}$$

onde Q é a P-lei da variável aleatória $\theta = \lim_{n\to\infty} \sum_{i=1}^{n} X_i/n$. Neste caso, $\mathcal{A}_0 = \sigma(\theta)$; $P(X_1 = x_1, \ldots, X_n = x_n | \mathcal{A}_0)(\mathbf{x}) = \Pi_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$ é um produto de distribuições Bernoulli com parâmetro θ e a Regra de Jeffrey dada pela expressão (2.39) poderá ser obtida da expressão (2.40) quando Q for substituída pela probabilidade atualizada arbitrariamente Q^* .

Exemplo 2.7: Um dos procedimentos cirúrgicos que vem sendo realizado comumente no Hospital das Clínicas é o transplante de fígado. Para evitar a rejeição do órgão transplantado os médicos têm usado a Droga A e desde o início observa-se uma variável aleatória X_i a qual recebe respectivamente os valores um ou zero se ocorre ou não uma rejeição do órgão transplantado no paciente i tendo ele recebido a droga A, onde $i \geq 1$. Com base em toda a sua experiência, Você (o médico responsável pela realização dos transplantes desta semana) acha que a chance de haver uma rejeição tendo tomado a droga A é um certo valor θ que Você não sabe precisar, mas sabe que é um valor pequeno que se comporta segundo uma distribuição Q. Além disto, Você acha que as X_i's são permutáveis.

Amanhã, serão realizados três transplantes e sob as considerações feitas sua opinião para

que ocorra rejeição nas duas primeiras cirurgias é representada da seguinte forma

$$P(X_1 = 1, X_2 = 1, X_3 = 0) = \int_0^1 (1-\theta)\theta^2 dQ(\theta),$$

Amanhece o dia e antes que Você entre no centro cirúrgico a enfermeira chefe comenta que chegou um comunicado oficial do laboratório produtor da droga A informando que o último lote da droga A enviado ao HC está fora dos padrões exigidos pela OMS. Houve uma falha na produção deste lote, só agora detectado, o que compromete a eficiência da droga A.

Ouvindo este comentário, Você passa a achar que uma vez que a eficiência da droga A está diminuida a chance atribuída inicialmente à rejeição aumenta e, portanto, o comportamento de θ passa a ser descrito por uma nova medida de probabilidades Q^*. Também nota-se que o recebimento desta informação não altera o julgamento inicial de permutabilidade que Você fazia sobre a seqüência $X_1, X_2, \ldots$. Desta forma, a probabilidade *a posteriori* para a ocorrência de duas rejeições nos transplantes que Você realizará daqui a pouco pode ser reavaliada através da Regra Preditivista de Jeffrey da seguinte forma

$$P^*(X_1 = 1, X_2 = 1, X_3 = 0) = \int_0^1 (1-\theta)\theta^2 dQ^*(\theta).$$

Perceba que graças a simplicidade das variáveis aleatórias envolvidas no exemplo anterior a função de verossimilhança fica completamente determinada e além disso aplicar a Regra de Jeffrey consiste apenas em reavaliar a medida de probabilidade atribuída inicialmente a θ, o qual neste particular exemplo é um objeto aleatório unidimensional.

Vimos no teorema 2.3.1 que, sob permutabilidade apenas, o parâmetro F envolvido na representação pode ser um objeto infinito dimensional e portanto a aplicação da Regra de Jeffrey exigiria a reavaliação da probabilidade atribuída inicialmente a este objeto, o que pode ser uma tarefa mental bastante complexa e algumas vezes humanamente impossível de ser realizada satisfatoriamente. Uma das maneiras de contornar este problema é impor condições adicionais à permutabilidade de tal forma que a medida misturadora do teorema 2.3.1 possa ser definida em um suporte mais simples, facilitando com isso as possíveis reavaliações das probabilidades iniciais. Para englobar estas situações bem como aquelas em que o manuseio

com funções de densidade é mais conveniente, apresentaremos uma reformulação do teorema 2.3.1 em termos de densidades e consideraremos um espaço paramétrico com dimensão finita.

Sejam Θ um espaço paramétrico finito-dimensional e $\mathcal{F} = \{F(.|\theta) : \theta \in \Theta\}$ a família de funções de distribuição definidas sobre $\Theta \subseteq \mathcal{R}$. Sob estas considerações o teorema de Representação de de Finetti para variáveis aleatórias permutáveis estabelece que existe uma medida de probabilidade Q definida sobre Θ tal que

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta \Pi_{i=1}^n F(x_i|\theta) dQ(\theta). \tag{2.41}$$

para qualquer sub-seqüência finita $X_1, \ldots, X_n$, $n > 0$, de uma seqüência infinita $X_1, X_2, \ldots$ de variáveis aleatórias permutáveis.

Defina uma função ψ de $\mathcal{F}$ em Θ de tal forma que $\psi(F(.|\theta)) = \theta$ e seja Q_ψ a lei de ψ. Considere P_F a medida de probabilidade induzida por $F \in \mathcal{F}$.

Corolário 2.3.1 *Suponha válidas as condições do teorema 2.3.1. Se P_F é dominada por uma medida σ-finita λ e se Q_ψ tem suporte em $\mathcal{F}$ tal que $Q_\psi(\mathcal{F}) = 1$, então*

$$p(x_1, \ldots, x_n) = \int_\Theta \Pi_{i=1}^n f(x_i|\theta) dQ_\psi(\theta), \tag{2.42}$$

onde $p = \frac{dP}{d\lambda}$ e $f(.|\theta) = \frac{dP_F(.|\theta)}{d\lambda}$

Prova: Por hipótese, $P_F(.|\theta) << \lambda$ para cada $\theta \in \Theta$. Então pelo teorema 2.3.1 temos que $P << \lambda$ e pelo teorema de Radon-Nikodin existem funções $f(.|\theta)$ e p tais que

$$p = \frac{dP}{d\lambda} \quad \text{e} \tag{2.43}$$

$$f(.|\theta) = \frac{dP_F(.|\theta)}{d\lambda}. \tag{2.44}$$

Queremos mostrar que

$$\int_\Theta f(x_1, \ldots, x_n|\theta) dQ_\psi(\theta) \tag{2.45}$$

é uma versão da derivada de Radon-Nikodin de P com respeito a λ.

Para isso considere $A \in \mathcal{B}^\infty$, onde $\mathcal{B}^\infty$ é a σ-álgebra de Borel de $\mathcal{R}^\infty$. Usando o teorema de Fubini e a expressão (2.44) temos:

$$\begin{aligned}\int_A \int_\Theta f(\mathbf{x}|\theta) dQ_\psi(\theta) d\lambda(\mathbf{x}) &= \int_\Theta \int_A f(\mathbf{x}|\theta) d\lambda(\mathbf{x}) dQ_\psi(\theta) \\ &= \int_\Theta P_F(A|\theta) dQ_\psi(\theta) \\ &= P(A) \end{aligned} \tag{2.46}$$

(Esta última igualdade segue pelo teorema 2.3.1, uma vez que $Q_\psi(\mathcal{F}) = 1$.)

Novamente, o teorema de Radon-Nikodin fornece que

$$\int_\Theta f(\mathbf{x}|\theta) dQ_\psi(\theta) = \frac{dP}{d\lambda}(\mathbf{x}). \tag{2.47}$$

Segue daí e de (2.43) a conclusão do resultado.

□

Exemplo 2.8: Seja $\mathcal{F} = \{F(.|\sigma^2) : \sigma^2 \in \mathcal{R}_+\}$ onde F é a função de distribuição acumulada de uma Normal com média zero e variância σ^2. Seja Q_ψ é a lei de ψ, onde ψ é definida como segue

$$\begin{aligned} \psi \ : \quad \mathcal{F} \quad &\longrightarrow \quad \mathcal{R}_+ \\ F(.|\sigma^2) \quad &\longrightarrow \quad \sigma^2, \end{aligned}$$

e suponha que $Q_\psi(\mathcal{F}) = 1$. Ou seja, *a priori* Você está certo que a função de verossimilhança é alguma distribuição normal com média zero e variância desconhecida.

Seja λ a medida de Lebesgue e P_F a medida induzida por F dominada por λ. Então,

$$\frac{dP_{F(.|\sigma^2)}}{d\lambda}(x) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}.$$

Por fim, usando o corolário anterior concluímos que para todo $n \geq 1$ a função de densidade conjunta de $X_1, \ldots, X_n$ é

$$p(x_1, \ldots, x_n) = \int_0^\infty \Pi_{i=1}^n \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\} dQ_\psi(\sigma^2). \tag{2.48}$$

Kingman(1972) estabelece a representação exibida na expressão (2.48) do exemplo anterior partindo de uma definição de invariância mais forte que permutabilidade - a invariância sob transformações ortogonais (ver cap.3). Neste caso, para obter a representação fornecida em (2.48) não foi necessário qualquer suposição adicional sobre a lei de ψ, Q_ψ, como a imposta aqui - a saber, $Q_\psi(\mathcal{F}) = 1$ - a qual depõe contra o caráter preditivista e operacional da representação. Outras representações de caráter preditivista como a estabelecida por Kingman(1972) podem ser encontradas em Smith(1981), Diaconis *et alii*(1992), Iglesias(1993) e Silva(1996), dentre outros.

Considerando o caso onde as quantidades aleatórias de interesse são contínuas, Diaconis e Freedman (1990) demonstram um teorema de representação geral que permite identificar as funções de verossimilhança envolvidas na representação quando estas são provenientes da família exponencial paramétrica. Este resultado fornece uma Regra Preditivista de Jeffrey particular para as situações em que as quantidades aleatórias em estudo puderem ser representadas como misturas de elementos da família exponencial.

Considere h uma função de Borel, localmente integrável, não-negativa e finita definida em $\mathcal{R}^d$, onde $\mathcal{R} = (-\infty, \infty)$ e d é um inteiro positivo fixo. Seja

$$D(\theta) = \int_{\mathcal{R}^d} \exp\{x \cdot \theta\} h(x) dx \quad \text{para} \quad \theta \in \mathcal{R}^d \tag{2.49}$$

onde $x \cdot \theta$ denota o produto interno de x por θ e considere o seguinte conjunto

$$\Theta = \{\theta \in \mathcal{R}^d : D(\theta) < \infty\}. \tag{2.50}$$

Suponha que Θ seja um conjunto não-vazio e convexo.

Consideremos como membro da família exponencial, qualquer medida de probabilidade P definida em $\mathcal{R}^d$, cuja função de densidade é da forma

$$dP_\theta(x) = \frac{\exp\{x \cdot \theta\} h(x)}{D(\theta)} dx, \quad \forall \theta \in \Theta \subseteq \mathcal{R}^d. \tag{2.51}$$

Sejam $X_1, X_2, \ldots$ uma seqüência de quantidades aleatórias assumindo valores em $\mathcal{R}^d$ e $S_n = X_1 + \ldots + X_n$. Denote por P_θ^∞ a medida de probabilidade definida sobre $(\mathcal{R}^d)^\infty$

sob a qual as variáveis aleatórias X_i's são independentes e identicamente distribuídas com distribuição P_θ.

Considere $C_n = \{s \in \mathcal{R}^d : 0 < h^{(n)}(s) < \infty\}$ onde $h^{(j)}$ é a j-ésima convolução de h com si mesmo. Tome $t = x_1 + \ldots + x_k$, $k < n$ e escolha $s \in C_n$. Então, a distribuição condicional de $X_1, \ldots, X_k | S_n = s$ - denotada por Q_{nsk} - é absolutamente contínua em $\mathcal{R}^d$ com densidade

$$\frac{h(x_1)\ldots h(x_k)h^{(n-k)}(s-t)}{h^{(n)}(s)}$$

e se $k = n$, Q_{nsk} é singular.

Assim definida, Q_{nsk} é a distribuição condicional regular de $X_1, \ldots, X_k | S_n = s$ relativa a P_θ^∞, para todo $\theta \in \Theta$, estabelecendo que S_n é uma estatística suficiente para θ.

Defina $\mathcal{M}_Q$ como a família de todas as medidas de probabilidade P definidas em $(\mathcal{R}^d)^\infty$, absolutamente contínuas, tal que para todo $n \geq 1$,

(i) $P(S_n \in C_n) = 1$

(ii) Q_{nsn} é a P-lei condicional de $X_1, \ldots, X_n | S_n = s$.

Informalmente, $\mathcal{M}_Q$ é a família de medidas de probabilidade definidas em $(\mathcal{R}^d)^\infty$ tais que para cada $n \geq 1$ as distribuições n-dimensionais $P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ têm S_n como estatística suficiente e Q_{nsn} como a lei de $X_1, \ldots, X_n | S_n = s$.

Teorema 2.3.2 *(Diaconis e Freedman(1990))*

Suponha válidas as considerações feitas sobre h e assuma as condições (2.49) e (2.50). Então, $P \in \mathcal{M}_Q$ se, e somente se,

$$P = \int_\Theta P_\theta^\infty dQ(\theta). \tag{2.52}$$

Além disto, a medida de probabilidade Q é definida unicamente sobre subconjuntos de Borel do conjunto convexo $\Theta \subseteq \mathcal{R}^d$.

Esse teorema garante que se uma seqüência infinita de quantidades aleatórias $X_1, X_2, \ldots$ tem a mesma distribuição condicional que uma quantidade aleatória com distribuição P_θ na família exponencial indexada por θ então, toda sub-seqüência $X_{(n)} = (X_1, \ldots, X_n)$ tem sua distribuição representada pela mistura de n variáveis aleatórias independentes e identicamente distribuídas com distribuição P_θ. Ou seja, para todo $n \geq 1$ a distribuição de qualquer subseqüência $X_{(n)}$ de tamanho n é representada por

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta P_\theta^n(X_1 \leq x_1, \ldots, X_n \leq x_n) dQ(\theta), \tag{2.53}$$

onde P_θ^n é a lei que faz de $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição P_θ.

Perceba que, como para o teorema 2.3.1, o teorema 2.3.2 visto acima pode ser traduzido em termos de densidades.

Corolário 2.3.2 *Sob as condições do teorema 2.3.2 e supondo que para cada $\theta \in \Theta$, P_θ é dominada por uma medida λ, σ-finita e positiva, temos que para todo $n \geq 1$,*

$$p(x_1, \ldots, x_n) = \int_\Theta p_\theta(x_1, \ldots, x_n) dQ(\theta), \tag{2.54}$$

onde $p = \frac{dP}{d\lambda}$ e $p_\theta = \frac{dP_\theta}{d\lambda}$

Prova: Será omitida por ser similar à prova do corolário 2.3.1.

Para tornar mais claras estas considerações veja o seguinte exemplo.

Exemplo 2.9: Seja $X_1, X_2, \ldots$ uma seqüência infinita de variáveis aleatórias positivas tais que para todo $k \geq 1$

$$P(X_{(k)} \in A) = P(X_{(k)} \in A + \mathbf{x}), \tag{2.55}$$

onde A é um boreliano qualquer de $\mathcal{R}_+^k$ e $\mathbf{x} \in \mathcal{R}^k$ é tal que $\sum_{i=1}^k x_i = 0$ e $A + \mathbf{x} \in \mathcal{R}_+^k$. Considere $S_n = X_1 + \ldots + X_n$. Então

$$X_1, \ldots, X_k | S_n = s \sim \mathcal{U}\left\{(x_1, \ldots, x_k) \in \mathcal{R}_+^k : \sum_{i=1}^k x_i = s\right\} = Q_{nsk}, \quad s > 0. \tag{2.56}$$

Q_{nsk} também é a lei de $Y_1, \ldots, Y_k | Y_1 + \ldots + Y_n = s$ quando as Y_i's são variáveis aleatórias independentes e identicamente distribuídas com distribuição exponencial $\mathcal{E}(\theta)$, $\theta > 0$, cuja densidade

$$dP_\theta(x) = \exp\{-\theta x + log(\theta)\} 1_{\{x>0\}} dx \tag{2.57}$$

pertencente à família exponencial, com $D(\theta) = 1/\theta$ e $\Theta = (0, \infty)$.

Note que $h(x_i) = 1_{\{x_i>0\}}$ para todo i e, conseqüentemente, a densidade induzida pela medida de probabilidade Q_{nsk} é

$$f_{Q_{nsk}} = \frac{h^{(n-k)}(s-t)}{h^{(n)}(s)} = \frac{(n-1)!}{(n-k-1)!} \frac{(s - \sum_{i=1}^k x_i)^{n-k-1}}{s^{n-1}}$$

A densidade $f_{Q_{nsk}}$ não depende de θ e por isso S_n é uma estatística suficiente para θ. Além disto, $C_n = (0, \infty)$ fazendo com que $P(S_n \in C_n) = 1$. Então, o corolário 2.3.2 garante a existência de uma medida de probabilidade Q construída sobre os borelianos de $\mathcal{R}_+$, tal que para todo $n \geq 1$

$$\begin{aligned} p(x_1, \ldots, x_n) &= \int_0^\infty \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\} dQ(\theta) \\ &= \int_0^\infty \Pi_{i=1}^n \theta \exp\{-\theta x_i\} dQ(\theta), \end{aligned} \tag{2.58}$$

ou seja, qualquer seqüência de variáveis aleatórias satisfazendo (2.55) é representada como uma mistura de distribuições exponencias com parâmetro finito dimensional pela distribuição *a priori* procedente.

O teorema de Diaconis e Freedman(1990) visto no teorema 2.3.2, embora permita reconhecer a função de verossimilhança em várias situações interessantes, é ineficiente para este fim quando a estatística suficiente S_n tem distribuição singular. Para esta situação Diaconis e Freedman(1990) exibem outro teorema utilizando-se da estrutura de espaços poloneses. Ao longo deste capítulo consideraremos apenas a versão mais simples do teorema de Diaconis e Freedman(1990) dado no teorema 2.3.2.

Proposição 2.3.2 *Suponha válidas as hipóteses do teorema 2.3.2. Sejam P e P^* duas*

medidas de probabilidade pertencentes à $\mathcal{M}_Q$. Então P^* é obtida de P através da Regra de Jeffrey, com respeito a sub- σ-álgebra $\mathcal{A}_0$ gerada por $\lim_{n\to\infty} S_n/n$.

Prova: Para qualquer $P' \in \mathcal{M}_Q$ o teorema 2.3.2 diz que existe uma única medida de probabilidade Q' definida sobre os borelianos de $\Theta \subseteq \mathcal{R}^d$, tal que para todo $n \geq 1$

$$P'(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta P_\theta^n(X_1 \leq x_1, \ldots, X_n \leq x_n) dQ'(\theta). \tag{2.59}$$

onde P_θ^n é a P-lei definida em $(\mathcal{R}^d)^n$ que faz das quantidades aleatórias X_i's independentes e com distribuição comum P_θ. Considerando este fato e a expressão (2.59) obtemos

$$P'(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta \Pi_{i=1}^n P_\theta(X_i \leq x_i) dQ'(\theta). \tag{2.60}$$

onde, para cada $\theta \in \Theta$, $\Pi_{i=1}^n P_\theta(X_i \leq x_i) = P'(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(x_1, x_2, \ldots)$ para todo $P' \in \mathcal{M}_Q$ e $\mathcal{A}_0 = \sigma(\theta) = \sigma(\lim_{n\to\infty} S_n/n)$. De fato, $\mathcal{A}_0$ é a σ-álgebra gerada pela interseção das σ-álgebras geradas por $\{S_n, X_{n+1}, X_{n+2}, \ldots\}$, $n \geq 1$.

Dessa forma, $\mathcal{A}_0$ é uma sub- σ-álgebra suficiente para $\mathcal{M}_Q$ e, particularmente, é suficiente para $\{P, P^*\}$. Com isto,

$$\begin{aligned} P^*(X_1 \leq x_1, \ldots, X_n \leq x_n) &= \int_{\mathcal{R}^{d\infty}} P'(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(x_1, x_2, \ldots) dP^*(x_1, x_2, \ldots) \\ &= \int_\Theta \Pi_{i=1}^n P_\theta(X_i \leq x_i) dQ^*(\theta). \end{aligned} \tag{2.61}$$

Notando da expressão (2.61) que P^* é dominada P, concluímos a prova do teorema.

□

Assim como na proposição 2.3.1, a expressão (2.61) fornece uma versão preditivista para a Regra de Jeffrey, agora permitindo identificar a forma funcional da função de verossimilhança.

Exemplo 2.10: Considere $X_1, X_2, \ldots$ uma seqüência constituída pelos tempos de vida das lâmpadas produzidas por uma indústria. De acordo com sua experiência passada, o dono desta indústria afirma em uma entrevista para a imprensa que para cada $i \geq 0$

$$X_i|\theta \sim \exp(\theta)$$

onde θ é a taxa de falha associada ao processo de produção de lâmpadas implantados na indústria.

O dono da indústria não conhece θ, mas afirma que em condições normais θ é muito pequeno e se comporta de acordo com uma distribuição de probabilidade Q.

Além disso, se uma amostra de n lâmpadas é escolhida e tempo de vida total desta amostra é um valor t conhecido, o dono da indústria confessa desconhecer qual amostra de tamanho n foi selecionada, ou seja, para ele

$$X_1, \ldots, X_n|T_n = t \sim \mathcal{U}\left\{(x_1, \ldots, x_n) \in \mathcal{R}_+ : \sum_{i+1}^{n} x_i = t\right\}. \tag{2.62}$$

Desta forma, temos por (2.58) que

$$P(x_1, \ldots, x_n) = \int_0^\infty \Pi_{i=1}^n \theta \exp\{-\theta x_i\}\, dQ(\theta), \; n \geq 1.$$

Antes do fim da entrevista, chega o secretário informando que foi descoberta uma falha na linha de produção e que as lâmpadas produzidas nos últimos quinze dias estão com sua qualidade comprometida. Esta informação (surpreendente para o dono da empresa pois ele havia feito recentemente um alto investimento na troca das máquinas de sua indústria por máquinas mais modernas) faz com que reavalie sua opinião inicial para a taxa de falha θ e ele se vê obrigado a declarar na entrevista que a taxa de falha das lâmpadas produzidas nos últimos quinze dias é maior do que o previamente declarado. Segundo ele a taxa de falha agora comporta-se segundo uma nova distribuição de probabilidade Q^* a qual favorece a valores maiores de θ. Sendo assim, sua nova opinião sobre o comportamento dos tempos de vida das n lâmpadas selecionadas pode ser obtida pela Regra Preditivista de Jeffrey como segue:

$$P^*(x_1, \ldots, x_n) = \int_0^\infty \Pi_{i=1}^n \theta \exp\{-\theta x_i\}\, dQ^*(\theta), \; n \geq 1.$$

O exemplo acima e o exemplo 2.7 ilustram situações onde informações recebidas de maneira imprevista modificam sua opinião *a priori* de forma arbitrária apenas parcialmente. Em ambos exemplos a condição de invariância percebida inicialmente foi mantida e por isso o uso da Regra Preditivista de Jeffrey para estender as modificações ocorridas em sua opinião inicial para todos os eventos de interesse pode ser usada.

Os resultados apresentados nesta seção serão posteriormente utilizados no capítulo 5 na formação de um algoritmo que permite incorporar informações não previamente consideradas para fazer previsões. A Regra Preditivista de Jeffrey também será conectada ao conceito de Jeffrey Conjugação que apresentaremos na próxima seção.

2.4 A Jeffrey Conjugação

Um conceito bastante explorado e também de utilidade prática estabelecido em inferência Bayesiana é o conceito de **conjugação**. Ao estabelecer-se o conceito de família de distribuições *a priori* conjugadas, pretendia-se determinar condições sobre a família de distribuições de probabilidade $\mathcal{P}$, definidas sobre um espaço paramétrico finito-dimensional Θ, de tal forma que a distribuição *a posteriori* fosse obtida de alguma distribuição *a priori* pertencente a $\mathcal{P}$ através do Condicionamento Bayesiano sem grandes esforços computacionais.

Inicialmente, definiu-se como uma família conjugada para um certo parâmetro θ, com respeito à função de verossimilhança $l(x|\theta)$, aquela família de distribuições de probabilidade $\mathcal{P}$ que continha tanto a distribuição *a priori* P para θ, quanto a distribuição *a posteriori* P^* obtida atualizando-se P através do Condicionamento Bayesiano. Essa definição, como se sabe, é muito abrangente e não responde à questão inicial de tornar mais imediato o cálculo da distribuição *a posteriori*. Por exemplo, considerando $\mathcal{P}$ como a família de todas as distribuições possíveis sobre θ, qualquer que seja a distribuição *a posteriori* obtida, esta pertencerá a $\mathcal{P}$.

Para atingir o objetivo proposto, mais restrições foram impostas sobre a família $\mathcal{P}$, está-

belecendo o conceito de **Conjugação Natural**. Este conceito é apresentado formalmente na seguinte definição.

Definição 2.4.1 *(Família Conjugada Natural)*

Seja $\mathcal{P}$ uma família de distribuições de probabilidade definidas sobre um espaço paramétrico Θ. Dizemos que $\mathcal{P}$ é uma família conjugada natural com respeito a uma função de verossimilhança $l(x, \theta)$ possuindo estatística suficiente $t(x)$, com dimensão fixa e independente de x, se:

(i) *a função de densidade ou função de probabilidade p associada a algum $P \in \mathcal{P}$ é proporcional a $l(\tau, \theta)$, onde τ tem a mesma dimensão que t, e é tal que $\int_\Theta l(\tau, \theta)p(\theta)d\theta < \infty$;*

(ii) *$\mathcal{P}$ é fechada por amostragem, isto é, a distribuição a priori $P \in \mathcal{P}$ e a distribuição a posteriori obtida de P através do Condicionamento Bayesiano pertencem a $\mathcal{P}$.*

Esta definição de conjugação provoca uma redução no número de elementos de $\mathcal{P}$ sem no entanto diminuir demasiadamente o número de escolhas possíveis para a distribuição *a priori*. Além disso, a distribuição *a posteriori* obtida difere da distribuição *a priori* apenas através dos parâmetros e, portanto, é facilmente determinável.

Ressaltamos que o uso de distribuições *a priori* conjugadas naturais não é obrigatório. Distribuições *a priori* devem refletir o seu conhecimento sobre θ e, por isso, podem ser quaisquer distribuições de probabilidade que descrevam adequadamente este conhecimento. Se por ventura a opinião declarada por Você for um membro da família conjugada natural, Você é premiado com uma distribuição *a posteriori* determinada sem muitos cálculos.

Mais detalhes sobre Conjugação e Conjugação Natural podem ser encontrados em DeGroot (1970), Bernardo e Smith(1994) e Gamerman e Migon(1993).

Como traduzir estes conceitos para aquelas situações onde o mecanismo escolhido para fazer as atualizações de probabilidade é a Regra de Jeffrey? Como retaxações são possíveis, não existe uma função de verossimilhança que atualize completamente sua opinião *a priori* como tínhamos antes. Desta forma, podemos pensar em uma conjugação de acordo com

Regra de Jeffrey - **Jeffrey Conjugação** ou simplesmente **J-Conjugação** - sempre que a informação recebida inesperadamente permitir uma atualização da distribuição *a priori* P para a distribuição *a posteriori* P^* através da Regra de Jeffrey de tal forma que P e P^* pertençam à mesma família de distribuições de probabilidade. Formalmente, definimos uma família de distribuições J-Conjugadas da seguinte maneira:

Definição 2.4.2 *(Família Jeffrey Conjugada ou J-Conjugada)*

Sejam $\mathcal{P}$ uma família de distribuições de probabilidade definidas sobre um espaço mensurável $(\Omega, \mathcal{A})$ e $\mathcal{A}_0$ uma sub - σ - álgebra de $\mathcal{A}$. Seja $P \in \mathcal{P}$ e considere P^* uma medida de probabilidade qualquer definida em $(\Omega, \mathcal{A})$ dominada por P. Suponha que P^* seja obtida de P via Regra de Jeffrey com relação a $\mathcal{A}_0$. Dizemos que $\mathcal{P}$ é uma família Jeffrey Conjugada ou J-Conjugada com respeito à $\mathcal{A}_0$ se $P^* \in \mathcal{P}$.

Especialmente no contexto preditivista o seguinte resultado pode ser estabelecido.

Proposição 2.4.1 *Seja $\mathcal{P}$ uma família de medidas de probabilidade definidas em $(\mathcal{R}^\infty, \mathcal{B}^\infty)$, permutáveis. Então, $\mathcal{P}$ é J-Conjugada com respeito à sub-σ-álgebra de eventos permutáveis $\mathcal{A}_0 \subseteq \mathcal{B}^\infty$.*

Prova: Seja $P \in \mathcal{P}$. Então pelo Teorema de de Finetti(1937), para todo $n \geq 1$

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \Pi_{i=1}^n F(x_i) dQ(F), \tag{2.63}$$

onde Q é a P-lei da quantidade aleatória $F = \lim_{n\to\infty} F_n$, F_n é como na expressão (2.36) e $\mathcal{F}$ é o conjunto de todas as distribuições de probabilidade definidas sobre $\mathcal{R}$.

Dessa forma,

$$\begin{aligned} \Pi_{i=1}^n F(x_i) &= P(X_1 \leq x_1, \ldots, X_n \leq x_n | F) \\ &= P(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(\mathbf{x}), \quad \forall P \in \mathcal{F} \quad \text{e} \quad \forall \mathbf{x} \in \mathcal{R}^\infty, \end{aligned} \tag{2.64}$$

onde $\mathcal{A}_0$ é a σ–álgebra de eventos permutáveis gerada por F. Note que $\mathcal{A}_0$ é suficiente para $\mathcal{P}$.

Seja P^* uma medida de probabilidade definida em $(\mathcal{R}^\infty, \mathcal{B}^\infty)$ obtida de P via Regra de Jeffrey com respeito a $\mathcal{A}_0$ e tal que $P^* << P$. Então, pela definição 2.3.1, para todo $n \geq 1$

$$P^*(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{R}^\infty} P(X_1 \leq x_1, \ldots, X_n \leq x_n | \mathcal{A}_0)(\mathbf{x}) dP^*(\mathbf{x}). \tag{2.65}$$

Como $P \in \mathcal{P}$, pela expressão (2.64) temos que

$$P^*(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \Pi_{i=1}^n F(x_i) dQ^*(F). \tag{2.66}$$

onde $Q^*(F)$ é uma distribuição marginal obtida de $P^*(\mathbf{x})$.

Aplicando novamente o teorema de de Finetti(1937), temos que P^* é permutável e, portanto, pertence a $\mathcal{P}$. Logo, $\mathcal{P}$ é J-conjugada com respeito a $\mathcal{A}_0$.

□

De fato, $\mathcal{P}$ é J-conjugada com respeito a $\mathcal{A}_0 \subseteq \mathcal{A}$ sempre que $\mathcal{P}$ é a maior família que faz de $\mathcal{A}_0$ suficiente. Seja $\mathcal{A}_0$ uma sub-σ-álgebra suficiente para $\mathcal{P}$, $\mathcal{P}$ dominada, e denote por E_Q a esperança calculada com respeito à medida de probabilidade Q. Então pela definição 3.1 de Pereira *at alii* para toda função f limitada e $\mathcal{A}$ mensurável existe uma função limitada f_1, $\mathcal{A}_0$- mensurável, tal que

$$f_1(w) = \mathrm{E}_Q(f|\mathcal{A}_0)(w), \quad \forall Q \in \mathcal{P}.$$

Dessa forma, para $f = 1_A$, $A \in \mathcal{A}$

$$\begin{aligned} f_1(w) &= \mathrm{E}_Q(1_A|\mathcal{A}_0)(w) \\ &= Q(A|\mathcal{A}_0)(w), \quad \forall Q \in \mathcal{P}. \end{aligned} \tag{2.67}$$

Seja P^* uma medida de probabilidade definida em $(\Omega, \mathcal{A})$ a qual é dominada por $P \in \mathcal{P}$ e é obtida de P através da Regra de Jeffrey com respeito a $\mathcal{A}_0$. Então para todo $A \in \mathcal{A}$

$$P^*(A|\mathcal{A}_0)(w) = P(A|\mathcal{A}_0)(w). \tag{2.68}$$

Decorre desse fato e da expressão (2.67) que

$$\begin{aligned} E_{P^*}(1_A|\mathcal{A}_0)(w) &= P^*(A|\mathcal{A}_0)(w) \\ &= P(A|\mathcal{A}_0)(w) = f_1(w). \end{aligned} \tag{2.69}$$

Similarmente pode-se provar o resultado anterior para qualquer função limitada f, $\mathcal{A}$-mensurável, levando-nos a concluir que $P^* \in \mathcal{P}$.

Se $\mathcal{A}_0$, na proposição 2.4.1, é a σ-álgebra das partes de Ω a família de medidas permutáveis não é necessariamente J-conjugada. Neste caso, teremos uma reavaliação completa das probabilidades inicialmente atribuídas a qual pode ser feita através de qualquer critério que Você julgar adequado.

A proposição 2.4.1 estabelece que qualquer medida de probabilidade obtida de uma medida permutável através da Regra de Jeffrey com respeito à σ-álgebra de eventos permutáveis é também permutável, sem imposição de restrições adicionais à medida misturadora Q. Ou seja, para termos J-conjugação basta que a informação imprevista atue sobre sua opinião mantendo a condição de permutabilidade inicialmente considerada.

Note ainda que, na proposição anterior, além da exigência de retaxações sobre um espaço infinito dimensional, a função de verossimilhança envolvida na mistura nem sempre é conhecida. Com isto a determinação da distribuição *a posteriori* é árdua pois, além de exigir a declaração de Q, requer dados com estrutura suficientemente simples para permitir a identificação da função de verossimilhança - o que ocorre em condições muito particulares.

No exemplo a seguir, apresentaremos uma situação onde a J-conjugação é verificada utilizando-nos de variáveis aleatórias permutáveis cuja estrutura - muito simples - permite identificar a função de verossimilhança da mistura. Além disso será considerado um espaço paramétrico finito-dimensional o que pode tornar mais simples a declaração de Q.

Exemplo 2.11: Seja $\mathcal{P}$ a família de distribuições de probabilidades permutáveis definidas sobre $(\{0,1\}^\infty, \mathcal{B}^\infty)$. Então, para toda $P \in \mathcal{P}$ e para $n \geq 1$ temos, pelo teorema de de

Finetti(1937), que

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_0^1 \Pi_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta) \tag{2.70}$$

onde $\Pi_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = P(X_1 = x_1, \ldots, X_n = x_n|\theta) = P(X_1 = x_1, \ldots, X_n = x_n|\mathcal{A}_0)(\mathbf{x})$ e $\mathcal{A}_0$ e $\mathbf{x}$ são como o definido para a situação descrita na expressão (2.40).

Seja P^* uma medida de probabilidade definida sobre $(\{0,1\}^\infty, \mathcal{B}^\infty)$ obtida de P através da Regra de Jeffrey com respeito a $\mathcal{A}_0$ e tal que $P^* << P$. Então para todo $n \geq 1$

$$\begin{aligned} P^*(X_1 = x_1, \ldots, X_n = x_n) &= \int_{\mathcal{F}} P(X_1 = x_1, \ldots, X_n = x_n|\mathcal{A}_0)(\mathbf{x}) dP^*(\mathbf{x}) \\ &= \int_0^1 \Pi_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} dQ^*(\theta), \end{aligned} \tag{2.71}$$

implicando, novamente pelo teorema de de Finetti(1937), que P^* é permutável e, portanto, que $\mathcal{P}$ é J-Conjugada com respeito à $\mathcal{A}_0$. Ou seja, mantida a condição de permutabilidade a distribuição *a posteriori* obtida via Regra de Jeffrey é dada pela expressão (2.71).

Perceba que no exemplo anterior, a distribuição *a posteriori* ficará completamente especificada quando a medida Q^*, definida sobre os elementos de Θ e atualizada arbitrariamente, for declarada.

Considerada como o apresentado na definição 2.4.2 a J-conjugação não se revela muito interessante pois permite famílias muito amplas de distribuições de probabilidades. Assim como na conjugação Bayesiana, a família $\mathcal{P}$ de todas as possíveis distribuições de probabilidades definidas sobre $(\Omega, \mathcal{A})$, onde $\mathcal{A}$ é formada pelas partes de Ω, é J-conjugada com respeito a própria σ - álgebra $\mathcal{A}$, que é suficiente para $\mathcal{P}$. Considere P^* como sendo uma distribuição de probabilidade qualquer definida em $(\Omega, \mathcal{A})$, dominada por $P \in \mathcal{P}$ e obtida de P através da Regra de Jeffrey com respeito a $\mathcal{A}$. Então, para todo $A \in \mathcal{A}$ temos

$$\begin{aligned} P^*(A) &= \int_\Omega P(A|\mathcal{A})(w) dP^*(w) \\ &= \int_\Omega E(1_A|\mathcal{A})(w) dP^*(w) \\ &= \int_{(\Omega,A)} dP^*(w) = P^*(A), \end{aligned}$$

o que consiste em uma reavaliação arbitrária completa da distribuição de probabilidade inicialmente atribuída a elementos de $\mathcal{A}$, revelando que P^* é um elemento de $\mathcal{P}$. Este fato depõe contra o valor prático da definição de J-conjugação estabelecida, uma vez que a tarefa de encontrar P^* poderá ser tão psicologicamente complexa quanto a tarefa de construir a distribuição *a priori*.

Outra situação pouco interessante é aquela que requer retaxações sobre um espaço infinito-dimensional. Por exemplo, quando a família $\mathcal{P}$ formada pelas medidas de probabilidades permutáveis é representada como a mistura dada na expressão (2.35), a J-conjugação é obtida após serem feitas retaxações sobre um espaço infinito-dimensional e demasiadamente complexo que é o espaço de todas as funções de distribuição definidas em $\mathcal{R}$- o que a princípio não aparenta ser uma tarefa muito simples.

Buscamos uma definição de conjugação segundo a Regra de Jeffrey que nos leve a uma identificação rápida da distribuição *a posteriori* nos obrigando o mínimo possível de retaxações arbitrárias de probabilidades. Consegue-se uma diminuição no número destas retaxações, sempre que o problema considerado puder ser representado como uma mistura regida por uma distribuição *a priori* definida sobre um espaço paramétrico de dimensão finita. Esta situação pode ser apreciada no exemplo 2.11 anteriormente visto, onde o espaço paramétrico era uni-dimensional e θ era interpretado como $\lim_{n\to\infty} \sum_{i=1}^{n} X_i/n$ (para muitos a probabilidade de sucesso num experimento binomial).

Com o intuito descrito no parágrafo anterior, apresentamos o seguinte conceito.

Definição 2.4.3 *(Família J-Conjugada Natural)*

Sejam $\mathcal{P}$ uma família de distribuições de probabilidade definidas sobre um espaço mensurável $(\Omega, \mathcal{A})$ e $\mathcal{A}_0$ uma sub-σ-álgebra de $\mathcal{A}$ gerada por um objeto aleatório $\theta = \theta(w)$, $w \in \Omega$ que assume valores em um espaço paramétrico Θ, finito dimensional. Dizemos que $\mathcal{P}$ é uma família J-Conjugada Natural com respeito a $\mathcal{A}_0$ se

(i) *a família $\mathcal{P}$ é J-conjugada com respeito a $\mathcal{A}_0$, isto é,*

$$P^*(\cdot) = \int_\Theta P(\cdot|\mathcal{A}_0)(w)dQ^*(\theta) \ \in \ \mathcal{P}, \tag{2.72}$$

onde $Q^* \in \mathcal{Q}$ é a restrição de P^* aos borelianos de Θ e $\mathcal{Q}$ é a família de medidas de probabilidade definidas sobre Θ; e se

(ii) *a família de medidas de probabilidades $\mathcal{Q}$ é conjugada natural no sentido Bayesiano usual com respeito a $P(.|\mathcal{A}_0)$.*

Como no caso Bayesiano usual, esta definição restringe o número de elementos de $\mathcal{P}$ sem, no entanto, tornar imperativa a escolha da distribuição *a priori*. Ao limitarmos a possibilidade de J-conjugação natural à existência de uma sub-σ-álgebra gerada por um objeto aleatório finito dimensional, provocamos uma "simplificação" na determinação das probabilidades retaxadas arbitrariamente e a exigência de proporcionalidade (condição (ii)) entre estas novas medidas e a função de verossimilhança induzida pela sub-σ-álgebra suficiente nos leva a uma rápida identificação da distribuição *a posteriori*.

Lembramos, mais uma vez, que apesar das facilidades trazidas por escolhas de distribuições *a priori* J-conjugadas, seu uso é justificável somente quando estas de fato reflitam sua opinião (coerente no instante em que é eliciada) sobre o objeto desconhecido.

Considere $\mathcal{F}$ como o considerado no corolário 2.3.1 e denote por $\mathcal{Q}$ a família das medidas de probabilidade Q_ψ consideradas neste corolário.

Proposição 2.4.2 *Seja $\mathcal{P}$ a família de medidas de probabilidades definidas sobre $(\mathcal{R}^\infty, \mathcal{B}^\infty)$, permutáveis e admita a representação dada na expressão (2.42). $\mathcal{P}$ é J-conjugada natural com respeito a sub-σ-álgebra de eventos permutáveis $\mathcal{A}_0 = \sigma(\theta)$ se, e somente se, a família de medidas de probabilidade $\mathcal{Q}$ é conjugada natural com respeito a $\mathcal{F}$.*

Prova: Suponha que $\mathcal{P}$ é J-conjugada natural. Então, a definição 2.4.3 garante a conjugação natural de $\mathcal{Q}$ com respeito a $\mathcal{F}$. Para a volta basta usar a proposição 2.4.1 e a hipótese.

□

Exemplo 2.12: Do exemplo 2.11 temos que a família de medidas de probabilidade $\mathcal{P}$ definidas sobre $(\{0,1\}^\infty, \mathcal{B}^\infty)$ é J-conjugada com respeito a σ-álgebra $\mathcal{A}_0 = \sigma\left(\lim_{n\to\infty} \sum_{i=1}^{n} X_i/n\right)$ e que

$$P^*(X_1 = x_1, \ldots, X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} dQ^*(\theta). \tag{2.73}$$

Considere $\mathcal{Q}$ como a família de distribuições Beta a qual é conjugada natural no sentido usual da família de função de verossimilhança $P(X_1 = x_1, \ldots, X_n = x_n | \mathcal{A}_0) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$. Tomando $Q^* \in \mathcal{Q}$, temos que $\mathcal{P}$ é J-conjugada natural com respeito a $\mathcal{A}_0$ e além disto para todo $P' \in \mathcal{P}$ e para todo $n \geq 1$ vale a seguinte representação

$$P'(X_1 = x_1, \ldots, X_n = x_n) = \frac{\text{Beta}(\alpha' + \sum_{i=1}^n x_i, n + \beta' - \sum_{i=1}^n x_i)}{\text{Beta}(\alpha', \beta')}, \tag{2.74}$$

onde $\text{Beta}(a, b)$ é a função Beta com parâmetros a e b.

Observe dos resultados acima, que se ao receber uma informação inesperada Você mantém inalterada a condição de invariância inicialmente percebida na seqüência e altera sua opinião inicial sobre elementos de Θ apenas reavaliando os índices da distribuição Q (sem alterar a forma funcional) sua distribuição *a posteriori* será obtida via Regra Preditivista de Jeffrey com respeito a σ-álgebra suficiente adequada e além disto terá sua forma funcional facilmente identificável.

Na próxima seção apresentaremos resultados referentes à J-Conjugação Natural para famílias de distribuições que podem ser representadas como misturas de distribuições pertencentes à família exponencial. Devido à estrutura especial dos elementos desta família, alguns resultados interessantes podem ser estabelecidos, como veremos a seguir.

2.5 J-Conjugação Natural para a Família Exponencial

Como vimos na seção 2.3, Diaconis e Freedman(1990) estabelecem um teorema de representação caracterizando seqüências de variáveis aleatórias contínuas como uma mistura

de elementos cuja distribuição (comum) pertence à família exponencial uni-paramétrica.

Esse tipo de representação adequa-se a várias situações de interesse - como por exemplo à situação abordada no capítulo 5 - e por isto pretendemos estabelecer condições para a J-conjugação natural para a família de medidas provenientes dessa mistura.

Nesta seção, seguindo as considerações de Bernardo e Smith(1994), consideraremos como membro da família exponencial d-paramétrica denotada por $\mathcal{E}_{f_d}$, $d \geq 1$, qualquer função densidade que satisfaça a seguinte definição.

Definição 2.5.1 *(Família Exponencial)*
Uma função de densidade ou função de probabilidade $p(x|\theta)$, $x \in \mathcal{X} \subseteq \mathcal{R}$, rotulada por $\theta \in \Theta \subseteq \mathcal{R}^d$, pertence à Família Exponencial d-paramétrica $\mathcal{E}_{f_d}$ se é da forma

$$p(x|\theta) = h(x)L(\theta)\exp\left\{\sum_{i=1}^{d} c_i\phi_i(\theta)f_i(x)\right\} \tag{2.75}$$

onde as funções h (positiva), $f = (f_1, \ldots, f_d)$ e $\phi = (\phi_1, \ldots, \phi_d)$ e as constantes c_i's são dadas e

$$[L(\theta)]^{-1} = \int_{\mathcal{X}} h(x)\exp\left\{\sum_{i=1}^{d} c_i\phi_i(\theta)f_i(x)\right\} d(x) < \infty. \tag{2.76}$$

Neste caso, $f = (f_1, \ldots, f_d)$ é uma estatística suficiente para a família $\mathcal{E}_{f_d}$ e tem dimensão fixa.

Considerando essa definição, Bernardo e Smith(1994) estabelecem o seguinte resultado o qual fornece a família de distribuições conjugada natural para $\mathcal{E}_{f_d}$.

Teorema 2.5.1 *(Família Conjugada Natural para $\mathcal{E}_{f_d}$)*

Se $X_1, \ldots, X_n$ são variáveis aleatórias condicionalmente independentes e identicamente distribuídas dado $\theta \in \Theta \subseteq \mathcal{R}^d$, com distribuição pertencente à $\mathcal{E}_{f_d}$, então a família conjugada natural para θ assume a seguinte forma:

$$p(\theta|\tau) = [k(\tau)]^{-1}[L(\theta)]^{\tau_0}\exp\{\sum_{i=1}^{d} c_i\phi_i(\theta)\tau_i\}, \tag{2.77}$$

onde $k(\tau) = \int_\Theta [L(\theta)]^{\tau_0} \exp\{\sum_{i=1}^d c_i\phi_i(\theta)\tau_i\}d\theta < \infty$ *e* $\tau = (\tau_0, \tau_1, \ldots, \tau_d)$ *são objetos que tornam* $p(\theta|\tau)$ *uma função de densidade ou função de probabilidade.*

Prova : Como $X_1, \ldots, X_n$ são variáveis aleatórias condicionalmente independentes e identicamente distribuídas dado θ, com distribuição dada na expressão (2.75) e considerando a estatística suficiente $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_d(\mathbf{x}))$, onde $\mathbf{x} = (x_1, \ldots, x_n)$, a função de verossimilhança adequada ao problema é

$$l(f(\mathbf{x}), \theta) \ \alpha \ [L(\theta)]^n \exp\left\{\sum_{i=1}^d c_i\phi_i(\theta) \sum_{j=1}^n f_i(x_j)\right\} \tag{2.78}$$

a qual é proporcional a distribuição *a priori* $p(\theta|\tau)$ dada na expressão (2.77), quando vista em função de θ.

Além disso, a distribuição *a posteriori* para θ é

$$\begin{aligned} p(\theta|\mathbf{x}, \tau) \ &\alpha \ l(f(\mathbf{x}), \theta)p(\theta|\tau) \\ &\alpha \ [L(\theta)]^{n+\tau_0} \exp\left\{\sum_{i=1}^d c_i\phi_i(\theta)\left[\tau_i + \sum_{j=1}^n f_i(x_j)\right]\right\} \end{aligned} \tag{2.79}$$

que pertence a mesma família de distribuições de probabilidade a que pertence a distribuição *a priori*. Portanto, a família de probabilidade constituída por elementos da forma apresentada em (2.77) é conjugada natural para a família exponencial $\mathcal{E}_{f_d}$.

□

Perceba que para esta situação, realizar a operação Bayesiana usual consiste em re-rotular os índices (parâmetros, se desejar) da distribuição *a priori* incorporando neles a informação trazida pela amostra. Se o índice da distribuição *a priori* é $\tau = (\tau_0, \tau_1, \ldots, \tau_d)$, o índice da distribuição *a posteriori* é facilmente obtido e assume a forma

$$\tau + f(\mathbf{x}) = \left(\tau_0 + n, \tau_1 + \sum_{j=1}^n f_1(x_j), \ldots, \tau_d + \sum_{j=1}^n f_d(x_j)\right). \tag{2.80}$$

Também nos é possível identificar com certa facilidade a distribuição preditiva *a posteriori* para observações futuras $Y_1, \ldots, Y_m$, como pode ser visto no seguinte teorema estabelecido por Bernardo e Smith(1994).

Teorema 2.5.2 *(Bernardo e Smith(1994))*

Se $X_1, \ldots, X_n$ são variáveis aleatórias condicionalmente independentes e identicamente distribuídas (ciid) dado $\theta \in \Theta \subseteq \mathcal{R}^d$, com distribuição pertencente a $\mathcal{E}_{f_d}$ e $p(\theta|\tau)$ é a distribuição a priori para θ conjugada natural com respeito a função de verossimilhança $l(f(\mathbf{x}); \theta)$, então a distribuição preditiva a posteriori para as observações futuras $Y_1, \ldots, Y_m$, que são ciid dado θ com distribuição pertencente a $\mathcal{E}_{f_d}$, é :

$$p(y_1, \ldots, y_m | x_1, \ldots, x_n, \tau) = \Pi_{l=1}^m h(y_l) \frac{k(\tau + t_n(\mathbf{x}) + t_m(\mathbf{y}))}{k(\tau + t_n(\mathbf{x}))}, \tag{2.81}$$

onde $t_m(\mathbf{y}) = (m, \sum_{l=1}^m f_1(y_l), \ldots, \sum_{l=1}^m f_d(y_l))$, $t_n(\mathbf{x}) = (n, \sum_{j=1}^n f_1(x_j), \ldots, \sum_{j=1}^n f_d(x_j))$ e $k(.)$ é como o definido para o teorema 2.5.1.

Prova: Considerando as independências condicionais em θ e usando cálculo de probabilidades, obtemos

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}, \tau) &= \int_\Theta p(\mathbf{y}, \theta | \mathbf{x}, \tau) d\theta \\ &= \int_\Theta p(\mathbf{y}|\theta, \tau) p(\theta|\mathbf{x}, \tau) d\theta \end{aligned}$$

Sob a hipótese de conjugação natural, determina-se:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}, \tau) &= \Pi_{l=1}^m \frac{h(y_l)}{k(\tau + t_n)} \int_\Theta [L(\theta)]^{n+m+\tau_0} \exp\left\{ \sum_{i=1}^d c_i \phi_i(\theta) \{\tau_i + \sum_{j=1}^n f_i(x_j) + \sum_{l=1}^m f_i(y_l)\} \right\} d\theta \\ &= \Pi_{l=1}^m h(y_l) \frac{k(\tau + t_n + t_m)}{k(\tau + t_n)}, \end{aligned}$$

concluíndo a prova do teorema.

□

A preditiva *a posteriori* também é obtida de uma re-rotulação dos valores que indexam a preditiva *a priori*, uma vez que esta é dada pela seguinte expressão:

$$p(\mathbf{x}|\tau) = \Pi_{j=1}^{n} h(x_j) \frac{k(\tau + t_n)}{k(\tau)}. \tag{2.82}$$

Como na seção 2.3, considere a família $\mathcal{M}_Q$ de medidas de probabilidades contínuas definidas sobre $(\mathcal{R}^d)^\infty$ tal que, para todo $n \geq 1$, $P(S_n \in C_n) = 1$ e Q_{nsn} representa a P - lei de $X_1, \ldots, X_n | S_n = s$, onde C_n e Q_{nsn} são como definidas para o teorema 2.3.2.

Proposição 2.5.1 *Suponha válidas as hipóteses do teorema 2.3.2. A família $\mathcal{M}_Q$ é J-conjugada natural com respeito à sub-σ-álgebra $\mathcal{A}_0$ gerada pela quantidade aleatória $\lim_{n\to\infty} S_n/n$ se, e somente se, a família $\mathcal{Q}$ das medidas misturadoras do teorema 2.3.2 é conjugada natural com respeito à função de verossimilhança induzida por $\mathcal{A}_0$, P_θ^∞.*

Prova: Omitida por ser similar a prova da proposição 2.4.1.

Suponha que sua opinião inicial sobre uma seqüência infinita de variáveis aleatórias $X_1, X_2, \ldots$ seja $P \in \mathcal{M}_Q$, a qual pode ser obtida através da mistura realizada por uma medida de probabilidade Q pertencente à família de distribuições indexadas por τ, conjugada natural com respeito a P_θ^∞ onde $\theta \in \Theta \subseteq \mathcal{R}$. Então, como conseqüência do teorema 2.5.2 temos que para qualquer sub-seqüência $X_{(n)}$, $n \geq 1$,

$$\begin{aligned} p(x_1, \ldots, x_n) &= \int_\Theta \Pi_{i=1}^{n} h(x_i) [D(\theta)]^{-n} \exp\left\{\sum_{i=1}^{n} x_i.\theta\right\} dQ(\theta) \\ &= \Pi_{i=1}^{n} h(x_i) \frac{k(\tau + t_n)}{k(\tau)}, \end{aligned}$$

onde $k(\tau) = \int_\Theta [D(\theta)]^{-\tau_0} \exp\{\tau_1.\theta\}\, d\theta < \infty$, $\tau = (\tau_0, \tau_1)$ e $t_n = (n, \sum_{i=1}^{n} x_i)$.

Admita que Você tenha recebido uma informação cuja possibilidade de ocorrência não havia sido previamente considerada por Você e que o levou a modificar sua opinião sobre $X_1, X_2, \ldots$ alterando apenas sua opinião sobre os elementos de Θ. Esta nova informação provoca a reavaliação de sua opinião *a priori* Q dando origem a sua nova opinião Q^* que indexada por τ^* e também pertencente a $\mathcal{Q}$. Então, pela proposição anterior sua distribuição *a posteriori* sobre $X_{(n)}$ é dada pela Regra de Jeffrey, através da seguinte expressão:

$$p^*(x_1, \ldots, x_n) = \Pi_{i=1}^{n} h(x_i) \frac{k(\tau^* + t_n)}{k(\tau^*)}. \tag{2.83}$$

Ou seja, se suas opiniões *a priori* e *a posteriori* são J-conjugadas naturais com respeito a sub-σ-álgebra gerada por $\lim_{n\to\infty} S_n/n$ (S_n é a estatística suficiente para a família exponencial envolvida na mistura), a distribuição *a posteriori* P^* é obtida imediatamente da distribuição *a priori* substituindo os índices iniciais pelos índices retaxados τ^*, o que simplifica bastante o problema de determinação da distribuição *a posteriori* e diminui sensivelmente o número de retaxações feitas arbitrariamente - basta Você declarar os parâmetros da distribuição *a posteriori* Q^*, pois sua forma funcional já será conhecida.

Perceba ainda das expressões (2.81) e (2.83) que neste caso a distribuição obtida através de uma atualização feita via Regra de Jeffrey por mera substituição de τ pelo valor reavaliado τ^* é igual àquela obtida utilizando-se o Condicionamento Bayesiano após observar-se os valores de n variáveis aleatórias $X_1, \ldots, X_n$ cuja função de distribuição condicional em θ tem função densidade na família exponencial $\mathcal{E}_{f_d}$ e estatística suficiente $f = (f_1, \ldots, f_d)$, desde que

$$\tau^* - \tau = t_n(\mathbf{x}) = \left(n, \sum_{i=1}^{n} f_1(x_i), \ldots, \sum_{i=1}^{n} f_d(x_i)\right).$$

Ou seja, se $\tau^* - \tau = t_n(\mathbf{x})$ a seguinte igualdade pode ser verificada

$$\begin{aligned} p^*(y_1, \ldots, y_m|\tau^*) &= \Pi_{l=1}^{m} h(y_l) \frac{k(\tau^* + t_m(\mathbf{y}))}{k(\tau^*)} \\ &= p(y_1, \ldots, y_m|x_1, \ldots, x_n, \tau), \end{aligned} \tag{2.84}$$

onde $\mathbf{y} = (y_1, \ldots y_m)$ são m observações futuras, identicamente distribuídas às variáveis X_i's, isto é, têm distribuição na família exponencial $\mathcal{E}_{f_d}$

Voltemos ao exemplo 2.9 da seção 2.3 para ilustrar a J-conjugação natural para a família exponencial.

Exemplo 2.13: Da seção 2.3 temos que toda medida de probabilidade P definida em $(\mathcal{R}_+^\infty, \mathcal{B}^\infty)$ que satisfaz a condição

$$P(X_{(n)} \in A) = P(X_{(n)} \in A + \mathbf{x}), \quad \forall\, n \geq 1 \tag{2.85}$$

onde A é um boreliano qualquer de $\mathcal{R}_+^n$ e $\mathbf{x} \in \mathcal{R}^n$ é tal que $\sum_{i=1}^{n} x_i = 0$ e $A + \mathbf{x} \in \mathcal{R}_+^n$, pode ser representada como a seguinte mistura de distribuições exponenciais:

$$p(x_1, \ldots, x_n) = \int_0^\infty \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\} dQ(\theta), \tag{2.86}$$

onde p é a função densidade associada a P. Diremos, neste caso, que $P \in \mathcal{M}_Q$, onde $D(\theta) = \frac{1}{\theta}$ e $h_i = 1$ para todo $i \geq 1$.

Se inicialmente sua opinião Q pertence à família de distribuições Gama com parâmetros $\alpha > 0$ e $\beta > 0$ - $\mathcal{G}(\alpha, \beta)$ - a qual é conjugada natural da família de distribuições exponenciais, então para todo $n \geq 1$, a distribuição preditiva torna-se

$$p(x_1, \ldots, x_n) = \frac{k\left((\alpha - 1, \beta) + (n, \sum_{i=1}^n x_i)\right)}{k\left(\alpha - 1, \beta\right)} \tag{2.87}$$

onde

$$k(\tau_0, \tau_1) = \int_0^\infty \theta^{\tau_0} \exp\{\tau_1 . \theta\} = \frac{\Gamma(\tau_0 + 1)}{\tau_1^{\tau_0 + 1}}. \tag{2.88}$$

Ou seja, teremos

$$p(x_1, \ldots, x_n) = \frac{\beta^\alpha}{(\beta + \sum_{i=1}^n x_i)^{n+\alpha}} \frac{\Gamma(n + \alpha)}{\Gamma(\alpha)}. \tag{2.89}$$

Sendo assim, se Você recebe uma informação que faz com que Você mude arbitrariamente de sua opinião inicial Q para uma nova opinião Q^* que é a distribuição $\mathcal{G}(\alpha^*, \beta^*)$, com $\alpha^* > 0$ e $\beta^* > 0$ onde pelo menos um destes novos parâmetros é diferente dos respectivos parâmetros iniciais, e tal que Você mantenha *a posteriori* a condição (2.85) para P^*, a proposição 2.5.1 garante que suas opiniões *a priori* P e *a posteriori* P^* são J-conjugadas naturais com respeito a $\mathcal{A}_0 = \sigma(\sum_{i=1}^n X_i/n)$ e sua opinião *a posteriori* é dada pela seguinte expressão

$$p^*(x_1, \ldots, x_n) = \frac{\beta^{*\alpha^*}}{(\beta^* + \sum_{i=1}^n x_i)^{n+\alpha^*}} \frac{\Gamma(n + \alpha^*)}{\Gamma(\alpha^*)}, \tag{2.90}$$

onde p^* é a densidade induzida por P^*.

Note que a distribuição obtida utilizando a Regra de Jeffrey dada na expressão (2.90) do exemplo acima é a mesma distribuição fornecida pelo Condicionamento Bayesiano para $X_{(n)}$

condicional em $\alpha^* - \alpha$ observações Y_i's satisfazendo a condição (2.85), desde que $\alpha^* - \alpha$ seja um inteiro positivo m e $\beta^* - \beta = \sum_{i=1}^{m} y_i$.

Num contexto um pouco mais preditivista e quando as quantidades aleatórias de interesse são contínuas, a Jeffrey conjugação natural para misturas de elementos da família exponencial poderá ser conseguida através de suposições sobre objetos observáveis apenas. (Como veremos mais adiante na proposição 2.5.2.)

Para demonstrar este fato, além do teorema de Diaconis e Freedman(1990) visto na seção 2.3.1, vamos fazer uso do seguinte teorema estabelecido por Diaconis e Ylvisaker(1979), o qual fornece condições para a caracterização de distribuições *a priori* conjugadas naturais para a família exponencial paramétrica que descreve o comportamento de dados contínuos.

Seja ξ uma medida $\sigma-$ finita sobre os borelianos de $\mathcal{R}^d$, fixa e seja $\mathcal{X}$ é o interior da casca convexa do conjunto suporte de ξ. Suponha que $\mathcal{X}$ é um conjunto aberto e não-vazio de $\mathcal{R}^d$.

Defina $L(\theta) = \log \int_{\mathcal{X}} \exp\{x \cdot \theta\}\, d\xi(x)$ onde $x \cdot y$ denota o produto interno usual em$\mathcal{R}^d$ de x por y e seja $\Theta = \{\theta \in \mathcal{R}^d : L(\theta) < \infty\}$ um conjunto convexo, aberto e não-vazio que representa o espaço paramétrico natural da família de densidades com respeito a ξ, dadas por:

$$dP_\theta(x) = \exp\{x \cdot \theta - L(\theta)\}\, d\xi(x), \quad \theta \in \Theta \subseteq \mathcal{R}^d. \tag{2.91}$$

Neste caso, diferenciando $L(\theta)$ com respeito a θ verificamos que

$$E(X|\theta) = L'(\theta). \tag{2.92}$$

Sob essas considerações, Diaconis e Ylvisaker (1979) estabelecem o seguinte teorema.

Teorema 2.5.3 *(Diaconis e Ylvisaker (1979))*
Suponha que Θ é um aberto de $\mathcal{R}^d$. Seja X uma amostra de tamanho um de P_θ. Suponha que o suporte de ξ contém um intervalo aberto em $\mathcal{R}^d$. Se θ tem um distribuição a priori própria π, a qual não está concentrada em um único ponto, e se

$$E\{L'(\theta)|X\} = aX + b \tag{2.93}$$

para alguma constante $a \neq 0$ e algum vetor constante b então π é absolutamente contínua com respeito a medida de Lebesgue e tem densidade

$$d\pi(\theta) = c\exp\left\{\frac{b}{a}\cdot\theta - \frac{(1-a)}{a}L(\theta)\right\}d\theta \tag{2.94}$$

Perceba que Diaconis e Ylvisaker(1979) determina a distribuição *a priori* conjugada natural para θ partindo basicamente de uma condição de linearidade sobre $E(L'(\theta)|X)$. Nosso primeiro passo, então, é encontrar condições sobre objetos observáveis que sejam equivalentes à condição (2.93).

Suponhamos que h seja uma função de Borel, integrável, não-negativa e finita definida em $\mathcal{R}^d$. Seja

$$D(\theta) = \int_{\mathcal{R}^d} h(x)\exp\{x\cdot\theta\}\,d(x) \tag{2.95}$$

onde $\theta \in \Theta = \{\theta \in \mathcal{R}^d : L(\theta) < \infty\}$ e Θ é um conjunto não-vazio, aberto e convexo (note que este conjunto também faz de $D(\theta)$ finito uma vez que $L(\theta) = \log D(\theta)$). Considere $\mathcal{M}_Q$ como a família de medidas de probabilidades definidas em $(\mathcal{R}^d)^\infty$ satisfazendo as exigências do teorema de Diaconis e Freedman (1990).

Proposição 2.5.2 *Seja $X_1, X_2, \ldots$ uma seqüência infinita de quantidades aleatórias definidas em $\mathcal{R}^d$ com lei $P \in \mathcal{M}_Q$. Então, se $E(X_2|X_1) = aX_1 + b$, onde a é uma constante não-nula e b é um vetor de constantes d-dimensional qualquer, temos que para todo $n \geq 1$*

$$p(x_1,\ldots,x_n) = \Pi_{i=1}^n h(x_i)\frac{k\left(n+\frac{1-a}{a};\frac{b}{a}+\sum_{i=1}^n x_i\right)}{k\left(\frac{1-a}{a};\frac{b}{a}\right)}, \tag{2.96}$$

onde $k(\tau_0,\tau_1) = \int_\Theta D(\theta)^{-\tau_0}\exp\{\theta\cdot\tau_1\}d\theta$ e p é a densidade associada a P.

Prova: Como por hipótese $P \in \mathcal{M}_Q$, pelo teorema de Diaconis e Freedman(1990), temos que para todo $n \geq 1$

$$P(X_1 \leq x_1,\ldots,X_n \leq x_n) = \int_\Theta P_\theta^n(X_1 \leq x_1,\ldots,X_n \leq x_n)dQ(\theta), \tag{2.97}$$

onde P_θ^n é a P-lei que faz de $X_1, \ldots, X_n$ condicionalmente independentes e identicamente distribuídas dado θ com distribuição

$$dP_\theta(x) = \frac{h(x)\exp\{x \cdot \theta\}}{D(\theta)} dx. \tag{2.98}$$

Então, considerando p como a densidade induzida por P, o corolário 2.3.2 nos permite a seguinte representação:

$$p(x_1, \ldots, x_n) = \int_\Theta \Pi_{i=1}^n h(x_i)[D(\theta)]^{-n} \exp\left\{\sum_{i=1}^n x_i \cdot \theta\right\} dQ(\theta), \quad \forall n \geq 1. \tag{2.99}$$

Seja X_1 uma amostra de tamanho um da distribuição dada em (2.91). Considerando $d\xi(x) = h(x)dx$ e $L(\theta) = \log D(\theta)$, temos:

$$E(X_1|\theta) = \frac{d}{d\theta} L(\theta) = L'(\theta) \tag{2.100}$$

pois $X_1|\theta$ tem distribuição na família exponencial. Usando propriedades de esperança condicional, a hipótese e o fato que condicional em θ, X_1 e X_2 são independentes e identicamente distribuídas,

$$\begin{aligned} E\{L'(\theta)|X_1\} &= E\{E\{X_1|\theta\}|X_1\} \\ &= E\{E\{X_2|\theta\}|X_1\} \\ &= E\{E\{X_2|\theta, X_1\}|X_1\} \\ &= E\{X_2|X_1\} = aX_1 + b. \end{aligned}$$

Então se a distribuição *a priori* para θ é própria e não concentra em um único ponto, pelo teorema de Diaconis e Ylvisaker(1979), Q é absolutamente contínua com respeito a medida de Lebesgue, é conjugada natural com respeito a P_θ^∞ e tem densidade dada por

$$dQ(\theta) = c\exp\left\{\frac{b}{a} \cdot \theta - \frac{1-a}{a} \log D(\theta)\right\} d\theta, \tag{2.101}$$

onde a constante $c = \frac{1}{k\left(\frac{1-a}{a}; \frac{b}{a}\right)}$ e $k(\tau_0, \tau_1) = \int_\Theta D(\theta)^{-\tau_0} \exp\{\theta \cdot \tau_1\} d\theta$.

Substituindo (2.101) em (2.99) obtemos

$$p(x_1, \ldots, x_n) = c\Pi_{i=1}^n h(x_i) k\left(n + \frac{1-a}{a}; \frac{b}{a} + \sum_{i=1}^n x_i\right), \tag{2.102}$$

concluíndo a prova do teorema.

□

Este resultado fornece uma maneira para caracterizar algumas distribuições de probabilidade para variáveis aleatórias contínuas cuja estatística suficiente é $S_n = X_1 + \ldots + X_n$ e sempre que a dimensão das observações X seja igual à dimensão do espaço paramétrico Θ . Se para Você a seqüência de variáveis aleatórias pode ser representada como uma mistura de elementos da família exponencial (em muitos casos isto é traduzido pela percepção de alguma condição de invariância sobre esta seqüência de objetos observáveis) e se Você acha que $E(X_2|X_1)$ é linear em X_1, então Você determina facilmente a distribuição *a priori* conjugada natural com respeito à função de verossimilhança da mistura e também a distribuição Preditiva.

Vejamos o seguinte exemplo para ilustrar essas idéias.

Exemplo 2.14 : Como no exemplo 2.9, considere $X_1, X_2, \ldots$ uma seqüência infinita de variáveis aleatórias tal que para Você ,

$$P(X_{(n)} \in A) = P(X_{(n)} \in A + \mathbf{x}), \quad \forall n \geq 1 \tag{2.103}$$

onde A é um boreliano qualquer de $\mathcal{R}_+^n$ e $\mathbf{x} \in \mathcal{R}^n$ é tal que $\sum_{i=1}^n x_i = 0$ e $A + \mathbf{x} \in \mathcal{R}_+^n$. Suponha que sob P, $E(X_2|X_1) = aX_1 + b$, onde a e b são números reais. A proposição 2.5.2 fornece-nos que a distribuição *a priori* conjugada natural para o parâmetro θ da distribuição exponencial envolvida na mistura explicitada na expressão (1.58) é

$$dQ(\theta) = \frac{1}{k\left(\frac{1-a}{a}; \frac{b}{a}\right)} \theta^{\frac{1-a}{a}} \exp\left\{\frac{b}{a} \cdot \theta\right\} \tag{2.104}$$

onde $k\left(\frac{1-a}{a}; \frac{b}{a}\right) = \frac{\Gamma(1/a)}{(b/a)^{1/a}}$ e, como conseqüência, obtem-se

$$p(x_1, \ldots, x_n) = \frac{k\left(\frac{1-a}{a} + n; \frac{b}{a} + \sum_{i=1}^n x_i\right)}{k\left(\frac{1-a}{a}; \frac{b}{a}\right)}$$

$$= \frac{\Gamma(1/a+n)}{\Gamma(1/a)} \frac{(b/a)^{1/a}}{(b/a+\sum_{i=1}^{n} x_i)^{1/a+n}} \tag{2.105}$$

Para o caso particular de quantidades aleatórias contínuas, a proposição anterior e os resultados sobre família exponencial anteriormente vistos dizem que as distribuições preditiva e *a posteriori* para θ ficam completamente determinados quando Você declara sua percepção sobre os objetos observáveis. Mais alguns exemplos e comentários a este respeito podem ser encontrados no capítulo 3.

A proposição 2.5.3 também nos permite concluir sobre a J-conjugação natural com base apenas em sua opinião sobre objetos observáveis, pois nela fica definida a forma funcional dos elementos pertencentes a família de distribuições que atuam como medidas misturadoras no teorema 2.3.2 estabelecido por Diaconis e Freedman(1990) a partir de uma condição de linearidade sobre a esperança condicional de X_n dado X_{n-1}, $n > 1$. Sempre que a informação recebida inesperadamente agir sobre seu estado de conhecimento modificando os valores atribuídos por Você aos coeficientes a e b da condição espacificada na proposição 2.5.2 para $E(X_2|X_1)$ e mantendo sua percepção sobre a lei de $X_1, \ldots, X_n|X_1 + \ldots + X_n$ teremos J-conjugação natural - isto é garantido pelas proposições 2.5.1 e 2.5.2.

Outros exemplos de J-Conjugação natural podem ser obtidos dos resultados apresentados no capítulo 3, onde apresentamos caracterizações para o modelo linear t partindo apenas de julgamentos de invariância sobre objetos observáveis, além de condições linearidade sobre esperanças condicionais.

Capítulo 3

Caracterização Preditivista de Modelos Lineares t

3.1 Introdução

Bruno de Finetti(1937) ao estabelecer o teorema de representação para quantidades aleatórias reais permutáveis, consegue eliminar do problema de inferência objetos de caráter completamente metafísico os quais, na literatura estatística corrente, são usualmente chamados de **parâmetros**. De fato, uma característica importante de Teoremas de Representação Tipo de Finetti é que descrevem como subjetivistas podem incorporar "parâmetros" em seu modelo (Wechsler(1993)). Parâmetros, dentro desta visão, são conseqüências de algum julgamento de invariância feito sobre a seqüência de observáveis e por isto são algumas vezes chamados de **parâmetros inter-subjetivos** (Dawid(1982)) . A esta linha subjetivista de pensamento denomina-se **preditivismo** - termo que parece ter sido utilizado primeiramente por Cifarelli e Regazzini(1982). Ver também seção 2.3.

Outra grande importância de Teoremas Tipo de Finetti é que eles reforçam o caráter subjetivista da função de verossimilhança, deixando claro que tanto a distribuição *a priori* quanto a função de verossimilhança são componentes da representação gerados unicamente por sua

opinião preditiva sobre a seqüência de observáveis e, portanto, são igualmente subjetivas. Apesar dessas qualidades, teoremas de representação são basicamente teoremas de existência e, sendo assim, nem sempre fornecem explicitamente a forma funcional da função de verossimilhança. Sob permutabilidade e considerando uma seqüência de variáveis aleatórias $0-1$, de Finetti(1937) estabelece que a função de verossimilhança proveniente deste julgamento de invariância é um produto de distribuições Bernoulli. Isto só foi possível neste caso por causa da simplicidade que envolvia as variáveis aleatórias. Geralmente, necessitamos de julgamentos de invariância mais fortes que permutabilidade para este fim. Kingman(1972), partindo do julgamento de invariância ortogonal sobre a seqüência de quantidades aleatórias (que denotaremos por q.a.) observáveis, obtém um teorema de representação onde a função de verossimilhança é um produto de distribuições $\mathcal{N}(0, V)$. Mas tarde, Smith(1981) determina que sob julgamento de $O(\mathbf{1})$-invariância, a distribuição preditiva para qualquer sub-seqüência finita de quantidades aleatórias observáveis é representada pela mistura de distribuições $\mathcal{N}(M, V)$ independentes deixando, com isso, explicita a forma da função de verossimilhança. Diaconis *et alii*(1992), usando teoria de grupos de invariância, representam uma seqüência de q.a. $O(\mathcal{M})$-invariantes, caracterizando o modelo linear usual. Neste caso, a função de verossimilhança também torna-se explicita como veremos mais adiante na seção 3.2. Além destes, vários outros trabalhos nesta linha têm sido desenvolvidos (ver, por exemplo, Iglesias *et alii*(1996),Iglesias(1993) e Silva(1997)). Os resultados supra citados e as definições e das condições de invariância utilizados em cada um deles serão apresentados detalhadamente na próxima seção.

Neste trabalho, trataremos de caracterizar preditivisticamente alguns modelos estatísticos (entenda-se função de verossimilhança e distribuição *a priori*) partindo dos resultados citados anteriormente. Por **caracterização preditivista de modelos** estamos entendendo a inevitabilidade da adoção de particulares modelos (e, conseqüentemente, de particulares distribuições *a priori*), quando certas características são percebidas na seqüência de observáveis, além do apropriado julgamento de invariância. Neste tipo de caracterizações, o modelo estatístico fica completamente determinado a partir de julgamentos sobre observáveis apenas. Uma vez que Você declara sua opinião sobre a seqüência de observáveis, Você está dispen-

sado de declarar opiniões sobre objetos metafísicos eventualmente envolvidos no problema, isto é, sobre aqueles objetos que são comumente chamado de parâmetros.

Resultados nesse sentido podem ser encontrados em Diaconis e Ylvisaker(1979,1985) e Arellano *et alii*(1994). Diaconis e Ylvisaker(1979) obtêm uma caracterização preditivista para algumas distribuições *a priori* Beta, partindo de suposições de permutabilidade e de linearidade da média a posteriori. Também acrescentando a condição de linearidade da média *a posteriori* à condição de invariância adequada, estes mesmos autores caracterizam em (1985) algumas distribuições *a priori* Gama. Em Arellano-Valle *et alii*(1994) pode-se encontrar caracterizações de algumas distribuições *a priori* Gama-inversa e Wishart-inversa.

Embora as caracterizações que obteremos aqui não sejam preditivistas no seu sentido mais puro - para obtê-las consideramos versões infinitas de Teoremas Tipo de Finetti, o que depõe contra o caráter operacional dos parâmetros inter-subjetivos - elas apresentam resultados importantes para a inferência (subjetivista). Através de julgamentos apropriados de invariância e de condições adicionais sobre a esperança e a variância *a posteriori*, obtivemos uma caracterização para algumas distribuições *a priori* Normal-gama-inversa e, conseqüentemente, de alguns modelos lineares t. Seguiremos a linha adotada em Arellano-Valle *et alii*(1994).

Na próxima seção, apresentaremos as definições e interpretações das condições de invariância que utilizaremos neste capítulo, apresentando uma demonstração alternativa para o Teorema de Representação para seqüências $O(\mathcal{M})$-invariantes estabelecido por Diaconis *et alii* (1992). Usando ferramentas matemáticas mais simples que as utilizadas por tais autores, obtivemos uma outra prova para a versão finita deste teorema - a qual é completamente preditivista - que usaremos posteriormente para obtenção da forma infinita. Na seção 3.3, apresentaremos uma caracterização preditivista de alguns modelos lineares t - e, conseqüentemente, de algumas distribuições *a priori* Normal–Gama–inversa - partindo dos teoremas de representação apropriados apresentados na seção 3.2.

3.2 Simetrias Esféricas: Definições e Interpretações

Apesar de sua importância para a Inferência Estatística, sabemos que julgamentos de permutabilidade são insuficientes para descrever todas as situações de interesse na prática estatística. Informalmente, permutabilidade traduz bem sua opinião sobre observáveis somente quando o que Você não consegue distingüir na seqüência de q.a. é a ordem em que elas ocorreram. Por exemplo, em uma seqüência de variáveis aleatórias $0-1$, uma vez conhecida a soma das n primeiras, Você não distingüe a ordem em que os zeros e os uns ocorreram - ou seja, para Você, qualquer permutação deles é igualmente possível de ocorrer. Para os casos que caracterizaremos na seção 3.3, necessitaremos de alguns julgamentos mais fortes de invariância.

Wechsler(1993) descreve a seguinte situação. Suponha que sua opinião sobre as variáveis aleatórias X e Y é traduzida por uma densidade conjunta sobre um plano a qual é considerada por Você como invariante sob rotações ao redor da origem, isto é, Você não tem qualquer preferência sobre pontos em um mesmo círculo centrado na origem. Ao julgar que X e Y se comportam como o descrito acima, Você de fato declara que X e Y são ortogonalmente invariantes. Neste caso, sua opinião sobre X e Y pode ser revelada por sua opinião sobre a distância de (X, Y) à origem.

Para generalizar este conceito, considere o vetor n-dimensional $X_{(n)} = (X_1, \ldots, X_n)^t$ e denote por $O = \{\Gamma, n \times n : \Gamma'\Gamma = I_n\}$ o grupo das matrizes ortogonais de ordem n.

Definição 3.2.1 *(Invariância Ortogonal)*
Uma seqüência finita de quantidades aleatórias reais, $X_1, X_2, \ldots, X_n$, é invariante sob transformações ortogonais - isto é, é Ortogonalmente Invariante - se para cada r a distribuição condicional de $X_{(n)}$ dado $\sum_{i=1}^n X_i^2 = r^2$ é uniformemente distribuída na superfície da esfera n-dimensional de raio r, $r > 0$,e centrada na origem. Ou seja, se para cada $r \in \mathcal{R}$

$$X_{(n)} \mid \sum_{i=1}^n X_i^2 = r^2 \sim \mathcal{U}\left\{(x_1, \ldots, x_n) \in \mathcal{R}^n : \sum_{i=1}^n x_i^2 = r^2\right\}, \quad r^2 > 0 \tag{3.1}$$

Em outras palavras, a distribuição de $X_{(n)}$ é Ortogonalmente Invariante se $X_{(n)} =^d \Gamma X_{(n)}$,

para toda $\Gamma \in O$.

Definição 3.2.2 *Uma seqüência infinita* $X_1, X_2, \ldots$ *de quantidades aleatórias é ortogonalmente invariante se para cada* $n \geq 1$ *e para cada* $r \in \mathcal{R}$ *vale* (3.1) *ou, em outras palavras, se para cada* $n \geq 1$, $X_{(n)} =^d \Gamma X_{(n)}, \quad \forall \ \Gamma \in O$.

OBS: O símbolo $=^d$ significa igual em distribuição

Kingman(1972) oferece uma maneira alternativa de apresentar o conceito de invariância ortogonal para seqüências infinitas de q.a. reais estabelecendo o seguinte teorema de representação.

Teorema 3.2.1 *(Kingman (1972))*
Seja $X_1, X_2, \ldots$ *uma seqüência infinita de quantidades aleatórias reais invariantes sob transformações ortogonais. Então existe uma quantidade aleatória não-negativa* V *tal que, condicional em* V, $X_1, X_2, \ldots,$ *são independentes e identicamente distribuídas com distribuição* $\mathcal{N}(0, V)$. *A recíproca também é válida.*

Uma conseqüência imediata do teorema de Kingman é que sob a hipótese de invariância sob transformações ortogonais, $X_{(n)}$ tem sua distribuição representada como uma mistura de distribuições Normais $\mathcal{N}(0, V)$ pela distribuição *a priori* atribuída à variável aleatória V, a qual pode ser interpretada como $\lim_{n\to\infty} \sum_{i=1}^{n} {x_i}^2/n$ (q.c).

Usando a mesma linguagem de Wechsler(1993), considere um ponto A qualquer (não necessariamente a origem) sobre a reta $X = Y$ e denote por R a reta passando por A e perpendicular à reta $X = Y$. Suponha, como antes, que sua opinião sobre (X, Y) é a densidade conjunta definida sobre o plano de tal forma que Você seja indiferente com respeito a pontos em R equidistantes de A. Então dizemos que X e Y são invariantes sob transformações ortogonais que preservam o vetor de uns. Note que neste caso, sua opinião sobre (X, Y) é a sua opinião sobre A e sobre a distância de (X, Y) a A. Além disto, A é o ponto médio entre tais pontos equidistantes (X, Y).

Mais geralmente, considere $O(\mathbf{1}) = \{\Gamma \in O : \Gamma'\mathbf{1} = \mathbf{1}\}$ o grupo das matrizes ortogonais de ordem n que preservam vetores de uns.

Definição 3.2.3 *($O(\mathbf{1})$-Invariância)*
Uma seqüência finita de q.a. reais $X_1, X_2, \ldots, X_n$ é invariante sob transformações ortogonais que preservam vetores de uns - é $O(\mathbf{1})$-invariante - se para cada $m \in \mathcal{R}$ e para cada r real, a distribuição de $X_{(n)}$ dado $(\bar{X}, \sum_{i=1}^{n}(X_i - \bar{X})^2) = (m, r^2)$ é uniforme na superfície da n-esfera de centro m e raio r, ou seja,

$$X_{(n)} \mid \bar{X} = m, \sum_{i=1}^{n}(X_i - \bar{X})^2 = r^2 \sim \mathcal{U}\left\{(x_1, \ldots, x_n) \in \mathcal{R}^n : \bar{x} = m, \sum_{i=1}^{n}(x_i - m)^2 = r^2\right\}, \tag{3.2}$$

para cada $m, r \in \mathcal{R}$. Ou ainda, podemos dizer que para cada $n \geq 1$ a distribuição de $X_{(n)}$ é $O(\mathbf{1})$-invariante se $X_{(n)} =^d \Gamma X_{(n)}, \quad \forall \ \Gamma \in O(\mathbf{1})$.

Definição 3.2.4 *Uma seqüência infinita $X_1, X_2, \ldots$ é $O(\mathbf{1})$-invariante se para todo $n \geq 1$, o vetor $X_{(n)}$ é distribuído como em (3.2), para cada m e r reais, ou em outras palavras, se para cada $n \geq 1$, $X_{(n)} =^d \Gamma X_{(n)}, \quad \forall \ \Gamma \in O(\mathbf{1})$.*

Assim como no teorema de Kingman(1972), Smith(1981) estabelece que, sob o julgamento de $O(\mathbf{1})$-invariância sobre a seqüência $X_1, X_2, \ldots$, o vetor $X_{(n)}$ tem distribuição representada como uma mistura de distribuições $\mathcal{N}(M, V)$, independentes, pela distribuição *a priori* atribuída à (M, V), fornecendo uma definição alternativa para o conceito de $O(\mathbf{1})$-invariância para seqüências infinitas de quantidades aleatórias reais. Neste caso, M e V são variáveis aleatórias e podem ser interpretadas como $\lim_{n\to\infty} \sum_{i=1}^{n} X_i/n$ (q.c.) e $\lim_{n\to\infty} \sum_{i=1}^{n}(X_i - \bar{X})^2/n$ (q.c.), respectivamente.

Teorema 3.2.2 *(Smith 1981))*
Seja $X_1, X_2, \ldots$ uma seqüência infinita de q.a. reais, $O(\mathbf{1})$-invariante. Então, existem q.a. reais M e V, V não-negativa, tais que condicional em M e V, X_1, X_2, ... são independentes e identicamentes distribuídas com distribuição $\mathcal{N}(M, V)$. A volta também se verifica.

Bernardo e Smith (1994), fornecem mais alguns detalhes sobre a interpretação geométrica desses conceitos, os quais são referenciados por "simetrias esféricas". Ver também Gasco(1997).

Sejam $P^{\mathcal{M}}$ a projeção ortogonal de um vetor n-dimensional em um sub-espaço $\mathcal{M}$ de $\mathcal{R}^n$, $\| \ \|$ a norma euclidiana definida em $\mathcal{R}^n$ e n um inteiro qualquer. Por exemplo, considere $n = 2$, $\mathcal{M}$ como uma reta qualquer passando pela origem e seja A um ponto qualquer nesta reta. Suponha que a sua opinião sobre o vetor bi-dimensional (X, Y) seja a densidade conjunta definida sob o plano e que para Você pontos equidistantes de A, localizados sobre a mesma reta que passa por A e é perpendicular à $\mathcal{M}$ são indistingíveis. Neste caso, dizemos que para Você X e Y são invariantes sob transformações ortogonais que preservam elementos de $\mathcal{M}$, ou seja, são $O(\mathcal{M})$-invariantes. Note que aqui, sua opinião sobre (X, Y) é obtida de sua opinião sobre A - que é $P^{\mathcal{M}}(X, Y)$ - e sobre a distância de (X, Y) à A - isto é, $\|(X, Y) - P^{\mathcal{M}}(X, Y)\|$. Formalmente, considere $O(\mathcal{M}) = \{\Gamma \in O : \Gamma' x = x, x \in \mathcal{M}\}$ o grupo das matrizes ortogonais de ordem n que preservam elementos de $\mathcal{M}$.

Definição 3.2.5 *($O(\mathcal{M})$-Invariância)*
Uma seqüência finita de q.a. reais $X_1, X_2, \ldots, X_n$ é invariante sob transformações ortogonais que preservam elementos de $\mathcal{M}$ - isto é, é $O(\mathcal{M})$-invariante - onde $\mathcal{M}$ é um sub-espaço de $\mathcal{R}^n$ com dimensão $d \leq n$, se a distribuição de $X_{(n)}$ dado $(P^{\mathcal{M}} x_{(n)}, \|x_{(n)} - \mathbf{m}\|^2) = (\mathbf{m}, r^2)$ é uniforme sobre a superfície da n-esfera de centro $\mathbf{m} \in \mathcal{R}^n$ e raio $r^2 > 0$, ou seja,

$$X_{(n)} \mid P^{\mathcal{M}}(X_{(n)}) = \mathbf{m}, \ \|X_{(n)} - \mathbf{m}\| = r^2 \sim \mathcal{U}\{x_{(n)} \in \mathcal{R}^n : P^{\mathcal{M}} x_{(n)} = \mathbf{m}, \|x_{(n)} - \mathbf{m}\|^2) = r^2\}, \tag{3.3}$$

para todo $\mathbf{m} \in \mathcal{R}^n$ e $r \in \mathcal{R}$ e para $P^{\mathcal{M}}$ representando a projeção ortogonal em $\mathcal{M}$. Alternativamente, a distribuição do vetor $X_{(n)}$ é $O(\mathcal{M})$-invariante, se $X_{(n)} =^d \Gamma X_{(n)}, \ \forall \ \Gamma \in O(\mathcal{M})$.

Definição 3.2.6 *Uma seqüência infinita de q.a. reais $X_1, X_2, \ldots$ é $O(\mathcal{M})$-invariante se para cada $n \geq 1$ vale a condição (3.3), ou alternativamente, se $X_{(n)} =^d \Gamma X_{(n)}, \ \forall \ \Gamma \in O(\mathcal{M})$.*

Em linguagem de planejamento de experimentos, dada uma matriz de delineamento $\mathbf{X}$, o vetor de resposta Y é $O(\mathcal{M})$-invariante se para Você todas as possíveis realizações y de

Y que produzem a mesma estimativa $P^{\mathcal{M}}y$ e o mesmo erro de estimação $\|y - P^{\mathcal{M}}y\|^2$ são igualmente prováveis.

Diaconis *et alii*(1992) apresentam este conceito através de um teorema de representação tipo de Finetti, o qual demonstraremos mais simplesmente aqui utilizando-nos basicamente de cálculo de probabilidades e de alguns resultados conhecidos da teoria de modelos lineares. Mostraremos inicialmente que a distribuição das k, $k < n - p$, primeiras coordenadas de um vetor $X_{(n)}$ com distribuição satisfazendo (3.3) tem distribuição Pearson tipo II e a seguir daremos as versões finita e infinita dos teoremas tipo de Finetti para variáveis aleatórias $O(\mathcal{M})$- invariantes.

Para simplificar a notação, chamaremos $\mathcal{P}_{II}(a, B, h)$ a lei de um vetor aleatório T, k-dimensional, cuja distribuição é uma Pearson tipo II com vetor de locação a, k-dimensional; matriz de dispersão B de ordem $k \times k$ e positiva definida; h é um parâmetro de forma e cuja função de densidade é

$$f_T(t) = \left(\frac{1}{\pi}\right)^{k/2} \frac{\Gamma(k/2 + h + 1)}{\Gamma(h + 1)} (\det B)^{-1/2} \left\{1 - (t - a)^t B^{-1}(t - a)\right\}^h, \tag{3.4}$$

onde $h > -1$ e $0 \leq (t - a)^t B^{-1}(t - a) \leq 1$.

Além disso, ao longo deste capítulo consideraremos $\mathcal{M}$ como sendo o espaço gerado pelas colunas de uma matriz de constantes conhecidas $\mathbf{X}$ de ordem $n \times p$ cujo posto é p e denotaremos por I_η a identidade de ordem η.

Considere $Y_{(n)} = (Y_1, \ldots, Y_n)^t$ e defina as funções

$$B_n = B(Y_{(n)}) = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t Y_{(n)} \qquad \text{e} \tag{3.5}$$

$$E_n = E(Y_{(n)}) = (Y_{(n)} - \mathbf{X}B_n)^t(Y_{(n)} - \mathbf{X}B_n). \tag{3.6}$$

Proposição 3.2.1 *Seja $Y_1, \ldots, Y_n$ uma seqüência finita de q.a. reais, $O(\mathcal{M})$-invariante e para cada $k < n - p$, considere a partição*

$$Y_{(n)} = \begin{pmatrix} Y_{(k)} \\ Y_{(n-k)} \end{pmatrix} \quad \text{e} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_k \\ \mathbf{X}_{n-k} \end{pmatrix},$$

onde $\mathbf{X}_{n-k}$ *é uma matriz formada por linhas de* $\mathbf{X}$ *e tem posto completo* p. *Então, para qualquer vetor* $\mathbf{b}$, *de dimensão* $p \times 1$ *e qualquer número real positivo* c *temos que*

$$Y_{(k)} \mid B_n = \mathbf{b}, E_n = c \sim \mathcal{P}_{II}\left(\mathbf{X}_k\mathbf{b};\ c[I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t];\ \frac{n-p-k-2}{2}\right).$$

Prova: Para provar o que desejamos basta verificar que a densidade condicional de $Y_{(k)}$ dados $B_n = \mathbf{b}$ e $E_n = c$ assume a forma dada pela expressão (3.4) com parâmetros de locação $\mathbf{X}_k\mathbf{b}$, matriz de dispersão $c[I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t]$ e parâmetro de forma $h = (n-p-k-2)/2$.

Usando propriedades de cálculo de probabilidades, sabemos que

$$f_{Y(k)|B_n=\mathbf{b},E_n=c}(y_{(k)}) = \frac{f_{E_n|B_n=\mathbf{b},Y_{(k)}=y_{(k)}}(c) f_{B_n|Y_{(k)}=y_{(k)}}(\mathbf{b}) f_{Y_{(k)}}(y_{(k)})}{f_{B_n,E_n}(\mathbf{b},c)}. \tag{3.7}$$

Como por hipótese $Y_1, \ldots, Y_n$ é uma seqüência $O(\mathcal{M})$-invariante, pela definição 3.2.5 temos que

$$Y_{(n)}|B_n = \mathbf{b}, E_n = c \sim \mathcal{U}\{y_{(n)} \in \mathcal{R}^n : B(y_{(n)}) = \mathbf{b}, E(y_{(n)}) = c,\ \ \mathbf{b} \in \mathcal{R}^n,\ c > 0\}, \tag{3.8}$$

qualquer que seja a P-lei de $Y_1, \ldots, Y_n$. Conseqüentemente, para todo $k < n - p$, a distribuição de $Y_{(k)}\ |B_n = \mathbf{b}, E_n = c$ será a mesma para qualquer lei $O(\mathcal{M})$-invariante. Desta forma, sem perda de generalidade, podemos supor que $Y_{(n)}$ tem distribuição $\mathcal{N}(0, I_n)$, ou seja,

$$f_{Y_{(n)}}(y_{(n)}) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right\}. \tag{3.9}$$

Sob condição de normalidade, B_n e E_n são independentes com distribuição $\mathcal{N}(0, (\mathbf{X}^t\mathbf{X})^{-1})$ e χ^2_{n-p}, respectivamente. Sendo assim,

$$f_{B_n,E_n}(b,c) = \left(\frac{1}{2}\right)^{\frac{n}{2}} \left(\frac{1}{\pi}\right)^{\frac{p}{2}} \frac{c^{\frac{n-p}{2}-1}}{\Gamma(\frac{n-p}{2})} \{\det[(\mathbf{X}^t\mathbf{X})^{-1}]\}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(c + \mathbf{b}^t(\mathbf{X}^t\mathbf{X})\mathbf{b})\right\} \tag{3.10}$$

Além disso, lembrando que $\mathbf{X}_{n-k}$ tem posto completo, defina $W = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_{n-k}^t Y_{(n-k)}$, $Q_2 = I_{n-k} - \mathbf{X}_{n-k}(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_{n-k}^t$ e considere $R = Y_{(n-k)}^t Q_2 Y_{(n-k)}$.

Como $Y_{(n-k)} \sim \mathcal{N}(0, I_{n-k})$ e como Q_2 é uma matriz idempotente com posto $n-p-k$, Searle(1982) determina que

$$W \sim \mathcal{N}(0, (\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})(\mathbf{X}^t\mathbf{X})^{-1}) \quad \text{e} \tag{3.11}$$

$$R \sim \chi^2_{n-p-k}. \tag{3.12}$$

Usando cálculo de probabilidades e equivalência de eventos, derivamos da expressão (3.11) que a densidade condicional de B_n dado $Y_{(k)}$ é

$$\begin{aligned} f_{B_n|Y_{(k)}=y_{(k)}}(\mathbf{b}) &= f_W(\mathbf{b} - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)}) \\ &= \left(\frac{1}{2\pi}\right)^{p/2} \frac{\det(\mathbf{X}^t\mathbf{X})}{\{\det(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})\}^{1/2}} \exp\left\{-\frac{1}{2}\left[\mathbf{b}^t(\mathbf{X}^t\mathbf{X})(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}(\mathbf{X}^t\mathbf{X})\mathbf{b}\right.\right. \\ &\quad \left.\left. -2\mathbf{b}^t(\mathbf{X}^t\mathbf{X})(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k^t y_{(k)} + y_{(k)}^t\mathbf{X}_k(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k^t y_{(k)}\right]\right\}. \end{aligned} \tag{3.13}$$

Defina $A = Y_{(n-k)}^t Y_{(n-k)} - Y_{(n-k)}^t\mathbf{X}_{n-k}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_{n-k}^t Y_{(n-k)} - 2Y_{(n-k)}^t\mathbf{X}_{n-k}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t Y_{(k)}$ e seja $g(.) = f_{E_n|B_n=\mathbf{b},Y_{(k)}=y_{(k)}}(.)$. Lembrando que sob normalidade R é independente de $S = (\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_{n-k}^t Y_{(n-k)}$ e usando o Princípio da Substituição verificamos que :

$$\begin{aligned} g(c) &= f_{A|S=(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}[\mathbf{X}^t\mathbf{X}\mathbf{b}-\mathbf{X}_k^t y_{(k)}];Y_{(k)}=y_{(k)}}\left(c - y_{(k)}^t y_{(k)} + y_{(k)}^t\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)}\right) \\ &= f_{Y_{(n-k)}^t Y_{(n-k)}|S=(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}[\mathbf{X}^t\mathbf{X}\mathbf{b}-\mathbf{X}_k^t y_{(k)}]}\left(c - y_{(k)}^t y_{(k)} + y_{(k)}^t\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)} + \right. \\ &\quad \left. +2\left[\mathbf{X}^t\mathbf{X}\mathbf{b}\mathbf{X}_k^t y_{(k)}\right]^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)}\left[\mathbf{X}^t\mathbf{X}\mathbf{b} - \mathbf{X}_k^t y_{(k)}\right]^t(\mathbf{X}^t\mathbf{X})^{-1}\left[\mathbf{X}^t\mathbf{X}\mathbf{b} - \mathbf{X}_k^t y_{(k)}\right]\right) \\ &= f_R\left(c - y_{(k)}^t y_{(k)} + y_{(k)}^t\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)} + 2\mathbf{b}^t\mathbf{X}_k^t y_{(k)} + [\mathbf{b}^t\mathbf{X}^t\mathbf{X} - y_{(k)}^t\mathbf{X}_k](\mathbf{X}^t\mathbf{X})^{-1}\right. \\ &\quad [\mathbf{X}^t\mathbf{X}\mathbf{b} - \mathbf{X}_k^t y_{(k)}] - [\mathbf{b}^t\mathbf{X}^t\mathbf{X} - y_{(k)}^t\mathbf{X}_k](\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}[\mathbf{X}^t\mathbf{X}\mathbf{b} - \mathbf{X}_k^t y_{(k)}] - \\ &\quad \left. -2y_{(k)}^t\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t y_{(k)}\right) \\ &= f_R\left(c - y_{(k)}^t y_{(k)} + \mathbf{b}^t\mathbf{X}^t\mathbf{X}\mathbf{b} - \mathbf{b}^t(\mathbf{X}^t\mathbf{X})(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}(\mathbf{X}^t\mathbf{X})\mathbf{b} + \right. \\ &\quad \left. +2\mathbf{b}^t(\mathbf{X}^t\mathbf{X})(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k^t y_{(k)} - y_{(k)}^t\mathbf{X}_k(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k^t y_{(k)}\right). \end{aligned}$$

De acordo com (3.12) verifica-se então que

$$g(c) = \left(\frac{1}{2}\right)^{\frac{(n-p-k)}{2}} \frac{1}{\Gamma\left(\frac{n-p-k}{2}\right)} \left\{c - (y_{(k)} - \mathbf{X}_k\mathbf{b})^t(I_k + \mathbf{X}_k(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k)(y_{(k)} - \mathbf{X}_k\mathbf{b})\right\}^{\frac{n-p-k}{2}-1}$$

$$\exp\left\{-\frac{1}{2}\left[c - (y_{(k)} - \mathbf{X}_k\mathbf{b})^t[I_k + \mathbf{X}_k(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})^{-1}\mathbf{X}_k](y_{(k)} - \mathbf{X}_k\mathbf{b})\right]\right\}. \tag{3.14}$$

Portanto, substituindo (3.9) com $n = k$, (3.10),(3.13) e (3.14) em (3.7), obtem-se:

$$f_{Y_{(k)}|B_n=b,E_n=c}(y_{(k)}) = \left(\frac{1}{\pi}\right)^{\frac{k}{2}} \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-p-k}{2}\right)} \left\{\frac{\det(\mathbf{X}^t\mathbf{X})}{c^k \det(\mathbf{X}_{n-k}^t\mathbf{X}_{n-k})}\right\}^{\frac{1}{2}} \tag{3.15}$$

$$\left\{1 - (y_{(k)} - \mathbf{X}_k\mathbf{b})^t[c(I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t)]^{-1}(y_{(k)} - \mathbf{X}_k\mathbf{b})\right\}^{\frac{n-p-k}{2}-1}.$$

Finalmente, notando que

$$\begin{aligned} V &= c[I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k] \\ &= \left(\frac{1}{c}[I_k + \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k]\right)^{-1}, \end{aligned} \tag{3.16}$$

conclui-se que $Y_{(k)}|B_n = b, E_n = c$ tem distribuição Pearson tipo II com os parâmetros anteriormente mencionados.

□

Os dois próximos resultados são fundamentais na prova da versão finita do teorema tipo de Finetti para seqüências $O(\mathcal{M})$-invariantes. O primeiro deles é um caso particular da proposição 2.3 fornecida por Diaconis *et alii*(1992). Optamos por apresentá-la apenas para o caso que nos é de interesse imediato, com o intuito de tornar mais clara a prova da forma finita.

Ao longo deste capítulo, $\mathcal{L}(U)$ denotará a lei de uma quantidade aleatória U e $\| \ \|_V$ denotará a distância de Variação Total.

Lema 3.2.1 *Sejam U e W vetores aleatórios k-dimensionais com distribuições $\mathcal{N}(0, sV_k)$ e $\mathcal{N}(0, sI_k)$, respectivamente, onde s é uma constante positiva e $V_k = I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t$. Então,*

$$\|\mathcal{L}(U) - \mathcal{L}(W)\|_V \leq 2\{(\det(V_k))^{-\frac{1}{2}} - 1\}. \tag{3.17}$$

Prova: Sejam f_U e f_W as funções de densidade de U e W, respectivamente. Usando propriedades da distância de Variação total, sabemos que

$$\|\mathcal{L}(U) - \mathcal{L}(W)\|_V = 2\int \left(\frac{f_U(\mathbf{x})}{f_W(\mathbf{x})} - 1\right) f_W(\mathbf{x})d\mathbf{x}.$$

Agora,

$$\frac{f_U(\mathbf{x})}{f_W(\mathbf{x})} = (\det(V_k))^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{n-p}{c}\mathbf{x}^t(V_k^{-1} - I_k)\mathbf{x}\right\}$$

e como $\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t$ é uma matriz definida não - negativa, $V_k^{-1} - I_k$ também o é. Portanto,

$$\frac{f_U(\mathbf{x})}{f_W(\mathbf{x})} \leq (det(V_k))^{-\frac{1}{2}},$$

seguindo daí o resultado.

□

Para percebermos melhor este resultado consideremos um exemplo simples. Suponha que $U \sim \mathcal{N}(0, V)$ e $W \sim \mathcal{N}(0, 1)$, onde V é um número real positivo menor ou igual que 1. (Note que esta exigência é feita para que $V^{-1} - 1$ seja não-negativo). Neste caso,

$$\|\mathcal{L}(U) - \mathcal{L}(W)\|_V \leq \frac{2}{\sqrt{V}} - 2.$$

Se V é muito próximo de zero $\|\mathcal{L}(U) - \mathcal{L}(W)\|_V$ é muito grande, ou seja, a aproximação entre as distribuições de U e W não parece muito adequada. Quando V é muito próxima de zero, a variável aleatória U é praticamente degenerada no ponto zero, o que nos leva a pensar que uma aproximação por uma distribuição $\mathcal{N}(0, 1)$ não é intuitivamente muito razoável.

Por outro lado, se V é muito próximo de 1, a distância de Variação Total entre as leis de U e W é muito próxima de zero, isto é, não erraremos muito ao fazermos uma aproximação entre elas.

Sendo assim, podemos pensar que a Distância de Variação Total fornece uma medida do quão bem sobrepostas estão duas distribuições de probabilidade.

O seguinte resultado estabelece que a lei das k primeiras coordenadas de um ponto uniformemente distribuído na superfície da n - esfera $\mathcal{S}_n = \{(x_1, \ldots, x_n) \in \mathcal{R}^n : P_{\mathcal{M}} x_{(n)} = \mathbf{m}, \|x_{(n)} - \mathbf{m}\|^2) = r^2\}$, $r \in \mathcal{R}$ e $\mathbf{m} \in \mathcal{R}^n$ é próxima (segundo a Distância de Variação Total) à lei de k v.a. independentes, distribuídas segundo uma $\mathcal{N}(\mathbf{X}_k\mathbf{b}; \frac{c}{n-p} I_k)$. Ou seja, usando os resultados da proposição 3.2.1, o teorema que segue estabelece a proximidade, no sentido da métrica assumida aqui, entre a distribuição $\mathcal{P}_{II}(\mathbf{X}_k\mathbf{b}; cV_k; (n-p-k-2)/2)$, onde V_k é definida como para o lema anterior, e uma distribuição $\mathcal{N}(\mathbf{X}_k\mathbf{b}; \frac{c}{n-p} I_k)$.

Seja $Q^k_{n\mathbf{b}c}$ a lei de $Y_{(k)}$ dados $B_n = \mathbf{b}$ e $E_n = c$ e considere P^k_{β,σ^2} a lei de um vetor aleatório com distribuição $\mathcal{N}(\mathbf{X}_k\beta, \sigma^2 I_k)$, $\beta \in \mathcal{R}^p$ e $\sigma^2 > 0$. Para simplificar a apresentação dos resultados, denote a distribuição $\mathcal{P}_{II}(\mathbf{X}_k\mathbf{b}; cV_k; (n-p-k-2)/2)$ por $\mathcal{P}_{II}(\mathbf{X}_k\mathbf{b}; cV_k)$.

Teorema 3.2.3 *Seja $Y_1, \ldots, Y_n$ uma seqüência finita de quantidades aleatórias $O(\mathcal{M})$- invariantes. Então, se $k+3 < n-p$,*

$$\|Q^k_{n\mathbf{b}c} - P^k_{\mathbf{b},\frac{c}{n-p}}\|_V \leq 2\left\{\frac{k+3}{n-p-k-3} + (\det(V_k))^{-\frac{1}{2}} - 1\right\}. \tag{3.18}$$

Prova : Usando propriedades da Distância de Variação Total, a proposição 3.2.1 e o lema anterior obtemos:

$$\begin{aligned}
\|Q^k_{n\mathbf{b}c} - P^k_{\mathbf{b},\frac{c}{n-p}}\|_V &= \|\mathcal{P}_{II}(\mathbf{X}_k\mathbf{b}; cV_k) - \mathcal{N}\left(\mathbf{X}_k\mathbf{b}; \frac{c}{n-p} I_k\right)\|_V \\
&\leq \|\mathcal{P}_{II}(0; cV_k) - \mathcal{N}\left(0; \frac{c}{n-p} V_k\right)\|_V + \|\mathcal{N}\left(0; \frac{c}{n-p} V_k\right) - \mathcal{N}\left(0; \frac{c}{n-p} I_k\right)\|_V \\
&\leq \|\mathcal{P}_{II}(0; I_k) - \mathcal{N}\left(0; \frac{1}{n-p} I_k\right)\|_V + 2\{(\det(V_k))^{-\frac{1}{2}} - 1\}. \qquad (3.19)
\end{aligned}$$

Considere f_N e f_P as funções densidades associadas às distribuições Normais e Pearson tipo II, respectivamente. Usando propriedades da distância de Variação Total temos:

$$\|\mathcal{P}_{II}(0;I_k) - \mathcal{N}\left(0;\frac{1}{n-p}I_k\right)\|_V = 2\int\left(\frac{f_P(y_{(k)})}{f_N(y_{(k)})} - 1\right) f_N(y_{(k)})dy_{(k)}, \tag{3.20}$$

onde

$$\frac{f_P(y_{(k)})}{f_N(y_{(k)})} = \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-p-k}{2}\right)}\left(\frac{n-p}{2}\right)^{-\frac{k}{2}}(1-y_{(k)}^t y_{(k)})^{\frac{n-p-k}{2}-1}\exp\left\{-\frac{1}{2}(n-p)y_{(k)}^t y_{(k)}\right\}. \tag{3.21}$$

Seguindo Eaton(1989) e considerando k par, temos que quando $k+3 < n-p$ a função $(1-u)^{\frac{n-p-k}{2}-1}\exp\{-1/2(n-p)u\}$ é maximizada ao tomarmos $u = \frac{k+2}{n-p}$ e, além disso,

$$\begin{aligned}
\log\left(\frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-p-k}{2}\right)}\left(\frac{n-p}{2}\right)^{-\frac{k}{2}}\right) &= \Sigma_{j=1}^{k/2}\log\left(1-\frac{2j}{n-p}\right) \\
&\leq -\left(\frac{n-p-k}{2}-1\right)\log\left(1-\frac{k+2}{n-p}\right) - \frac{k+2}{2} - \log\left(1-\frac{k+2}{n-p}\right) \\
&= -\frac{n-p-k}{2}\log\left(1-\frac{k+2}{n-p}\right) - \frac{k+2}{2}. \qquad (3.22)
\end{aligned}$$

Daí,

$$\log\frac{f_P(y_{(k)})}{f_N(y_{(k)})} \leq -\log\frac{n-p-k-2}{n-p},$$

e, portanto,

$$\frac{f_P(y_{(k)})}{f_N(y_{(k)})} - 1 \leq \frac{k+2}{n-p-k-2}.$$

Se k é ímpar, por argumentos algébricos similares, temos que

$$\frac{f_P(y_{(k)})}{f_N(y_{(k)})} - 1 \leq \frac{k+3}{n-p-k-3}.$$

E como $\frac{k+2}{n-p-k-2} > \frac{k+3}{n-p-k-3}$, temos que para todo $k+3 < n-p$, $\|Q_{n\mathbf{b}c}^k - P_{\mathbf{b},\frac{c}{n-p}}^k\|_V$ é limitada pela expressão à direita em (3.18).

□

Diaconis *et alii*(1992) em seu teorema 3.1 demonstra que

$$\|Q^k_{\mathbf{nbc}} - P^k_{\mathbf{b},\frac{c}{n-p}}\|_V \leq 2\left\{\left(1-\frac{k+3}{n-p}\right)^{-\frac{1}{2}} - 1\right\} + 2\left\{(\det(V_k))^{-\frac{1}{2}} - 1\right\}. \tag{3.23}$$

Fazendo a diferença entre a cota exibida no teorema 3.2.3 e aquela fornecida por Diaconis *et alii*(1992) na expressão acima verifica-se que, uma vez que $k+3 < n-p$,

$$\frac{k+3}{n-p-k-3} \geq \left(1-\frac{k+3}{n-p}\right)^{-\frac{1}{2}} - 1, \tag{3.24}$$

significando que a cota obtida por Diaconis *et alii*(1992) é mais refinada que a obtida no teorema anterior.

Considerando k e p fixos, notamos que as duas cotas anteriormente vistas convergem uniformemente para zero quando n cresce. Portanto, essa diferença entre elas não tem qualquer efeito para a obtenção da versão infinita de um teorema tipo de Finetti para uma seqüência $O(\mathcal{M})$-invariante. Aqui, daremos preferência à limitante obtida no teorema 3.2.3.

Para formalizarmos a forma finita para uma seqüência $O(\mathcal{M})$-invariante, denotemos por P^k a P-lei do vetor $O(\mathcal{M})$-invariante $Y_{(k)}$, $1 \leq k \leq n$, e consideremos , como antes, P^k_{β,σ^2} denotando a lei de um vetor aleatório k-dimensional com distribuição $\mathcal{N}(\mathbf{X}_k\beta, \sigma^2 I_k)$, $\beta \in \mathcal{R}^p$ e $\sigma^2 > 0$. Note, ainda que a forma finita é uma conseqüência direta do teorema anterior.

Corolário 3.2.1 *(Forma Finita) Seja $Y_1, \ldots, Y_n$ uma seqüência finita de quantidades aleatórias reais $O(\mathcal{M})$-invariantes. Então, existe uma medida de probabilidade μ definida sobre $\mathcal{R}^p \times (0,\infty)$ tal que para cada k com $k+3 < n-p$,*

$$\|P^k - \int_{\mathcal{R}^p\times(0,\infty)} P^k_{\beta,\sigma^2} d\mu(\beta,\sigma^2)\|_V \leq 2\left\{\frac{k+3}{n-p-k-3} + (\det(V_k))^{-\frac{1}{2}} - 1\right\} \tag{3.25}$$

Prova: Sejam ν_n a P-lei de (B_n, E_n) e μ_n a P-lei de $(B_n, E_n/(n-p))$. Como $Y_1, \ldots, Y_n$ é uma seqüência finita de quantidades aleatórias reais $O(\mathcal{M})$-invariantes, pela proposição 3.2.1 temos que

$$\begin{aligned} P^k &= \int_{\mathcal{R}^p\times(0,\infty)} Q^k_{n\mathbf{bc}} d\nu_n(\mathbf{b}, c) \\ &= \int_{\mathcal{R}^p\times(0,\infty)} Q^k_{n\mathbf{bc}^*} d\mu_n(\mathbf{b}, c), \end{aligned} \tag{3.26}$$

onde $c^* = c(n-p)$.

Escolhendo $\mu = \mu_n$,

$$\begin{aligned}\int_{\mathcal{R}^p\times(0,\infty)} P^k_{\beta,\sigma^2} d\mu(\beta,\sigma^2) &= \int_{\mathcal{R}^p\times(0,\infty)} P^k_{\beta,\sigma^2} d\mu_n(\beta,\sigma^2) \\ &= \int_{\mathcal{R}^p\times(0,\infty)} P^k_{\mathbf{b}c} d\mu_n(\mathbf{b},c). \qquad (3.27)\end{aligned}$$

Usando (3.26), (3.27) e a propriedade de convexidade da distância de Variação Total, obtemos

$$\begin{aligned}\|P^k - \int_{\mathcal{R}^p\times(0,\infty)} P^k_{\beta,\sigma^2} d\mu(\beta,\sigma^2)\|_V &= \|\int_{\mathcal{R}^p\times(0,\infty)} \left(Q^k_{n\mathbf{b}c^*} - P^k_{\mathbf{b}c}\right) d\mu(\mathbf{b},c)\|_V \\ &\leq \int_{\mathcal{R}^p\times(0,\infty)} \|Q^k_{n\mathbf{b}c^*} - P^k_{\mathbf{b}c}\|_V d\mu(\mathbf{b},c) \\ &\leq 2\left\{\frac{k+3}{n-p-k-3} + (det(V_k))^{-\frac{1}{2}} - 1\right\},\end{aligned}$$

onde a última desigualdade segue pelo teorema anterior.

□

O próximo passo será exibir a versão infinita do teorema tipo de Finetti para uma seqüência $O(\mathcal{M})$-invariante e um esboço de sua prova. As idéias da prova são fornecidas por Diaconis *et alii*(1992), que não explicitam a demonstração para o caso particular que está sendo considerado aqui. Ressaltamos que a forma infinita será muito útil para a obtenção da particular caracterização (completamente) preditivista para modelos $O(\mathcal{M})$-invariantes como aquelas que veremos na próxima seção. Como definimos anteriormente, sejam $\mathbf{X}$ uma matriz de constantes conhecidas, de dimensão $n \times p$, com posto p, $\mathbf{X}_k$ uma matriz de dimensão $k \times p$ formada a partir de linhas da matriz $\mathbf{X}$ e $V_k = I_k - \mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t$. Considere o seguinte lema:

Lema 3.2.2 *Sejam k e p fixos. Se todos os autovalores de V_k são menores ou iguais a um, então* $\lim_{n\to\infty}\{\det(V_k)\}^{-\frac{1}{2}} = 1$ *se, e somente se* $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0$.

Prova : Denote por $\lambda_j(A)$ o j-ésimo autovalor da matriz não nula A e seja B uma matriz com p colunas.

Suponha que $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0$. Então, para todo j, temos que $\lim_{n\to\infty}\lambda_j(\mathbf{X}^t\mathbf{X})^{-1} = 0$ se, e somente se, $\lim_{n\to\infty}\lambda_{max}(\mathbf{X}^t\mathbf{X})^{-1} = 0$. Mais ainda,

$$\begin{aligned}
\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0 \quad &\Rightarrow \quad \lim_{n\to\infty} B(\mathbf{X}^t\mathbf{X})^{-1}B^t = 0 \\
&\Rightarrow \quad \lim_{n\to\infty} \lambda_j\left(B(\mathbf{X}^t\mathbf{X})^{-1}B^t\right) = 0, \qquad \forall j \\
&\Rightarrow \quad \lim_{n\to\infty} \Pi_{j=1}^{k}\left[1 - \lambda_j\left(B(\mathbf{X}^t\mathbf{X})^{-1}B^t\right)\right] = 1 \\
&\Leftrightarrow \quad \lim_{n\to\infty} \det\left[I_k - \left(B(\mathbf{X}^t\mathbf{X})^{-1}B^t\right)\right] = 1,
\end{aligned}$$

e, conseqüentemente, $\lim_{n\to\infty}\det\left[I_k - (\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t)^{-\frac{1}{2}}\right] = 1$.

Por outro lado, suponha que $\lim_{n\to\infty}\{\det[I_k - (\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t)]\}^{-\frac{1}{2}} = 1$. Como, por hipótese, $0 \leq \lambda_j(V_k) \leq 1$, para todo j, temos que

$$0 \leq \lambda_j(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t) \leq 1 \quad \forall j.$$

Sendo assim, temos

$$\begin{aligned}
\lim_{n\to\infty}\det\left[I_k - \left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right)\right] = 1 \quad &\Leftrightarrow \quad \lim_{n\to\infty}\left[1 - \lambda_j\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right)\right] = 1, \quad \forall j \\
&\Leftrightarrow \quad \lim_{n\to\infty}\lambda_j\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right) = 0, \quad \forall j \\
&\Leftrightarrow \quad \lim_{n\to\infty}\lambda_{max}\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right) = 0.
\end{aligned}$$

Agora, $\lambda_{max}\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right) = \max_{\|d\|=1} d^t\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right)d$ (ver Rao(1973)-pg 591). Então, como desejávamos

$$\begin{aligned}
\lim_{n\to\infty}\max_{\|d\|=1} d^t\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right)d = 0 \quad &\Rightarrow \quad \lim_{n\to\infty} d^t\left(\mathbf{X}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}_k^t\right)d = 0 \\
&\Rightarrow \quad \lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0.
\end{aligned}$$

□

(A prova deste lema nos foi fornecida por del Pino(1996)).

Teorema 3.2.4 *(Forma Infinita) Seja $Y_1, Y_2, \ldots$ uma seqüência infinita de quantidades aleatórias reais $O(\mathcal{M})$-invariantes, onde $\mathcal{M}$ é o espaço gerado pelas colunas da matriz $\mathbf{X}$ com posto completo p. Se $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0$, então existe uma medida de probabilidade μ definida sob $\mathcal{R}^p \times (0,\infty)$ tal que, para cada $k > p$, a P-lei de $Y_{(k)}$ tem a seguinte representação*

$$P^k = \int_{\mathcal{R}^p\times(0,\infty)} \mathcal{N}(\mathbf{X}_k\mathbf{b}, vI_k)d\mu(\mathbf{b}, v). \tag{3.28}$$

Prova: (Aqui faremos apenas um esboço da prova deste teorema. Leitores interessados em seus detalhes podem consultar Gasco(1997)).

Seja μ_n a P-lei da quantidade aleatória$\left(B_n, \frac{E_n}{n-p}\right)$. Considere o conjunto compacto

$$C_K = \left\{ \|B_n\| > K \text{ ou } \frac{E_n}{n-p} > K,\ K \in \mathcal{R}^+ \right\}. \tag{3.29}$$

Para todo $\epsilon > 0$ é possível encontrar C_{K_ϵ} tal que se verifica a condição

$$\mu_n(\bar{C}_{K_\epsilon}) < \epsilon,$$

onde $\bar{C}_{K_\epsilon}$ é o complementar de C_{K_ϵ}. Isto é, temos que a seqüência $\{\mu_n\}$ é tight. Então existe uma sub–seqüência $\{\mu_{n_j}\}$ e uma medida de probabilidade μ tal que $\lim_{j\to\infty}\{\mu_{n_j}\} = \mu$.

Como a expressão à direita em (3.25) converge uniformemente para zero, sempre que $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0$, tomando o limite através da sub–seqüência conclui-se a prova do teorema.

□

Observações :

1. Considere B_n e E_n como definidos nas equações (3.5) e (3.6), respectivamente. Defina $E_n^* = \frac{E_n}{n-p}$. Então, o vetor aleatório (B_n, E_n^*) converge quase certamente, pois ao considerarmos P como a P-lei de uma seqüência de quantidades aleatórias reais cujas marginais k-dimensionais P^k_{β,σ^2} denotam a P-lei da distribuição $\mathcal{N}(\mathbf{X}_k\beta, \sigma^2 I_k)$ e ao definirmos o evento

$$H = \{\mathbf{y} : (B_n(\mathbf{y}), E_n^*(\mathbf{y})) \text{ converge}\}$$

temos que

$$P(H) = \int_{\mathcal{R}^p \times (0,\infty)} P^{\infty}_{\beta,\sigma^2}(H) d\mu(\beta, \sigma^2). \tag{3.30}$$

Sob a P-lei da distribuição Normal dada acima, temos que os Y_i's são independentes e normalmente distribuídos e como, por hipótese, $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = 0$ Anderson e Taylor (1976) provam que B_n é um estimador fortemente consistente, isto é, converge quase certamente quando $n \to \infty$. Sob suposição de normalidade e considerando que $\lim_{n\to\infty} \frac{X^tX}{n} = A$ onde A é uma matriz finita e positiva definida, E^*_n também converge quase certamente (Ver Singer e Sen(1995)-pag 323).

Então,

$$P^{\infty}_{\beta,\sigma^2}(H) = 1$$

e, portanto, $P(H) = 1$.

(Perceba que ao supor que $\lim_{n\to\infty} \frac{X^tX}{n} = A$, garantimos a validade da condição $\lim_{n\to\infty}(X^tX)^{-1} = 0$.)

Além disso, como (B_n, E^*_n) converge quase certamente, existe um único vetor aleatório (B, V) tal que

$$\lim_{n\to\infty}(B_n, E^*_n) = (B, V). \tag{3.31}$$

Seja C um boreliano qualquer de $\mathcal{R}^p \times (0, \infty)$. Como para $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 I)$, $\lim_{n\to\infty}(B_n, E^*_n) = (\beta, \sigma^2)$,

$$\begin{aligned} P\{(B,V) \in C\} &= \int_{\mathcal{R}^p \times (0,\infty)} P^{\infty}_{\beta,\sigma^2}(\lim_{n\to\infty}(B_n, E^*_n) \in C) d\mu(\beta, \sigma^2) \\ &= \int_{\mathcal{R}^p \times (0,\infty)} 1_{\{(\beta,\sigma^2) \in C\}} d\mu(\beta, \sigma^2) \\ &= \mu(C) \end{aligned}$$

onde $1_{\{c \in C\}}$ é a função indicadora e assume o valor 1 se $c \in C$ e assume o valor zero, caso contrário.

Portanto, a P-lei de (B, V) é μ e, pela unicidade do limite na expressão (3.31), μ é única.

2. A volta do teorema também é válida. Considere Γ uma matriz ortogonal que preserva elementos de $\mathcal{M}$. Para todo $A \in \mathcal{M}$ e para todo k

$$P^k(\Gamma A) = \int_{\mathcal{R}^p \times (0,\infty)} P^k_{\beta,\sigma^2}(\Gamma A) d\mu(\beta, \sigma^2). \tag{3.32}$$

Como a distribuição Normal é $O(\mathcal{M})$-invariante,

$$P^k_{\beta,\sigma^2}(\Gamma A) = P^k_{\beta,\sigma^2}(A). \tag{3.33}$$

E, portanto, de (3.32) e (3.33) conclui-se que

$$P^k(\Gamma A) = P^k(A). \tag{3.34}$$

Como a suposição $\lim_{n\to\infty} \frac{X^t X}{n} = A$ implica que $\lim_{n\to\infty}(X^t X)^{-1} = 0$, a observação 1 permite-nos formular o teorema anterior da seguinte maneira:

Corolário 3.2.2 *Seja $Y_1, Y_2, \ldots$ uma seqüência infinita de quantidades aleatórias reais $O(\mathcal{M})$-invariantes, onde $\mathcal{M}$ é o espaço gerado pelas colunas da matriz $\mathbf{X}$, $n \times p$ com posto completo p. Se $\lim_{n\to\infty} \frac{\mathbf{X}^t\mathbf{X}}{n} = A$, então existe um vetor aleatório $\mathbf{B}$, $p \times 1$, e uma variável aleatória V, não negativa, tal que para cada $k > p$*

$$Y_{(k)}|\mathbf{B}, V \sim \mathcal{N}(\mathbf{X}_k\mathbf{B}, VI_k). \tag{3.35}$$

A volta também é verdadeira.

Esse resultado exibe uma nova maneira de apresentar o conceito de $O(\mathcal{M})$-invariância, conforme mencionado anteriormente. Além disso, como $\mathbf{X}$ é uma matriz qualquer de constantes conhecidas, vários modelos lineares comuns na prática estatística ficam preditivisticamente caracterizados. Por exemplo, podemos citar os modelos de Análise de Variância, Modelos de Regressão e Modelos de Análise de Covariância.

Exemplo 3.1: Uma pesquisa é realizada com o objetivo de estudar o efeito de alguns fertilizantes na produtividade do Pimentão Magali. Neste estudo estão sendo considerados p fertilizantes diferentes e supõe-se que cada pé-de-pimentão receberá apenas um destes

fertilizantes. Chamemos de Y_i a quantidade de pimentões produzidas pelo i-ésimo pé-de-pimentão, após o recebimento de algum fertilizante de interesse e por $Y_{(n)}$ uma amostra de tamanho n constituída sucessivamente pela quantidade de pimentões produzida cada um dos n_i pés que receberam o fertilizante $i = 1, \ldots, p$, de tal forma que $n = n_1 + \ldots + n_p$.

O pesquisador não tem qualquer preferência por amostras $Y_{(n)}$ que forneçam a mesma inferência sobre o efeito dos tratamentos e o mesmo erro experimental.

Chame de n o número máximo de pés plantados e considere n_j o número destes pés que receberam o fertilizante j, tal que $n = n_1 + \ldots + n_p$.

Denote por X_{ij} uma quantidade que assume o valor 1 se o i-ésimo pé-de-pimentão recebeu o tratamento j e assume zero, caso contrário. Desta forma a matriz de delineamento $\mathbf{X}$ é uma matriz $n \times p$ formada por n_1 linhas com um 1 na primeira coluna e zero nas demais, n_2 linhas com um 1 na segunda coluna e zero nas demais, e assim sucessivamente.

Com essa estrutura, verifica-se que

$$(\mathbf{X}^t\mathbf{X})^{-1} = \operatorname{diag}\left\{\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right\} \tag{3.36}$$

$$\mathbf{X}^t Y_{(n)} = \left(n_1\bar{Y}_1, \ldots, n_p\bar{Y}_p\right)^t \tag{3.37}$$

onde $\bar{Y}_j$ é a produtividade média dos pés que receberam o tratamento j. Temos, portanto, que as expressões (3.5) e (3.6) assumem a seguinte forma:

$$B_n = \left(\bar{Y}_1, \ldots, \bar{Y}_p\right)^t \quad \text{e} \tag{3.38}$$

$$E_n = \sum_{i=1}^{n} Y_i^2 - \sum_{j=1}^{p} n_j \bar{Y}_j^{\,2}, \tag{3.39}$$

ou seja, B_n é o vetor com as produtividades médias para cada fertilizante.

Suponha que para cada $j = 1, \ldots, p$ apenas $k_j < n_j$ pés estejam disponíveis para análise, de forma que a matriz $\mathbf{X}_{n-k}$ tenha posto completo ($k = k_1 + \ldots + k_p$). Então pode-se escrever $\mathbf{X}_k = \operatorname{diag}\left\{\mathbf{1}_{k_1 \times 1}, \ldots, \mathbf{1}_{k_p \times 1}\right\}$, onde $\mathbf{1}_{a \times b}$ denota uma matriz de uns de ordem $a \times b$.

Com isso, a distribuição condicional em $B_n = \mathbf{b}$ e $E_n = c$, do vetor $Y_{(k)}$ composto pelas k variáveis respostas observáveis $Y_1, \ldots, Y_k$ é a Person do tipo II fornecida na proposição 3.2.1 com parâmetro de locação

$$\mathbf{X_k b} = \begin{pmatrix} b_1 \mathbf{1}_{k_1 \times 1} \\ \vdots \\ b_p \mathbf{1}_{k_p \times 1} \end{pmatrix}$$

e matriz de dispersão

$$D = \operatorname{diag}\left\{ I_{k_1} - \frac{1}{n_1}\mathbf{1}_{k_1 \times k_1}, \ldots, I_{k_p} - \frac{1}{n_p}\mathbf{1}_{k_p \times k_p} \right\}.$$

Qual seria o erro se utilizassemos para inferências não a distribuição exata encontrada acima mas uma aproximação pela $\mathcal{N}(\mathbf{X}_k\mathbf{b}, \frac{c}{n-p} I_k)$? O teorema 3.2.3 fornece que a proximidade (no sentido da Distância de Variação Total) entre estas duas distribuições se $k+3 < n-p$ é:

$$\|Q^k_{n\mathbf{b}c} - P^k_{\mathbf{b}, \frac{c}{n-p}}\|_V \leq 2 \left\{ \frac{k+3}{n-p-k-3} + \left\{ \frac{1}{\Pi_{j=1}^p \left(1 - \frac{k_j}{n_j}\right)} \right\}^{\frac{1}{2}} - 1 \right\}. \tag{3.40}$$

Perceba que se para cada fertilizante j a população de pés-de-pimentão (que recebeu este fertilizante) pudesse ser ampliada de tal forma que $\forall j$, $n_j \to \infty$, teríamos

$$\lim_{\forall j, n_j \to \infty} \Pi_{j=1}^p \left(1 - \frac{k_j}{n_j}\right) = 1$$

e, conseqüentemente,

$$\lim_{n \to \infty} \|Q^k_{n\mathbf{b}c} - P^k_{\mathbf{b}, \frac{c}{n-p}}\|_V = 0,$$

ou seja, a aproximação pela distribuição Normal seria, neste caso, muito boa.

Note ainda que quando $n_j \to \infty$, $\forall j = 1 \ldots p$, os Y_i's são quase independentes (com

distribuição $\mathcal{N}(\mathbf{X}_k\mathbf{b}, \frac{c}{n-p}I_k)$) e com isto,

$$\lim_{n\to\infty} B_n = \lim_{\forall j, n_j\to\infty} \left(\bar{Y}_1, \ldots, \bar{Y}_p\right)^t = \mathbf{B} = (B_1, \ldots, B_p)^t \quad \text{(q.c)} \tag{3.41}$$

$$\lim_{n\to\infty} E_n^* = \lim_{\forall j, n_j\to\infty} \frac{1}{n-p}\left\{\sum_{i=1}^{n} Y_i^2 - \sum_{j=1}^{p} n_j \bar{Y}_j^{\,2}\right\} = V \quad \text{(q.c.)}, \tag{3.42}$$

onde B_j pode ser interpretado como a produtividade média da população de pimentões que receberam o fertilizante j e V é a variância da distribuição do efeito aleatório envolvido no problema.

Mais ainda, $\lim_{n\to\infty}(\mathbf{X}^t\mathbf{X})^{-1} = \lim_{\forall j, n_j\to\infty} \operatorname{diag}\left\{\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right\} = 0$, fornecendo (pelo corolário 3.2.2), que para qualquer $k < p$,

$$Y_{(k)}|\mathbf{B}, V \sim \mathcal{N}(\mathbf{X}_k\mathbf{B}, VI_k),$$

onde

$$\mathbf{X_k B} = \begin{pmatrix} B_1\mathbf{1}_{k_1\times 1} \\ \vdots \\ B_p\mathbf{1}_{k_p\times 1} \end{pmatrix}.$$

Observe que a distribuição de Y_{n_j} dados B_n e E_n, para qualquer $j = 1, \ldots, p$, não é dada pela proposição 3.2.1. Nestes casos, a matriz $\mathbf{X}_{n-n_j}$ não terá posto completo, obrigando-nos a usar outros métodos para determiná-la.

Leitores interessados em mais alguns exemplos podem encontrá-los em Diaconis *et alii* (1992). Na próxima seção, usaremos os resultados e definições apresentados aqui para fornecer caracterizações completamente preditivistas de alguns modelos estatísticos(isto é, distribuição *a priori* e função de verossimilhança).

3.3 Preditivismo e o Modelo Linear t

Em Arellano-Valle *et alii*(1994) são apresentadas caracterizações preditivistas para alguns modelos probabilísticos. Particularmente, caracterizam uma distribuição preditiva t central - e, conseqüentemente, uma distribuição *a priori* Gama-Inversa para o parâmetro inter-subjetivo envolvido no teorema 3.2.1 (teorema de Kingman(1972)) - adicionando à condição de invariância ortogonal percebida na seqüência infinita de variáveis aleatórias uma condição de linearidade sobre a variância *a posteriori*. Ou seja, asssumindo que um vetor aleatório n-dimensional $X_{(n)}$ é distribuído segundo uma distribuição t n-variada, com parâmetros m, c, D e d - abreviadamente, $X_{(n)} \sim t_n(\mathbf{m}; C; D; d)$ - quando sua densidade é dada por:

$$f_{X_{(n)}}(\mathbf{x}) = \frac{\Gamma(\frac{d+n}{2})D^{\frac{d}{2}}|C|^{-\frac{1}{2}}}{\pi^{n/2}\Gamma(\frac{d}{2})}\left\{D + (\mathbf{x}-\mathbf{m})^t C^{-1}(\mathbf{x}-\mathbf{m})\right\}^{-\frac{d+n}{2}} \tag{3.43}$$

onde $\mathbf{m}$ é o parâmetro de locação cuja ordem é $n \times 1$; C é a matriz de dispersão com ordem $n \times n$; D e d são reais positivos - com d igual ao número de graus de liberdade e D sendo o parâmetro de escala (note que quando $D = d$, temos a distribuição t de Student usual) , Arellano-Valle *et alii*(1994) estabelecem o seguinte teorema:

Teorema 3.3.1 *(Arellano-Valle; Bolfarine, Iglesias (1994))*
Seja $X_1, X_2, \ldots$ uma seqüência infinita de quantidades aleatórias reais Ortogonalmente Invariantes sob P. Então,

$$\mathrm{Var}(X_2|X_1) = aX_1^2 + b, \quad a \in (0,1) \quad \text{e} \quad b > 0 \tag{3.44}$$

se, e somente se, $X_{(n)} \sim t_n(\mathbf{0}; I_n; \frac{b}{a}; \frac{(a+1)}{a})$.

Para a prova desse teorema, além do teorema 3.2.1 estabelecido por de Kingman(1972), os autores utilizaram-se do teorema 2.5.3 estabelecido por Diaconis e Ylvisaker(1979) visto no capítulo 2, o que permitiu concluir que a distribuição misturadora para o teorema de Kingman(1972), sob a condição (3.44) para a variância *a posteriori*, é a distribuição $\mathcal{GI}\left(\frac{b}{2a}; \frac{a+1}{2a}\right)$ cuja densidade é

$$f_V(v) \quad \propto \quad (v)^{-\{\frac{a+1}{2a}+1\}} \exp\{-\frac{b}{2a}v^{-1}\}. \tag{3.45}$$

Note que esse resultado retira do problema de inferência a arbitrariedade da adoção de particulares distribuições *a priori* conjugadas para o parâmetro V do teorema 3.2.1 estabelecido por Kingman(1972), as quais são escolhidas meramente com o intuito de facilitar cálculos. A distribuição *a priori* Gama-Inversa, conjugada natural com respeito a distribuição $\mathcal{N}(0, V)$, dada na expressão (3.45) surge como conseqüência de julgamentos que Você faz *a priori* sobre a seqüência de observáveis refletindo, portanto, sua opinião sobre V. Essa distribuição *a priori* certamente não será adequada para o problema se a condição (3.44) não for uma avaliação razoável para a seqüência de variáveis aleatórias considerada.

Os resultados que seguem serão utilizados para provar o resultado que veremos mais adiante, na proposição 3.3.1.

Considere $\mathbf{1}_{i\times j}$ uma matriz de un's de ordem $i \times j$, i, j inteiros positivos quaisquer.

Teorema 3.3.2 *Seja $X_{(n)} = (X_1, \ldots, X_n)^t$. Se $X_{(n)}$ dado μ e σ^2, $\mu \in \mathcal{R}$ e $\sigma^2 > 0$, tem distribuição $\mathcal{N}(\mu\mathbf{1}_{n\times 1}, \sigma^2 I_n)$ e se a distribuição a priori de (μ, σ^2) é a Normal-Gama-Inversa tal que $\mu|\sigma^2 \sim \mathcal{N}(m, \sigma^2\tau)$, $m \in \mathcal{R}$ e $\tau > 0$, e $\sigma^2 \sim \mathcal{GI}(\frac{a}{2}, \frac{d}{2})$, $a > 0$ e $d > 0$, então*

$$X_{(n)} \sim t_n(m\mathbf{1}_{n\times 1}; V; a; d) \tag{3.46}$$

onde $V = (I_n + \tau\mathbf{1}_{n\times n})$.

Uma prova deste resultado pode ser encontrada em O'Hagan(1994).

Para o próximo resultado consideraremos a notação previamente estabelecida nesta seção para a distribuição t n-variada.

Lema 3.3.1 *Seja $X_{(n)} \sim t_n(\mathbf{m}; C; D; d)$. Então,*

i. $E(X_{(n)}) = \mathbf{m}$ *se* $d > 1$ *e* $V(X_{(n)}) = \frac{D}{d-2}C$ *se* $d > 2$;

ii. *Sejam A uma matriz $m \times n$ e $\mathbf{b}$ um vetor $m \times 1j$ Se A e $\mathbf{b}$ são não-estocásticos*

$$AX_{(n)} + \mathbf{b} \sim t_n(A\mathbf{m} + \mathbf{b}; ACA^t; D; d);$$

iii. *Considerando a partição:*

$$X_{(n)} = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}; \qquad \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \quad \text{e} \qquad C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

onde $X_{(1)}$ *e* $\mathbf{m}_1$ *são vetores* $m \times 1$ *e* C_{11} *é uma matriz* $m \times m$ *temos,*

$$X_{(1)}|X_{(2)} = x_{(2)} \sim t_m(\mathbf{m}_{1.2}; C_{11.2}; D_1; d_1),$$

onde

$$\begin{aligned} \mathbf{m}_{1.2} &= \mathbf{m}_1 + C_{12}C_{22}^{-1}(x_{(2)} - \mathbf{m}_2); \\ C_{11.2} &= C_{11} - C_{12}C_{22}^{-1}C_{21}; \\ D_1 &= D + (x_{(2)} - \mathbf{m}_2)^t C_{22}^{-1}(x_{(2)} - \mathbf{m}_2) \\ d_1 &= d + n - m, \quad 1 \leq m \leq n. \end{aligned}$$

Este lema pode ser encontrado em Arellano e Bolfarine(1994) que o estendem para distribuições Elípticas das quais a distribuição t é um caso particular.

A seguinte proposição estabelece uma caracterização completamente preditivista para uma distribuição t não-central. Não encontramos um resultado similar na literatura, o que sugere que o próximo resultado seja uma das principais contribuições neste capítulo.

Proposição 3.3.1 *Seja* $X_1, X_2, \ldots$ *uma seqüência de q.a reais,* $O(\mathbf{1})$*-invariante, tal que*

$$E(X_3^2|X_1, X_2) = a(X_1^2 + X_2^2) + b_2 \qquad \text{e} \tag{3.47}$$

$$E(X_3|X_1, X_2) = a(X_1 + X_2) + b_1 \tag{3.48}$$

onde $a \in (0, 1/2)$, $b_1 \in \mathcal{R}$ *e* $b_2 \in \mathcal{R}^+$ *tal que* $b_2 > \frac{b_1^2}{1-2a}$. *Então, para cada* $n \geq 3$

$$X_{(n)} \sim t_n\left(\frac{b_1}{1-2a}\mathbf{1}_{n\times 1}; I_n + \frac{a}{1-2a}\mathbf{1}_{n\times n}; \frac{1}{a}\left(b_2 - \frac{b_1^2}{1-2a}\right); \frac{1+a}{a}\right). \tag{3.49}$$

A volta também é verdadeira.

Prova: Como por hipótese, $X_1, X_2, \ldots$ é uma seqüência $O(\mathbf{1})$-invariante, pelo teorema 3.2.2 de Smith(1981), existem quantidade aleatórias reais μ e σ^2, σ^2 não-negativa tais que

$$X_{(n)}|\mu,\ \sigma^2 \sim \mathcal{N}(\mu \mathbf{1}_{n\times 1},\ \sigma^2 I_n),$$

onde $X_{(n)} = (X_1, \ldots, X_n)^t$.

Sejam $\bar{X} = (X_2 + X_1)/2$ e $S = \sum_{i=1}^{2}(X_i - \bar{X})^2$. Dados μ e σ^2, $\bar{X}$ e S são q.a. reais independentes com

$$\bar{X}|\mu, \sigma^2 \sim \mathcal{N}\left(\mu, \tfrac{\sigma^2}{2}\right) \quad \text{e}$$

$$S|\mu, \sigma^2 \sim \mathcal{G}\left(\tfrac{1}{2}, \tfrac{1}{2\sigma^2}\right).$$

Desta forma, a distribuição condicional do vetor aleatório $(\bar{X}, S)$ dado (μ, σ^2) é

$$f_{\bar{X},S|\mu,\sigma^2}(\bar{x}, s) = \frac{1}{\pi\sigma^2}\frac{1}{\sqrt{2}}s^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}[2(\bar{x}-\mu)^2 - s]\right\}.$$

Definindo $M = \sum_{i=1}^{2} X_i$ e $Q = \sum_{i=1}^{2} X_i^2$ e usando o método do Jacobiano, temos da expressão anterior que

$$f_{M,Q|\mu,\sigma^2}(m, q) = \tfrac{1}{\sqrt{2}\pi}(q + \tfrac{m^2}{2})^{-\frac{1}{2}} \quad \exp\{-\tfrac{\mu^2}{\sigma^2} - \log 2\sigma^2\}\exp\{(\tfrac{\mu}{\sigma^2}, -\tfrac{1}{2\sigma^2}) \cdot (m, q)\}.$$

Denote por $\theta = (\theta_1, \theta_2) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ o parâmetro natural da distribuição de (M, Q) dado (μ, σ^2) e considere P_θ a distribuição condicional de (M, Q) dado θ.

Sendo assim,

$$dP_\theta(m, q) = \exp\{(\theta_1, \theta_2) \cdot (m, q) - D(\theta)\}d\xi(m, q),$$

onde $d\xi(m, q) = \frac{1}{\pi\sqrt{2}}(q - m^2/2)^{-\frac{1}{2}}d\lambda$; λ é a medida de Lebesgue definida sobre $\mathcal{R}^2$ e $D(\theta) = -\frac{\theta_1^2}{2\theta_2} - \log(-\theta_2)$. (Note que ξ é uma medida σ–finita dominada por λ).

Agora, é fácil verificar que o vetor de derivadas parciais da função $D(\theta)$ com respeito aos parâmetros naturais θ_1 e θ_2 é:

$$D'(\theta) = \left(-\frac{\theta_1}{\theta_2};\ \frac{\theta_1^2}{2\theta_2^2} - \frac{1}{\theta_2}\right) = E\{(M, Q)|\theta\}.$$

Usando propriedades de esperança condicional e as condições (3.47) e (3.48), temos que

$$\begin{aligned}
E\{D'(\theta)|(M,Q)\} &= E\{E\{(M,Q)|\theta_1,\theta_2\}|(M,Q)\} \\
&= E\{E\{(M,Q)|\mu,\sigma^2\}|(M,Q)\} \\
&= E\{E\{(M,Q)|\mu,\sigma^2\}|X_2,X_1\} \\
&= 2E\{E\{(X_3,X_3^2)|(\mu,\sigma^2)\}|X_2,X_1\} \\
&= 2E\{E\{(X_3,X_3^2)|\mu,\sigma^2,X_2,X_1\}|X_2,X_1\} \\
&= 2E\{(X_3,X_3^2)|X_1,X_2\} = \text{ (por hipótese)} \\
&= 2a(X_2+X_1;\ X_2^2+X_1^2)+2(b_1,b_2).
\end{aligned}$$

Pelo teorema 2.5.3 de Diaconis e Ylvisaker(1975) visto no capítulo 2 temos que a distribuição *a priori* para o parâmetro natural θ é

$$d\pi(\theta) = k\exp\left\{\frac{2b_1\theta_1+2b_2\theta_2}{2a}+\frac{(1-2a)}{2a}\left(\frac{\theta_1^2}{2\theta_2}-\log\theta_2\right)\right\}d\theta_1 d\theta_2$$

e, aplicando o método do Jacobiano a esta expressão, encontramos a seguinte distribuição *a priori* para (μ,σ^2):

$$d\pi((\mu,\sigma^2)) = k\left\{\tfrac{1}{\sigma^2}\right\}^{\frac{1}{2a}+\frac{3}{2}}\exp\left\{-\tfrac{1}{\sigma^2}\tfrac{1}{2a}\left(b_2-\tfrac{b_1^2}{(1-2a)}\right)\right\}\left\{\tfrac{(1-2a)}{a\sigma^2}\right\}^{\frac{1}{2}}\exp\left\{-\tfrac{(1-2a)}{2a\sigma^2}\left(\mu-\tfrac{b_1}{(1-2a)}\right)^2\right\}d\mu d\sigma^2. \tag{3.50}$$

Note que função densidade fornecida na expressão (3.50) caracteriza a seguinte distribuição Normal–Gama–Inversa como distribuição *a priori* para (μ,σ^2):

$$\mu|\sigma^2 \sim \mathcal{N}\left(\tfrac{b_1}{1-2a};\tfrac{a\sigma^2}{(1-2a)}\right)$$

$$\sigma^2 \sim \mathcal{GI}\left(\tfrac{1}{2a}\left(b_2-\tfrac{b_1^2}{1-2a}\right);\tfrac{1+a}{2a}\right).$$

Com isso e usando o teorema 3.2.3 segue que

$$X_{(n)} \sim t_n\left(\frac{b_1}{1-2a}\mathbf{1}_{n\times 1};I_n+\frac{a}{1-2a}\mathbf{1}_{n\times n};\frac{1}{a}\left(b_2-\frac{b_1^2}{1-2a}\right);\frac{1+a}{a}\right).$$

Por outro lado, como toda sub-seqüência finita é distribuída segundo uma distribuição t, sua distribuição pode ser representada como uma mistura de normais independentes. Sendo assim, o teorema 3.2.2 de Smith(1981) garante que a seqüência infinita $X_1, X_2, \ldots$ é $O(\mathbf{1})$-invariante.

Além disso, pelo lema 3.3.1, segue que para $n \geq 3$

$$X_3|X_1 = x_1, X_2 = x_2 \sim t\left(a(x_1 + x_2) + b_1;\ 1 + a;\ \lambda_{q(x_1,x_2)};\ (1 + 3a)/2\right)$$

onde $\lambda_{q(x_1,x_2)} = (1 - a)(x_1^2 + x_2^2) - 2b_1(x_1 + x_2) - 2ax_1x_2 + \frac{b_2}{a} - \frac{b_1^2}{a}$.

Daí, temos que

$$\begin{aligned} E(X_3|X_1, X_2) &= a(X_1 + X_2) + b_1, \\ \\ \mathrm{Var}(X_3|X_1, X_2) &= (a - a^2)(X_1^2 + X_2^2) - 2ab_1(X_1 + X_2) - 2a^2X_1X_2 + b_2 - {b_1}^2 \end{aligned}$$

e, conseqüentemente,

$$E(X_3^2|X_1, X_2) = a(X_1^2 + X_2^2) + b_2,$$

concluíndo assim, a prova do teorema.

□

Perceba que, nesse caso, os julgamentos feitos sobre a seqüência de observáveis determina que a função misturadora para o parâmetro inter-subjetivo envolvido no teorema 3.2.2 de Smith(1981) é a Normal–Gama–Inversa fornecida na expressão (3.50) e, como foi comentado para o teorema 3.3.1, sem qualquer tipo de arbitrariedade.

O próximo resultado é a maior contribuição deste capítulo por fornecer uma caracterização completamente preditivista para o Modelo Linear usual, o que significa banir de um amplo leque de problemas estatísticos objetos sem sentido operacional. Além disso, note que este resultado oferece uma extensão da proposição 3.3.1, por perceber que $O(\mathbf{1})$-invariância é um caso particular de $O(\mathcal{M})$-invariância.

Fazendo uso apenas do julgamento de $O(\mathcal{M})$-invariância sobre a seqüência infinita de quantidades aleatórias de interesse e de condições de linearidade para a esperança *a posteriori*

das estatísticas suficientes, foi possível caracterizar uma distribuição Normal–Gama–Inversa específica como sendo a distribuição *a priori* para os parâmetros B e V envolvidos no corolário 3.2.2. Neste caso, a distribuição *a priori* que obtivemos é a distribuição *a priori* conjugada natural - no sentido definido em Bernardo e Smith(1994) e no cap. 2 - para tais parâmetros.

Considere $\mathcal{M}$ o espaço linear gerado pelas colunas da seguinte matriz $\mathbf{X}_m$ de ordem $m \times p$;

$$\mathbf{X_m} = \begin{pmatrix} \mathbf{1}_{\lfloor m/k \rfloor \times 1} \otimes \mathbf{X}_k \\ (I_{Res(m,k)} \quad \mathbf{0})\mathbf{X}_k \end{pmatrix},$$

onde $\mathbf{X}_k$, $p \leq k \leq m$, é uma matriz $k \times p$ e pode ser interpretada como a matriz do delineamento de fato realizado; $\lfloor m/k \rfloor$ é o menor inteiro menor ou igual que m/k; $Res(m,k)$ é o resto da divisão inteira de m por k e $(I_a \quad \mathbf{0})$ é uma matriz de ordem $a \times k$ constituída pela matriz identidade de ordem a e pela matriz de zeros $\mathbf{0}$, cuja dimensão é $a \times (k-a)$.

Defina $Y^i_{(k)} = (Y_{(i-1)k+1}, \ldots, Y_{ik})^t$, $i = 1, 2, \ldots$, e considere as respectivas estatísticas suficientes para os parâmetros B e V do corolário 3.2.2,

$$T_1 = X_k^t Y^1_{(k)} \tag{3.51}$$

$$T_2 = (Y^1_{(k)})^t Y^1_{(k)}. \tag{3.52}$$

Considerando a matriz $\mathbf{Y}_{(nk)} = (Y_1, \ldots, Y_{nk}) = (Y^1_{(k)}, \ldots, Y^n_{(k)})$ de dimensão $n \times k$, estabelecemos o seguinte resultado:

Proposição 3.3.2 *Seja $Y_1, Y_2, \ldots$ uma seqüência infinita de quantidades aleatórias reais $O(\mathcal{M})$-invariantes. Sejam a e b_2 números reais positivos e $\mathbf{b}_1$ um vetor de constantes qualquer com dimensão $p \times 1$ com $p \leq m$. Se $\lim_{m\to\infty}(\mathbf{X}_m^t \mathbf{X}_m)^{-1} = 0$, então*

$$E(X_k^t Y^2_{(k)} | Y^1_{(k)}) = aT_1 + \mathbf{b}_1 \quad \text{e} \tag{3.53}$$

$$E((Y^2_{(k)})^t Y^2_{(k)} | Y^1_{(k)}) = aT_2 + b_2, \tag{3.54}$$

se, e somente se, para cada par de inteiros positivos n e k, com $k > p$ tais que X_{n-k} tenha posto completo, obtiver-se

$$\mathbf{Y}_{(nk)} \sim t\left(\mathbf{1}_{n\times 1}\otimes \frac{\mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{1-a};\quad I_n\otimes I_k+(\mathbf{1}_{n\times 1}\otimes \mathbf{X}_k)\frac{a}{1-a}(\mathbf{X}_k^t\mathbf{X}_k)^{-1}(\mathbf{1}_{n\times 1}\otimes \mathbf{X}_k)^t;\right.$$
$$\left.\frac{b_2}{a}-\frac{\mathbf{b}_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{a(1-a)};\quad \frac{k(1-a)+(p+2)a}{a}\right). \tag{3.55}$$

Prova : Uma vez que por hipótese $\lim_{m\to\infty}(\mathbf{X}_m^t\mathbf{X}_m)^{-1}=0$, pelo corolário 3.2.2 existem um vetor aleatório B e uma variável aleatória positiva V tal que

$$\mathbf{Y}_{(nk)}|B,V \sim \mathcal{N}\left(\mathbf{1}_{n\times 1}\otimes \mathbf{X}_kB;\quad VI_n\otimes I_k\right). \tag{3.56}$$

Sendo assim, fixando $\mathbf{X}_k$ e dados B e V, a distribuição conjunta das estatísticas T_1 e T_2 anteriormente definidas é caracterizada pela seguinte função de densidade:

$$f_{T_1,T_2|B,V,\mathbf{X}_k}(T_1,T_2) = \mathcal{H}\exp\left\{(T_1,T_2).\left(\frac{B^t}{V};-\frac{1}{2V}\right)+\frac{k}{2}\log\left(\frac{k}{2}\log\left(\frac{1}{2V}\right)-\frac{B^t\mathbf{X}_k^t\mathbf{X}_kB}{2V}\right)\right\} \tag{3.57}$$

onde $\mathcal{H}$ é uma função qualquer dos Y_i's e $U.W$ representa o produto interno entre os vetores U e W.

Denote por $\theta=(\theta_1^t,\theta_2)=\left(\frac{B^t}{V},\quad -\frac{1}{2V}\right)$ o parâmetro natural da distribuição condicional de (T_1,T_2) dados (B,V) e $\mathbf{X}_k$ e chame de $P_{\theta,\mathbf{X}_k}$ a distribuição condicional de (T_1,T_2) dado θ e $\mathbf{X}_k$.

Dados (B,V) e $\mathbf{X}_k$, (T_1,T_2) é uma amostra de tamanho um da família

$$dP_{\theta,\mathbf{X}_k}(T_1,T_2) = \exp\left\{(T_1,T_2).(\theta_1^t,\theta_2)-D(\theta)\right\}d\xi(T_1,T_2), \tag{3.58}$$

onde $d\xi(T_1,T_2)=\mathcal{H}d\lambda$; λ é a medida de Lebesgue definida sobre $\mathcal{R}^{p+1}$ e

$$D(\theta)=-\frac{k}{2}\log(-\theta_2)-\frac{(\mathbf{X}_k\theta_1)^t(\mathbf{X}_k\theta_1)}{4\theta_2}.$$

(Note que ξ é uma medida σ–finita dominada por λ).

Com isso, verificamos que

$$D'(\theta)=\left(-\frac{\mathbf{X}_k^t\mathbf{X}_k\theta_1}{2\theta_2};\quad \frac{k}{2\theta_2}+\frac{(\mathbf{X}_k\theta_1)^t(\mathbf{X}_k\theta_1)}{4\theta_2^2}\right)=E\{(T_1,T_2)|\theta,\mathbf{X}_k\}.$$

Usando as hipóteses (3.53) e (3.54) e propriedades de esperança condicional, temos que

$$
\begin{aligned}
E\{D'(\theta)|(T_1,T_2)\} &= E\{E\{(T_1,T_2)|\theta,\mathbf{X}_k\}|(T_1,T_2)\} \\
&= E\{E\{(T_1,T_2)|B,V,\mathbf{X}_k\}|(T_1,T_2)\} \\
&= E\{E\{(T_1,T_2)|B,V,\mathbf{X}_k\}|Y^1_{(k)}\} \\
&= \left(\mathbf{X}_k^t E\{Y^2_{(k)}|Y^1_{(k)}\};\quad E\{(Y^2_{(k)})^t Y^2_{(k)}|Y^1_{(k)}\}\right) \\
&= \left(E\{\mathbf{X}_k^t Y^2_{(k)}|Y^1_{(k)}\};\quad E\{(Y^2_{(k)})^t Y^2_{(k)}|Y^1_{(k)}\}\right) \\
&= a(T_1,T_2)+(\mathbf{b}_1,b_2).
\end{aligned}
$$

Dessa forma, pelo teorema 2.5.3 estabelecido por Diaconis e Ylvisaker(1979) as hipóteses estão verificadas e, portanto, a distribuição *a priori* para o parâmetro natural θ é absolutamente contínua com relação à medida de Lebesgue e tem a seguinte densidade

$$
d\pi(\theta) = l\exp\left\{\frac{1}{a}\left(\mathbf{b}_1\theta_1^t+b_2\theta_2\right)-\frac{(1-a)}{a}\left(-\frac{k}{2}\log(-\theta_2)-\frac{(\mathbf{X}_k\theta_1)^t(\mathbf{X}_k\theta_1)}{4\theta_2}.\right)\right\}d\theta_1 d\theta_2,
$$

onde l é uma constante apropriada.

Considerando esse resultado e aplicando o método do Jacobiano, a seguinte distribuição *a priori* para (B,V) é obtida

$$
\begin{aligned}
d\pi((B,V)) = {} & l\left\{\frac{1}{V}\right\}^{\frac{k(1-a)+(p+2)a+1}{2a}}\exp\left\{-\frac{1}{2V}\left(\frac{b_2}{a}-\frac{\mathbf{b}_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{a(1-a)}\right)\right\}\left(\frac{1}{V}\right)^{\frac{p}{2}} \\
& \exp\left\{-\frac{1}{2}\left(B-\frac{(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{1-a}\right)^t\frac{\mathbf{X}_k^t\mathbf{X}_k(1-a)}{aV}\left(B-\frac{(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{1-a}\right)\right\}dBdV.
\end{aligned}
\tag{3.59}
$$

Logo, o vetor (B,V) tem como distribuição *a priori* uma distribuição Normal–Gama–

Inversa de tal forma que

$$V \sim \mathcal{GI}\left(\frac{b_2}{2a} - \frac{\mathbf{b}_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{2a(1-a)};\quad \frac{k(1-a)+(p+2)a}{2a}\right) \quad \text{e} \tag{3.60}$$

$$B|V \sim \mathcal{N}\left(\frac{(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{1-a};\quad \frac{aV}{1-a}(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\right). \tag{3.61}$$

Com isso, pelo teorema 3.3.2 verificamos que para todo inteiro $k > p$, $\mathbf{Y}_{(nk)}$ tem a distribuição t nk-variada desejada.

Por outro lado, como toda sub–seqüência $\mathbf{Y}_{(nk)}$ é distribuída segundo uma distribuição t, usando o lema 3.3.1, determinamos que;

$$Y_{(k)}^2|Y_{(k)}^1 \sim t\,(\mathbf{m}_{2.1};\quad C_{2.1};\quad D_2;\quad d_2)\,,$$

onde

$$\begin{aligned}
\mathbf{m}_{2.1} &= \mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}(aT_1+\mathbf{b}_1) = E(Y_{(k)}^2|Y_{(k)}^1),\\
C_{2.1} &= I_k + a\mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{X}_k^t,\\
d_2 &= \frac{k+ap+2a}{a},\\
D_2 &= b_2 + aT_2 - 2aT_1^t\mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1 - a^2T_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}T_1 - \frac{\mathbf{b}_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{a}.
\end{aligned}$$

Com isso obtém-se que

$$\begin{aligned}
\mathrm{Var}(Y_{(k)}^2|Y_{(k)}^1) &= \frac{a(b_2+aT_2-2aT_1^t\mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1 - a^2T_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}T_1) - \mathbf{b}_1^t(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{b}_1}{k+ap}.\\
&\qquad .\left(I_k + a\mathbf{X}_k(\mathbf{X}_k^t\mathbf{X}_k)^{-1}\mathbf{X}_k^t\right)
\end{aligned}$$

e, conseqüentemente,

$$\begin{aligned}
E(\mathbf{X}_k^tY_{(k)}^2|Y_{(k)}^2) &= \mathbf{X}_k^tE(Y_{(k)}^2|Y_{(k)}^1) = aT_1+\mathbf{b}_1, \qquad \text{e}\\
E((Y_{(k)}^2)^tY_{(k)}^2|Y_{(k)}^1) &= \mathrm{tr}(V(Y_{(k)}^2|Y_{(k)}^1)) + (E(Y_{(k)}^2|Y_{(k)}^1))^tE(Y_{(k)}^2|Y_{(k)}^1) = aT_2+b_2.
\end{aligned}$$

Agora, lembrando que toda sub-seqüência distribuída segundo uma distribuição t pode ser representada como uma mistura de Normais, o corolário 3.2.2 garante que $Y_1, Y_2, \ldots$ é uma seqüência $O(\mathcal{M})$-invariante, concluíndo a prova.

□

Exemplo 3.2: Voltando ao exemplo 3.1 da seção anterior, vamos identificar aqui, as distribuições *a priori* para o vetor aleatório $(\mathbf{B}, V)$ do corolário 3.2.2 e o modelo adequado quando as hipóteses da proposição 3.3.2 estiverem satisfeitas.

Assuma que o vetor constante dessa proposição é $\mathbf{b}_1 = (\mathbf{b}_1^1, \ldots, \mathbf{b}_p^1)^t$ e considere como antes que $k = k_1 + \ldots + k_p$. Para cada n inteiro positivo e para cada $k > p$ tal que X_{n-k} tenha posto completo temos que o modelo adequado para quantidades aleatórias $O(\mathcal{M})$-invariantes satisfazendo as condições (3.53) e (3.54) é:

$$\mathbf{Y}_{(\mathbf{nk})} \sim t\left(\mathbf{1}_n \otimes \begin{pmatrix} \frac{b_1^1}{k_1}\mathbf{1}_{k_1\times 1} \\ \vdots \\ \frac{b_p^1}{k_p}\mathbf{1}_{k_p\times 1} \end{pmatrix}; \quad I_n \otimes I_k + \frac{a}{1-a}S; \right.$$

$$\left. \frac{b_2}{a} - \frac{\sum_{j=1}^p \frac{(b_j^1)^2}{k_j}}{a(1-a)}; \quad \frac{k(1-a)+(p+2)a}{a}\right), \tag{3.62}$$

onde

$$S = \begin{pmatrix} \operatorname{diag}\left\{\frac{1}{k_1}\mathbf{1}_{k_1\times k_1}, \ldots, \frac{1}{k_p}\mathbf{1}_{k_p\times k_p}\right\} & \ldots & \operatorname{diag}\left\{\frac{1}{k_1}\mathbf{1}_{k_1\times k_1}, \ldots, \frac{1}{k_p}\mathbf{1}_{k_p\times k_p}\right\} \\ \vdots & & \vdots \\ \operatorname{diag}\left\{\frac{1}{k_1}\mathbf{1}_{k_1\times k_1}, \ldots, \frac{1}{k_p}\mathbf{1}_{k_p\times k_p}\right\} & \ldots & \operatorname{diag}\left\{\frac{1}{k_1}\mathbf{1}_{k_1\times k_1}, \ldots, \frac{1}{k_p}\mathbf{1}_{k_p\times k_p}\right\} \end{pmatrix}.$$

Neste caso, a distribuição *a priori* para $(\mathbf{B}, V)$ é a distribuição Normal-Gama-Inversa dada por

$$V \sim \mathcal{GI}\left(\frac{b_2}{2a} - \frac{\sum_{j=1}^p \frac{(b_j^1)^2}{k_j}}{2a(1-a)}; \quad \frac{k(1-a)+(p+2)a}{2a}\right) \quad \text{e}$$

$$\mathbf{B}|V \sim \mathcal{N}\left(\frac{1}{1-a}\begin{pmatrix} \frac{b_1^1}{k_1}\mathbf{1}_{k_1\times 1} \\ \vdots \\ \frac{b_p^1}{k_p}\mathbf{1}_{k_p\times 1} \end{pmatrix}; \frac{aV}{1-a}\text{diag}\left\{\frac{1}{k_1},\ldots,\frac{1}{k_p}\right\}\right).$$

Alguns dos resultados apresentados aqui serão utilizados posteriormente para fazermos previsões de retornos de preços de ações dentro do contexto de mercados eficientes. Junto com a Regra de Jeffrey Preditiva vista no capítulo 2, estes resultados propiciarão a criação de um algoritmo para previsão incorporando imprevistos(ver capítulo 5).

Devemos notar ainda que os resultados apresentados na seção 3.3 deste capítulo - especificamente, o teorema 3.3.1 e as proposições 3.3.1 e 3.3.2 - fornecem mais alguns exemplos de J-conjugação, anteriormente visto no capítulo 2.

Capítulo 4

Imprevistos e Pontos de Mudança

4.1 Introdução

Como vimos no capítulo 2, a Regra de Jeffrey é um mecanismo de atualização de probabilidades que permite incorporar acontecimentos imprevistos em sua opinião no momento em que Você toma conhecimento de sua ocorrência . Isto só é possível por que a Regra de Jeffrey permite a retaxação arbitrária das probabilidades inicialmente atribuídas a todos os eventos de uma partição. Como conseqüência a Regra Preditivista de Jeffrey fornece um procedimento para se fazer previsões sobre observações futuras, envolvendo neste processo a ocorrência de imprevistos ou informações não previamente consideradas que ocorrem instantes antes de Você declarar suas opiniões.

Neste capítulo voltaremos nossos estudos para uma outra situação de interesse que é o problema de inferência sobre pontos de mudanças. Muitas vezes a ocorrência destes pontos de mudança está relacionada com o acontecimento de eventos imprevistos e a detecção destes pode ser esclarecedora e muitas vezes informativa a respeito do fenômeno que está sendo estudado. Um problema famoso que ilustra a importância da detecção de pontos de mudança é o caso *Watergate* cuja ocorrência provocou uma alteração importante no comportamento do mercado financeiro norte americano e ainda hoje é estudado pelos economistas com o

intuito de descobrir quais fatos realmente levaram o mercado a instabilizar-se. Um outro exemplo que reforça a importância de detecção de pontos de mudança é quando estudamos a evolução de pacientes que sofreram transplantes e tentamos explicar as possíveis causas de rejeição do órgão transplantado. Saber o momento exato em que a rejeição ocorre pode auxiliar na determinação das possíveis causas da rejeição e, conseqüentemente, ajudar na adequação do tratamento de pacientes futuros. Perceba que esta rejeição pode ter ocorrido antes que se manifestasse aparentemente.

Um **ponto de mudança** é um ponto onde se observa uma modificação na estrutura da distribuição de probabilidade que descreve sua incerteza sobre objetos observáveis. Esta mudança pode ser marcada por uma alteração na forma funcional da distribuição ou simplesmente por uma modificação nos parâmetros que indexam tal distribuição sem, no entanto, modificar a sua forma funcional.

Do ponto de vista Bayesiano, o problema de como fazer inferências sobre pontos de mudanças tem sido considerado por vários autores. Por exemplo, Smith(1975) aborda o problema de inferir sobre um único ponto de mudança fornecendo resultados particulares para mudanças na média considerando tanto dados normais como binomiais. Em ambas situações Smith(1975) considera apenas mudanças na média e não na forma funcional da distribuição. Posteriormente, Menzefricke(1981) usa procedimentos similares para fazer inferências sobre um ponto de mudança na precisão usando dados normais com médias e variâncias desconhecidas. Nesta mesma linha Hsu(1982) faz inferências sobre um ponto de mudança permitindo localizar também modificações na forma funcional da distribuição dentro da particular classe das distribuições poder exponencial definida por Box e Tiao(1973). Hartigan(1990) define o Modelo Partição Produto cuja estrutura permite inferir sobre vários pontos de mudança não só nos parâmetros como também na forma funcional da distribuição. Este modelo generaliza todos os casos vistos anteriormente e é utilizado por Barry e Hartigan(1993) para concluir sobre vários pontos de mudança na média de dados normais com variância comum (conhecida). Este problema também foi abordado recentemente por Crowley(1997). Correa(1998) utiliza o Modelo Partição Produto para detectar vários pontos de mudança na variância de dados normais com média conhecida e aplica os resultados obtidos para concluir sobre o

comportamento estocástico da volatilidade do mercado financeiro chileno. Stephens(1994) fornece uma metodologia para inferir sobre vários pontos de mudança quando os pontos de mudança podem ocorrer em tempo contínuo e Green(1995) discute o problema para processos de Poisson.

O Modelo Partição Produto (neste trabalho denotado por MPP) definido por Hartigan(1990) tem-se mostrado bastante eficiente nos estudos de identificação de pontos de mudança e por esta razão dedicaremos a próxima seção deste capítulo ao estudo detalhado deste modelo, destacando algumas de suas propriedades - como a conjugação - e apresentando a sua versão paramétrica de forma menos prolixa do que a exibida em Barry e Hartigan(1992,1993). Na seção 4.3 apresentaremos o procedimento computacional proposto por Yao(1984) para o cálculo das relevâncias, o qual é vital para tornar computacionalmente viável o uso do MPP. Na seção 4.4 estudaremos o MPP para dados condicionalmente normais com média e variância desconhecida e os resultados obtidos nesta seção serão posteriormente utilizados no capítulo 5 para inferir sobre várias mudanças na volatilidade e retornos diários esperados para dados do mercado chileno, estendendo assim os resultados desenvolvidos por Correa(1998) e Crowley(1997). Finalmente, na seção 4.5 apresentaremos o procedimento computacional para o cálculo da distribuição *a posteriori* para vários pontos de mudanças, utilizando técnicas de simulação via cadeia de Markov como foi sugerido em Barry e Hartigan(1993).

4.2 Modelo Partição Produto - MPP

Nesta seção apresentaremos o modelo partição produto (MPP) para o caso particular em que o conjunto de observações de interesse são obtidas seqüencialmente e exploraremos algumas de suas propriedades.

Os componentes básicos dos MPP, no caso que será discutido aqui, são um conjunto de observações $X_1, \ldots, X_n$, $n \geq 1$, feitas seqüencialmente e uma partição aleatória do conjunto de índices $I = \{1, \ldots, n\}$ que será denotada por ρ.

O propósito desse modelo é descrever sua incerteza sobre o objeto aleatório $(X_1, \ldots, X_n, \rho)$

dentro de uma estrutura bem particular que deve traduzir tanto a percepção que Você tem da similaridade (em algum sentido) existente entre elementos de subconjuntos disjuntos destas observações, como a independência que Você julga existir entre estes subconjuntos ou blocos de observações. Uma apresentação formal do MPP poderá ser vista nas definições 4.2.1 e 4.2.2 fornecidas mais adiante nesta seção.

Por simplicidade não diferiremos a notação usada para valores observados daquela usada para valores a observar - todos serão grafados com a letra xis maiúscula, em itálico.

Considere B como sendo a variável aleatória que conta o número de blocos existentes na partição ρ. Então B assume valores no conjunto I.

Como estamos assumindo uma estrutura seqüencial para as observações, o valor $\{i_o, i_1, \ldots, i_b\}$ de ρ satisfazendo a condição $0 = i_o < i_1 < \ldots < i_b = n$, onde b é um valor assumido por B, dividirá o conjunto de observações $X_1, \ldots, X_n$ em b blocos contígüos da seguinte forma

$$[X_1, \ldots, X_{i_1}]\,[X_{i_1+1}, \ldots, X_{i_2}] \ldots [X_{i_{b-1}+1}, \ldots, X_{i_b}]$$

e dentro desta estrutura não será permitido a existência de blocos formados por observações não sucessivas como por exemplo

$$[X_1, X_4, X_6]\,[X_2, X_3, X_5]\,[X_7, \ldots, X_n].$$

(Ver Hartigan(1990) para uma definição mais geral.)

O que faz de duas observações quaisquer X_i e X_j, $i \neq j$, elementos de um mesmo bloco é a percepção que Você tem sobre a semelhança ou associação existente entre elas. Hartigan(1990) define como **coesão** o valor numérico e não-negativo que traduz este grau de associação ou semelhança que Você julga existir entre as observações pertencentes a um mesmo bloco. No caso que estamos considerando, as coesões podem ser interpretadas como as probabilidades de transição da cadeia de Markov $\{Z_k, k \in \mathcal{N}\}$, onde Z_k representa o instante em que ocorre a k-ésima mudança na estrutura da distribuição de $X_1, \ldots, X_n$, isto é, para $k = 0$, Z_k assume o valor $i_0 = 0$ e para $k > 0$, Z_k assume valores no conjunto de inteiros $\{Z_{k-1} + 1, \ldots, n\}$, se $Z_{k-1} \neq n$ e $Z_k = n$ se $Z_{k-1} = n$. Perceba que $\{Z_k, k \in \mathcal{N}\}$ é

a cadeia $\{i_o, i_1, \ldots, i_b\}$ formada pelos pontos finais dos blocos e que ao eliciar suas coesões Você expressa indiretamente sua opinião sobre os pontos $\{i_1+1, \ldots, i_{b-1}+1\}$ onde ocorrem mudanças no comportamento da seqüência observada, isto é, sobre os pontos de mudança da série.

Vendo dessa forma é intuitivo pensar que sua opinião *a priori* sobre a partição aleatória ρ é uma função das coesões previamente determinadas. De fato, assumindo que as observações são seqüencialmente feitas e considerando a estrutura markoviana descrita acima com $P(Z_o = i_o) = 1$ obtemos

$$\begin{aligned} P(\rho = \{i_o, \ldots, i_b\}) &= P(Z_o = i_o, Z_1 = i_1, \ldots, Z_b = i_b) \\ &= P(Z_b = i_b | Z_{b-1} = i_{b-1}) \ldots P(Z_1 = i_1 | Z_o = i_o) P(Z_o = i_o) \\ &= \Pi_{k=1}^{b} p_{i_{k-1} i_k}, \end{aligned} \tag{4.1}$$

onde $p_{i_{k-1} i_k}$ denota a probabilidade de transição do estado i_{k-1} para o estado i_k na cadeia de Markov apresentada anteriormente, $k = 1, \ldots, b$.

Denote por X_{ij} o bloco formado pelas observações $X_{i+1}, \ldots, X_j$ e por c_{ij} a coesão associada ao bloco X_{ij} onde $i, j \in I$.

Definição 4.2.1 *(Modelo Partição Produto)*
Considera-se como Modelo Partição Produto (MPP) qualquer distribuição conjunta sobre as observações $X_1, \ldots, X_n$ e a partição aleatória ρ satisfazendo as seguintes condições:

(i) *a distribuição a priori para ρ é uma* **distribuição produto**, *isto é*

$$P(\rho = \{i_o, \ldots, i_b\}) = K \Pi_{j=1}^{b} c_{i_{j-1} i_j} \tag{4.2}$$

onde $K = \{\Sigma_{\mathcal{C}} \Pi_{j=1}^{b} c_{i_{j-1} i_j}\}^{-1}$ e $\mathcal{C}$ é o conjunto de todas as possíveis partições do conjunto de índices I em b blocos contígüos com pontos finais $\{i_o, \ldots, i_b\}$ satisfazendo a condição $0 = i_o < i_1 < \ldots < i_b = n$, $\forall b \in I$;

(ii) *Condicional em ρ, as observações em diferentes blocos são independentes, isto é,*

$$f(X_1, \ldots, X_n | \rho = \{i_o, \ldots, i_b\}) = \Pi_{j=1}^{b} f_{i_{j-1} i_j}(X_{i_{j-1} i_j}) \tag{4.3}$$

onde $f_{ij}(X_{ij})$ é chamado de fator dado e $b \in I$.

Perceba que um modelo partição produto é adequado para descrever sua opinião sempre que a semelhança existente entre os elementos de um mesmo bloco puder ser traduzida, por exemplo, através de algum julgamento de invariância (ver cap.3 para alguns exemplos), além de ser considerado razoável o julgamento de independência feito ao descrever o comportamento entre os blocos.

Assumida a estrutura do modelo partição produto, a distribuição *a priori* para a variável aleatória número de blocos B pertencentes à partição pode ser obtida por notar a equivalência entre os eventos $\{B = b\}$ e $\bigcup_{\mathcal{C}}[\rho = \{i_0, i_1, \ldots, i_b\}]$ para cada $b \in I$. Deste modo,

$$\begin{aligned} P(B = b) &= P\left(\bigcup_{\mathcal{C}'}[\rho = \{i_0, i_1, \ldots, i_b\}]\right) \\ &= \sum_{\mathcal{C}'} P\left(\rho = \{i_0, i_1, \ldots, i_b\}\right) \\ &= K' \sum_{\mathcal{C}'} \Pi_{j=1}^{b} c_{i_{j-1} i_j}, \end{aligned} \tag{4.4}$$

onde $\mathcal{C}'$ é o conjunto de todas as partições do conjunto de índices I em b(fixo) blocos contígüos com pontos finais $\{i_0, i_1, \ldots, i_b\}$ satisfazendo a condição $0 = i_0 < i_1 < \ldots < i_b = n$.

Uma característica importante dos modelos partição produto é que a distribuição *a posteriori* para ρ tem a mesma forma que a distribuição *a priori*, como pode ser visto no seguinte resultado.

Proposição 4.2.1 *(Barry e Hartigan(1992))*

Se o MPP é assumido como distribuição conjunta das observações $X_1, \ldots, X_n$ e ρ, então a distribuição a posteriori para ρ é uma distribuição produto com coesões

$$c^*_{ij} = c_{ij} f_{ij}(X_{ij}). \tag{4.5}$$

Prova: Supondo válidas as condições (i) e (ii) da definição 4.2.1, temos que para cada $b \in I$,

$$P(\rho = \{i_o, \ldots, i_b\} | X_1, \ldots, X_n) = \frac{f(\rho = \{i_o, \ldots, i_b\}, X_1, \ldots, X_n)}{f(X_1, \ldots, X_n)}$$

$$
\begin{aligned}
&= \frac{f(X_1, \ldots X_n | \rho = \{i_o, \ldots, i_b\}) P(\rho = \{i_o \ldots, i_b\})}{f(X_1, \ldots, X_n)} \\
&= \frac{K}{f(X_1, \ldots, X_n)} \Pi_{j=1}^{b} f_{i_{j-1} i_j}(X_{i_{j-1} i_j}) c_{i_{j-1} i_j} \\
&= K^* \Pi_{j=1}^{b} c^*_{i_{j-1} i_j},
\end{aligned}
$$

como se pretendia demonstrar.

□

Uma vez que a distribuição *a posteriori* para ρ é um produto de coesões *a posteriori* c^*_{ij}, por argumentos análogos aos usados para obter a distribuição *a priori* para B derivamos a sua distribuição *a posteriori*, a qual é dada por

$$
P(B = b | X_1, \ldots, X_n) = K^* \sum_{\mathcal{C}'} \Pi_{j=1}^{b} c^*_{i_{j-1} i_j}, \ \forall b \in I, \tag{4.6}
$$

onde $\mathcal{C}'$ é o conjunto definido na expressão (4.4).

Em muitas situações práticas, como a que abordaremos no capítulo 5, os pontos de mudança são definidos por modificações nos parâmetros que indexam a distribuição dos X_i's e não por alterações em sua forma funcional. Por exemplo, não é pouco razoável supor que em algumas situações ocorra um aumento na volatilidade do mercado mas que permaneçam inalterados o retorno diário esperado e a forma funcional da distribuição que descreve a sua percepção sobre o comportamento dos retornos diários. Por esta razão, torna-se interessante incluir no MPP os parâmetros que indexam as distribuições das observações.

Considerando X_i como a variável aleatória cuja distribuição é indexada pelo parâmetro θ_i, $i \in I$, podemos definir um modelo partição produto envolvendo os parâmetros θ_i's e a partição aleatória ρ da seguinte forma:

Definição 4.2.2 *(Modelo Partição Produto Paramétrico)*
Diz-se que a distribuição conjunta dos objetos aleatórios $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ *e da partição aleatória* ρ *é um Modelo Partição Produto Paramétrico se*

(i) *a distribuição a priori para ρ é uma* **distribuição produto**, *isto é*

$$P(\rho = \{i_o, \ldots, i_b\}) = K\Pi_{j=1}^{b} c_{i_{j-1}i_j}, \tag{4.7}$$

onde $K = \{\Sigma_{\mathcal{C}}\Pi_{j=1}^{b} c_{i_{j-1}i_j}\}^{-1}$ *e* $\mathcal{C}$ *é o conjunto de todas as possíveis partições do conjunto de índices* I *em* b *blocos contígüos com pontos finais* $\{i_o, \ldots, i_b\}$ *satisfazendo a condição* $0 = i_o < i_1 < \ldots < i_b = n$, $b \in I$;

(ii) *Condicional em* ρ, *os objetos aleatórios* $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ *localizados em diferentes blocos são independentes e dentro de um mesmo bloco* X_{ij} *os parâmetros são todos iguais a um parâmetro comum* θ^{ij}, *isto é,*

$$f((X_1, \theta_1), \ldots, (X_n, \theta_n)|\rho = \{i_o, \ldots, i_b\}) = \Pi_{j=1}^{b} f_{i_{j-1}i_j}(X_{i_{j-1}i_j}, \theta^{i_{j-1}i_j}), \tag{4.8}$$

onde $\theta^{ij} = \theta_{i+1} = \ldots = \theta_j$, $f_{ij}(X_{ij}, \theta^{ij}) = f_{ij}(X_{ij}|\theta^{ij})f_{ij}(\theta^{ij})$ *e* $f_{ij}(\theta^{ij})$ *é a distribuição a priori para o parâmetro comum* θ^{ij} *que indexa a distribuição das observações dentro do bloco* X_{ij} *para todo* $i = 1, \ldots, n$ *e* $j > i$.

Note que se conhecéssemos os parâmetros θ_i's, a partição ρ seria conhecida e conseqüentemente seriam conhecidos os pontos de mudança.

Como supostamente as quantidades aleatórias (X_i, θ_i)'s em blocos distintos são condicionalmente independentes dado ρ, as quantidades aleatórias X_i's em diferentes blocos também serão condicionalmente independentes dado ρ, por que estas são funções disjuntas das quantidades aleatórias (X_i, θ_i)'s no bloco correspondente. Desta forma, se *a priori* assumirmos o MPP definido acima a distribuição *a posteriori* para ρ é a distribuição produto dada no teorema 4.2.1, isto é,

$$P(\rho = \{i_o, \ldots, i_b\}) = K^*\Pi_{j=1}^{b} c^*_{i_{j-1}i_j}, \tag{4.9}$$

onde $c^*_{ij} = c_{ij}f_{ij}(X_{ij})$, $f_{ij}(X_{ij}) = \int_{\Theta_{ij}} f(X_{ij}, \theta)d\theta$ e Θ_{ij} é o espaço paramétrico do parâmetro comum θ^{ij} que indexa a distribuição do bloco de observações X_{ij}.

Também verificamos que a distribuição condicional de $\theta_1, \ldots, \theta_n$ dado $\rho = \{i_o, i_1, \ldots, i_b\}$ e $X_1, \ldots, X_n$ é

$$f(\theta_1, \ldots, \theta_n|\rho = \{i_o, \ldots, i_b\}, X_1, \ldots, X_n) = \frac{f((X_1, \theta_1), \ldots, (X_n, \theta_n)|\rho = \{i_o, \ldots, i_b\})}{f(X_1, \ldots, X_n|\rho = \{i_o, \ldots, i_b\})}$$

$$= \frac{\Pi_{j=1}^{b} f_{i_{j-1}i_j}(X_{i_{j-1}i_j}|\theta^{i_{j-1}i_j}) f_{i_{j-1}i_j}(\theta^{i_{j-1}i_j})}{\Pi_{j=1}^{b} f_{i_{j-1}i_j}(X_{i_{j-1}i_j})}$$

$$= \Pi_{j=1}^{b} f_{i_{j-1}i_j}(\theta^{i_{j-1}i_j}|X_{i_{j-1}i_j}), \qquad (4.10)$$

onde $f_{ij}(\theta^{ij}|X_{ij})$ é a distribuição para o parâmetro comum que indexa as distribuições das observações pertencentes ao bloco de observações X_{ij}, após estas serem observadas - isto é, é a distribuição *a posteriori* dentro do bloco X_{ij}.

Esse fato, junto com o resultado obtido no teorema 4.2.1, nos permite concluir que a distribuição conjunta *a posteriori* para os parâmetros $\theta_1, \ldots, \theta_n$ e para a partição aleatória ρ é também um MPP. Ou seja, o MPP paramétrico também preserva *a posteriori* sua estrutura inicial, conduzindo-nos à conjugação. Desta forma, quando é razoável a adoção de um modelo partição produto para a descrição inicial de sua incerteza, inferências pós-observações sobre pontos de mudança são feitas através do produto das coesões *a posteriori* c_{ij}^* anteriormente definidas na expressão (4.9).

Uma vez encontrados os pontos de mudança torna-se importante estimar os parâmetros dentro de cada bloco para se ter uma idéia da dimensão da mudança ocorrida. Para isso é necessário encontrar o comportamento *a posteriori* para cada parâmetro θ_k, $k = 1, \ldots, n$. Obviamente, o comportamento de θ_k depende do bloco a que ele pertence. Consideremos um exemplo simples para entender como derivar a distribuição *a posteriori* de θ_k.

Denote por $[ij]$ o bloco formado pelos índices $\{i+1, \ldots, j\}$.

Exemplo 4.1: Considere uma seqüência X_1, X_2, X_3 onde X_i é uma variável aleatória cuja distribuição está indexada por θ_i, $i = 1, 2, 3$.

Primeiramente, tome $i = 1$. Todas as partições possíveis do conjunto $\{\theta_1, \theta_2, \theta_3\}$ que possuem θ_1 como o possível parâmetro do bloco ao qual pertence são

$$[\theta_1] \ [\theta_2] \ [\theta_3]$$

$$[\theta_1] \ [\theta_2, \theta_3]$$

$$[\theta_1, \theta_2] \ [\theta_3]$$

$$[\theta_1, \theta_2, \theta_3].$$

Nas duas primeiras partições θ_1 pertence a blocos idênticos. Portanto, seu comportamento deve ser descrito em ambas partições por uma mesma distribuição de probabilidade, independentemente de como é particionado o restante do conjunto. Em cada uma das demais partições, θ_1 pertence a blocos distintos dos anteriores e também diferentes em cada uma destas partições. Por esta razão, seu comportamento deverá ser descrito diferentemente em cada uma delas e também distintamente da descrição feita nas duas primeiras. Desta forma, a distribuição *a posteriori* para θ_1 será

$$\begin{aligned} f(\theta_1|X_1, X_2, X_3) &= f_{01}(\theta_1|X_{01})P([01] \in \rho|X_1, X_2, X_3) + f_{02}(\theta_1|X_{02})P([02] \in \rho|X_1, X_2, X_3) \\ &\quad + f_{03}(\theta_1|X_{03})P([03] \in \rho|X_1, X_2, X_3) \\ &= \Sigma_{j=1}^{3} f_{0j}(\theta_1|X_{0j})P([0j] \in \rho|X_1, X_2, X_3), \end{aligned}$$

onde $P([ij] \in \rho|X_1, X_2, X_3)$ denota a probabilidade *a posteriori* do bloco $\{i+1, \ldots, j\}$ pertencer a partição ρ.

Para $i = 2$ obtemos as mesmas partições do conjunto $\{\theta_1, \theta_2, \theta_3\}$ que as obtidas no caso anterior, isto é

$$[\theta_1] \ [\theta_2] \ [\theta_3]$$

$$[\theta_1] \ [\theta_2, \theta_3]$$

$$[\theta_1, \theta_2] \ [\theta_3]$$

$$[\theta_1, \theta_2, \theta_3].$$

Diferente do que ocorria com θ_1, θ_2 sempre pertence a blocos distintos e por isso seu comportamento deve ser descrito diferentemente em cada partição. Fazendo isto obtemos

$$\begin{aligned} f(\theta_2|X_1, X_2, X_3) &= f_{12}(\theta_2|X_{12})P([12] \in \rho|X_1, X_2, X_3) + f_{02}(\theta_2|X_{02})P([02] \in \rho|X_1, X_2, X_3) \\ &\quad + f_{03}(\theta_2|X_{03})P([03] \in \rho|X_1, X_2, X_3) + f_{13}(\theta_2|X_{13})P([13] \in \rho|X_1, X_2, X_3) \\ &= \Sigma_{i=0}^{1}\Sigma_{j=2}^{3} f_{ij}(\theta_2|X_{ij})P([ij] \in \rho|X_1, X_2, X_3). \end{aligned}$$

Analogamente, derivamos que

$$f(\theta_3|X_1, X_2, X_3) = \Sigma_{i=0}^{2} f_{i3}(\theta_3|X_{i3}) P([i3] \in \rho|X_1, X_2, X_3).$$

O resultado mostrado no exemplo acima foi generalizado por Barry e Hartigan(1992) os quais, usando argumentos análogos aos usados no exemplo, estabelecem que

$$\begin{aligned} f(\theta_k|X_1, \ldots, X_n) &= \Sigma_{\mathcal{C}} f(\theta_k, \rho|X_1, \ldots, X_n) \\ &= \Sigma_{\mathcal{C}} f(\theta_k|\rho, X_1, \ldots, X_n) P(\rho|X_1, \ldots, X_n) \\ &= \Sigma_{i=0}^{k-1} \Sigma_{j=k}^{n} f_{ij}(\theta_k|X_{ij}) r([ij]|X_1, \ldots, X_n), \end{aligned} \tag{4.11}$$

onde $r([ij]|X_1, \ldots, X_n)$ é definida como a **relevância** *a posteriori* para o bloco $[ij] = \{i+1, \ldots, j\}$ e representa a probabilidade *a posteriori* do evento "o bloco $[ij]$ pertence a partição aleatória ρ".

Se a regra de penalização adotada é a função de perda quadrática, a estimativa *a posteriori* para cada parâmetro θ_k, $k = 1, \ldots, n$, é a média da distribuição *a posteriori* para θ_k, a qual pode ser determinada através da seguinte expressão

$$\begin{aligned} E(\theta_k|X_1, \ldots, X_n) &= E_{\rho|X_1,\ldots,X_n} \left(E(\theta_k|\rho, X_1, \ldots, X_n)\right) \\ &= \Sigma_{\mathcal{C}} E(\theta_k|\rho, X_1, \ldots, X_n) P(\rho|X_1, \ldots, X_n) \\ &= \Sigma_{i=0}^{k-1} \Sigma_{j=k}^{n} E(\theta_k|X_{ij}) r([ij]|X_1, \ldots, X_n). \end{aligned} \tag{4.12}$$

A expressão (4.12) é obtida utilizando-se de argumentos semelhantes aos usados para derivar a distribuição *a posteriori* de θ_k, $k = 1, \ldots, n$, dada em (4.11). (Ver Correa(1998)).

Como calcular a relevância *a posteriori* do bloco X_{ij}? Devido à complexidade dos cálculos devemos recorrer a procedimentos computacionais para o cálculo desta distribuição *a posteriori*. Estes procedimentos, que foram desenvolvidos por Yao(1984) e também divulgados em Barry e Hartigan(1992,1993), serão apresentados na próxima seção.

4.3 O Cálculo da Relevância

Nesta seção, descreveremos o algoritmo recursivo proposto por Yao(1984) para o cálculo da relevância quando consideramos partições em blocos contígüos que é o caso que nos interessa neste trabalho. Para isto, considere como antes que c_{ij} é a coesão atribuída ao bloco X_{ij} e denote por $r([ij])$ a relevância *a priori* para o bloco X_{ij}, $i, j \in I = \{1, \ldots, n\}$.

Seja $\lambda_{ij} = \Sigma\Pi_{k=1}^{b} c_{i_{k-1}i_k}$, onde a soma é feita sobre todo conjunto de inteiros satisfazendo $i = i_o < i_1 < \ldots < i_b = j$ que particiona o conjunto $\{i+1, \ldots, j\}$ em b blocos com pontos finais $i_o, i_1, \ldots, i_b$, $b \in I$, e denote por B o evento " o bloco $[ij]$ pertence à partição ρ.

Proposição 4.3.1 *Sejam $i, j \in I$. Então a relevância a priori para o bloco X_{ij} é dada por*

$$r([ij]) = \frac{\lambda_{oi} c_{ij} \lambda_{jn}}{\lambda_{on}}. \tag{4.13}$$

Prova: Sejam $\mathcal{C}_1$ o conjunto de todas as partições possíveis do conjunto de índices $I = \{1, \ldots, n\}$ satisfazendo $0 = i_o < i_1 < \ldots < i_k = i < j = i_{k+1} < i_{k+2} < \ldots < i_b = n$, $\mathcal{C}_2$ o conjunto de todas as partições possíveis do conjunto de índices $\{1, \ldots, i_k = i\}$ satisfazendo $0 = i_o < i_1 < \ldots < i_k = i$ e $\mathcal{C}_3$ o conjunto de todas as partições possíveis do conjunto de índices $\{j, \ldots, n\}$ satisfazendo $j = i_{k+1} < i_{k+2} < \ldots < i_b = n$.

Desta forma,

$$\begin{aligned} r([ij]) &= \Sigma_{\mathcal{C}_1} P(\rho = \{i_o, \ldots, i_k = i, j = i_{k+1}, \ldots, i_b\}) \\ &= K \Sigma_{\mathcal{C}_1} \Pi_{r=1}^{b} c_{i_{r-1} i_r} \\ &= K (\Sigma_{\mathcal{C}_2} \Pi_{r=1}^{k} c_{i_{r-1} i_r}) c_{ij} (\Sigma_{\mathcal{C}_3} \Pi_{r=k+2}^{b} c_{i_{r-1} i_r}) \\ &= K \lambda_{oi_k} c_{ij} \lambda_{i_l i_b} = K \lambda_{oi} c_{ij} \lambda_{jn}, \end{aligned}$$

onde K é uma constante obtida da seguinte forma

$$\begin{aligned} K &= \{\Sigma_{\mathcal{C}} P(\rho = \{i_o, \ldots, i_b\})\}^{-1} \\ &= \frac{1}{\lambda_{0n}} \end{aligned} \tag{4.14}$$

e $\mathcal{C}$ é o conjunto apresentado na definição 4.2.1.

□

A proposta feita por Yao(1984) consiste em fazer cálculos recursivos das quantidades λ_{or} e λ_{rn} executando-se o seguinte algoritmo

$$\begin{aligned}
\lambda_{01} &= c_{i_o i_1} = c_{01} \\
\lambda_{0r+1} &= c_{or+1} + \Sigma_{t=1}^{r} \lambda_{ot} c_{tr+1} \\
\lambda_{n-1n} &= c_{n-1n} \\
\lambda_{rn} &= c_{rn} + \Sigma_{t=r+1}^{n-1} \lambda_{tn} c_{tn}.
\end{aligned}$$

Note que usando este algoritmo podemos determinar as estimativas *a posteriori* para os parâmetros θ_k's sem conhecer explicitamente a distribuição *a posteriori* para ρ. Isto significa um grande ganho em tempo computacional por requerer neste processo menos passos do que os exigidos na derivação da distribuição *a posteriori* para ρ.

Se desejamos as relevâncias *a posteriori* para o bloco $[ij]$, basta considerarmos as coesões *a posteriori*

$$c_{ij}^{*} = c_{ij} f(X_{ij})$$

e procedermos similarmente.

4.4 Aplicações a Modelos t

Nesta seção aplicaremos os MPP apresentados anteriormente na seção 4.2 a dados julgados condicionalmente normais com média e variância desconhecidas - isto é, partiremos supondo que cada observação X_i, dado $\theta_i = (\mu_i, \sigma_i^2)$, tem distribuição normal com média μ_i e variância σ_i^2 ambas desconhecidas. Trataremos este problema aqui pois nos interessa aplicá-lo a dados do mercado financeiro chileno para inferir sobre pontos de mudanças tanto no retorno diário esperado como na volatilidade para alguma ação particular deste mercado. Por isso adotaremos algumas distribuições *a priori* particulares as quais serão posteriormente justificadas no capítulo 5.

Seja $\theta^{ij} = (\mu^{ij}, (\sigma^{ij})^2)$ o parâmetro comum que indexa a distribuição conjunta das variáveis aleatórias pertencentes ao bloco X_{ij}, isto é $\theta^{ij} = \theta_{i+1} = \ldots = \theta_j$, $j = 1, \ldots, n$ e $j > i$. Denote por $\mathcal{GI}(a/2; d/2)$, $a > 0$ e $d > 0$, a distribuição Gama Inversa cuja função de densidade é

$$f_V(v) = \frac{(a/2)^{d/2}}{\Gamma(d/2)} v^{-\frac{d+2}{2}} \exp\{-\frac{a}{2v}\}, \quad v > 0 \tag{4.15}$$

e com valor esperado e variância dados, respectivamente, por

$$E(V) = \frac{a}{d-2} \tag{4.16}$$

$$\mathrm{Var}(V) = \frac{2a^2}{(d-2)^2(d-4)}. \tag{4.17}$$

Suponha que sua incerteza *a priori* sobre θ^{ij} seja descrita através da seguinte distribuição Normal-Gama-Inversa

$$\mu^{ij}|(\sigma^{ij})^2 \sim \mathcal{N}(m^{ij}, \tau(\sigma^{ij})^2) \tag{4.18}$$

$$(\sigma^{ij})^2 \sim \mathcal{GI}\left(\frac{a^{ij}}{2}, \frac{d^{ij}}{2}\right). \tag{4.19}$$

onde $m^{ij} \in \mathcal{R}$, τ é um número real positivo, $a^{ij} > 0$ e $d^{ij} > 0$.

Com estas suposições e utilizando o teorema 3.3.2 obtemos que o modelo que traduz sua percepção sobre o comportamento das observações pertencentes ao bloco X_{ij} é o seguinte modelo t, $(j-i)$-variado

$$X_{ij} \sim t_{(j-i)}(m^{ij}\mathbf{1}_{(j-i)\times 1}; v_{ij}; a^{ij}; d^{ij}) \tag{4.20}$$

onde $v_{ij} = I_{(j-i)} + \tau \mathbf{1}_{(j-i)\times(j-i)}$.

Neste caso, se $d^{ij} > 2$ obtém-se que

$$E(X_{ij}) = m^{ij}\mathbf{1}_{(j-i)\times 1} \tag{4.21}$$

$$V(X_{ij}) = \frac{a^{ij}}{d^{ij}-2} V_{ij} \tag{4.22}$$

e a densidade associada ao bloco X_{ij} é dada pela seguinte expressão

$$f(X_{ij}) = \frac{\Gamma\left(\frac{d^{ij}+j-i}{2}\right)(a^{ij})^{d^{ij}/2}}{\pi^{\frac{j-i}{2}}\Gamma(d^{ij}/2)} (1+(j-i)\tau)^{-1/2} \left(a^{ij} + M_{ij}\right)^{-\frac{d^{ij}+(j-i)}{2}} \tag{4.23}$$

onde $M_{ij} = \Sigma_{k=i+1}^{j}(X_k - m^{ij})^2 - \frac{\tau}{1+\tau(j-i)}(\Sigma_{k=i+1}^{j}(X_k - m^{ij}))^2$.

Usando resultados de O'Hagan(1994) - pg. 244 e seguintes - obtemos que para cada parâmetro θ^{ij}, a distribuição condicional nas observações feitas dentro do bloco X_{ij} é a distribuição Normal-Gama-Inversa

$$\mu^{ij}, (\sigma^{ij})^2 | X_{ij} \sim \mathcal{NGI}(a_{ij}^*; d_{ij}^*; m_{ij}^*; v_{ij}^*) \tag{4.24}$$

onde

$$\begin{aligned}
v_{ij}^* &= \frac{\tau}{(j-i)\tau + 1} \\
m_{ij}^* &= \frac{\tau}{(j-i)\tau + 1} \sum_{r=i+1}^{j} X_r + \frac{m^{ij}}{(j-i)\tau + 1} \\
d_{ij}^* &= d^{ij} + j - i \\
a_{ij}^* &= a^{ij} + \frac{(j-i)(m^{ij})^2}{(j-i)\tau + 1} + \Sigma_{r=i+1}^{j} X_r^2 - \frac{2m^{ij}}{(j-i)\tau + 1} \Sigma_{r=i+1}^{j} X_r - \frac{\tau}{(j-i)\tau + 1} (\Sigma_{r=i+1}^{j} X_r)^2
\end{aligned}$$

o que nos permite determinar após alguns cálculos que, dentro de cada bloco,

$$\mu^{ij} | X_{ij} \sim t_1(m_{ij}^*; v_{ij}^*; a_{ij}^*; d_{ij}^*) \tag{4.25}$$

$$(\sigma^{ij})^2 | X_{ij} \sim \mathcal{GI}\left(\frac{a_{ij}^*}{2}; \frac{d_{ij}^*}{2}\right), \tag{4.26}$$

se $a_{ij}^* > 0$

Assumindo como regra de penalidade a função de perda quadrática e considerando $d_{ij}^* > 2$ temos, respectivamente, como estimativas para μ^{ij} e $(\sigma^{ij})^2$ após observarmos o bloco X_{ij},

$$E(\mu^{ij} | X_{ij}) = m_{ij}^* \tag{4.27}$$

$$E((\sigma^{ij})^2 | X_{ij}) = \frac{a_{ij}^*}{d_{ij}^* - 2}, \tag{4.28}$$

para as quais estão associadas, respectivamente, os seguintes erros de estimação

$$\mathrm{Var}(\mu^{ij} | X_{ij}) = \frac{\tau}{(j-i)\tau + 1} \frac{a_{ij}^*}{d_{ij}^* - 2} \tag{4.29}$$

$$\mathrm{Var}((\sigma^{ij})^2 | X_{ij}) = \frac{2(a_{ij}^*)^2}{(d_{ij}^* - 2)^2 (d_{ij}^* - 4)}, \tag{4.30}$$

desde que $d^*_{ij} > 4$.

Conseqüentemente, usando os resultados apresentados na expressão (4.12), obtemos como estimativas *a posteriori* para cada μ_k e para cada σ_k^2, $k = 1, \ldots, n$, respectivamente

$$E(\mu_k|X_1, \ldots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} \left(\frac{\tau \sum_{r=i+1}^{j} X_r}{(j-i)\tau + 1} + \frac{m^{ij}}{(j-i)\tau + 1} \right) r([ij]|X_1, \ldots, X_n) \quad (4.31)$$

$$E(\sigma_k^2|X_1, \ldots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} \frac{a^*_{ij}}{d^*_{ij} - 2} r([ij]|X_1, \ldots, X_n), \quad (4.32)$$

as quais estarão completamente determinadas após calcularmos as relevâncias *a posteriori* $r([ij]|X_1, \ldots, X_n)$.

Para que nossas análises se completem devemos inferir sobre os instantes em que ocorreram mudanças - isto é, necessitamos fornecer a distribuição *a posteriori* para a partição ρ.

Considere p como sendo a probabilidade que ocorra uma mudança em um instante qualquer, $0 \leq p \leq 1$. Então, a coesão *a priori* para o bloco X_{ij} corresponderá à probabilidade de ocorrer uma nova mudança após $(j-i)$ instantes, dado que houve uma mudança no instante i - ou seja,

$$c_{ij} = \begin{cases} p(1-p)^{j-i-1} & \text{se } j < n \\ (1-p)^{j-i-1} & \text{se } j = n. \end{cases} \quad (4.33)$$

Essas coesões *a priori* foram sugeridas por Yao(1984) e também serão adotadas neste trabalho por descreverem razoavelmente bem a situação que apresentaremos no capítulo 5. Perceba que essa escolha de coesões faz com que a seqüência dos pontos de mudança seja um processo de renovação discreto com tempos entre ocorrências identica e geometricamente distribuídos. Assumir que as coesões *a priori* são distribuições geométricas truncadas é considerar que o conhecimento de pontos de mudanças já ocorridos antes de um instante t não é relevante para inferências sobre pontos de mudanças futuros dado que ocorreu uma mudança neste instante t. Além disto é considerar que somente a mudança ocorrida no tempo t infuencia no comportamento das observações $X_{t+1}, X_{t+2}, \ldots$ a serem feitas a partir deste instante e não importa o que foi observado antes de t.

Perceba que esta escolha de coesões favorece a formação de grandes blocos de observações se a probabilidade de ocorrência de uma mudança p for pequena e teremos pequenos blocos se esta probabilidade for grande.

Feitas essas considerações, temos que a distribuição *a priori* para ρ é dada por

$$\begin{aligned} P(\rho = \{i_o, i_1, \ldots, i_b\}) &= \frac{1}{\lambda_{0n}}(1-p)^{i_b - i_{b-1}} \Pi_{k=1}^{b-1} p(1-p)^{i_k - i_{k-1}} \\ &= \frac{1}{\lambda_{0n}} p^{b-1}(1-p)^{n-b}, \end{aligned} \tag{4.34}$$

$b = 1, \ldots, n$ e $n > 0$, a qual depende apenas do número de observações n e do número de blocos existentes na partição b e não das posições onde ocorreram os pontos de mudança.

Com estas particulares escolhas de coesões qualquer partição do conjunto de índices $I = \{1, \ldots, n\}$ em b blocos contígüos tem igual chance de ocorrer. Sendo assim, a distribuição *a priori* para a variável aleatória B que representa o número de blocos pertencentes à partição é dada por

$$P(B = b) = \begin{pmatrix} n-1 \\ b-1 \end{pmatrix} \frac{p^{b-1}(1-p)^{n-b}}{\lambda_{0n}}, \quad \forall\, b \in I, \tag{4.35}$$

onde $p \in (0,1)$ e $\begin{pmatrix} n-1 \\ b-1 \end{pmatrix}$ fornece o número de partições distintas em b blocos contígüos de I.

Também sob estas suposições e usando os resultados apresentados na proposição 4.2.1 e na expressão (4.6) pode-se determinar as distribuições *a posteriori* de ρ e de B. Para este fim serão necessárias as coesões *a posteriori*, as quais no caso normal assumem a seguinte forma

$$c_{ij}^* = \begin{cases} p(1-p)^{j-i-1} \dfrac{\Gamma\left(\frac{d^{ij}+j-i}{2}\right)(a^{ij})^{d^{ij}/2}}{\pi^{\frac{j-i}{2}}\Gamma(d^{ij}/2)} (1+(j-i)\tau)^{-1/2} (a^{ij}+M_{ij})^{-\frac{d^{ij}+(j-i)}{2}} & \text{se } j < n \\ (1-p)^{j-i-1} \dfrac{\Gamma\left(\frac{d^{ij}+j-i}{2}\right)(a^{ij})^{d^{ij}/2}}{\pi^{\frac{j-i}{2}}\Gamma(d^{ij}/2)} (1+(j-i)\tau)^{-1/2} (a^{ij}+M_{ij})^{-\frac{d^{ij}+(j-i)}{2}} & \text{se } j = n. \end{cases} \tag{4.36}$$

Perceba que o cálculo das distribuições *a posteriori* para ρ e B demanda um grande esforço computacional. Na próxima seção desenvolveremos o procedimento computacional sugerido por Barry e Hartigan(1993) para calcular as distribuições *a posteriori* para a partição aleatória ρ e para B o qual esta baseado na geração de amostras através do método de Gibbs.

4.5 Gibbs e MPP: A Distribuição *a Posteriori* para ρ

Determinar a distribuição *a posteriori* para ρ consiste em estabelecer a probabilidade de ocorrência de cada uma das possíveis partições em b blocos contígüos do conjunto de índices I, para todo $b = 1, \ldots, n$, o que envolve o cálculo das 2^{n-1} probabilidades fornecidas pela seguinte expressão

$$P(\rho = \{i_o, i_1, \ldots, i_b\}) = \frac{1}{\lambda^*_{on}} \Pi_{j=1}^{b} c^*_{i_{j-1} i_j}, \quad b = 1, \ldots, n. \tag{4.37}$$

onde c^*_{ij} é a coesão *a posteriori* do bloco $[ij]$, $i, j \in I$ e $j > i$, a qual, no caso normal estudado na seção anterior, é dada pela expressão (4.36).

Devido a sua complexidade, essa tarefa pode ser humanamente impossível de ser satisfatoriamente realizada em situações onde o número n de observações é grande. Portanto faz-se necessário o uso de algum procedimento computacional para realizar tais cálculos. A proposta sugerida por Barry e Hartigan(1993) é derivar a distribuição *a posteriori* para ρ usando o amostrador de Gibbs (ver, por exemplo, Gamerman(1997) para descrição geral do método). Para isso estes autores consideram uma quantidade aleatória auxiliar U_i que reflete se ocorreu ou não um ponto de mudança no instante i, ou seja,

$$U_i = \begin{cases} 1 & \text{se } \theta_i = \theta_{i+1} \\ 0 & \text{se } \theta_i \neq \theta_{i+1}. \end{cases} \tag{4.38}$$

Desta forma, o elemento aleatório ρ é perfeitamente identificável por uma seqüência destas quantidades aleatória $(U_1, \ldots, U_{n-1})$, $n > 0$.

Por exemplo, considerando $n = 3$ percebemos que os seguintes eventos são equivalentes

$$\{\rho = \{0, 1, 2, 3\}\} \quad \Leftrightarrow \quad \{U_1 = 0, U_2 = 0\}$$

$$\{\rho = \{0,2,3\}\} \quad \Leftrightarrow \quad \{U_1 = 1, U_2 = 0\}.$$

Então, genericamente, temos que a probabilidade *a posteriori* para uma particular partição com b blocos contígüos $\{i_o, i_1, \ldots, i_b\}$ do conjunto de índices I é dada por

$$P(\rho = \{i_o, i_1, \ldots, i_b\}|X_{(n)}) = P\,(U_1 = 1, \ldots, U_{i_1-1} = 1, U_{i_1} = 0, U_{i_1+1} = 1, \ldots, U_{i_2-1} = 1,$$

$$U_{i_2} = 0, \ldots, U_{i_{b-1}} = 0, U_{i_{b-1}+1} = 1, \ldots, U_{i_b-1} = 1|X_{(n)}\big) \quad (4.39)$$

onde $X_{(n)}$ é o vetor constituído pelas observações $X_1, \ldots, X_n$.

Uma vez definido este instrumento, a distribuição *a posteriori* para ρ é obtida da seguinte forma. Considere uma amostra inicial $(U_1^0, \ldots, U_{n-1}^0)$ do objeto aleatório $(U_1, \ldots, U_{n-1})$, convenientemente escolhida. Partindo desta amostra inicial, para cada $k > 0$, gere a k - ésima amostra deste objeto aleatório procedendo recursivamente de acordo com o seguinte procedimento. O primeiro elemento da k-ésima amostra U_1^k é um valor gerado da distribuição $P(U_1|U_2^{k-1}, \ldots, U_{n-1}^{k-1}, X_{(n)})$; o segundo U_2^k é um valor gerado da distribuição $P(U_2|U_1^k, U_3^{k-1}, \ldots, U_{n-1}^{k-1}, X_{(n)})$; o terceiro U_3^k é um valor gerado da distribuição $P(U_3|U_1^k, U_2^k, U_4^{k-1} \ldots, U_{n-1}^{k-1}, X_{(n)})$; proceda assim sucessivamente até gerar o $(n-1)$-ésimo elemento da amostra U_{n-1}^k partindo da distribuição $P(U_{n-1}|U_1^k, U_2^k, \ldots, U_{n-2}^k, X_{(n)})$.

Após gerar um número suficientemente grande de amostras do objeto aleatório $(U_1, \ldots, U_{n-1})$ estimamos a probabilidade $P(\rho = \{i_o, i_1, \ldots, i_b\}|X_{(n)})$ para cada possível partição em b blocos contígüos de I pela proporção de seqüências $(U_1, \ldots, U_{n-1})$ da forma especificada na expressão (4.39) que aparecem na amostra utilizada, a qual é formada por seqüências $(U_1, \ldots, U_{n-1})$ escolhidas convenientemente da amostra inicialmente gerada.

Também por esse processo podemos determinar a distribuição *a posteriori* para B por notar que B relaciona-se com as variáveis aleatórias U_i's da seguinte forma

$$B = 1 + \sum_{i=1}^{n-1}(1 - U_i). \quad (4.40)$$

Para cada vetor gerado $(U_1^k, \ldots, U_{n-1}^k)$ obtém-se um valor $B^k = 1 + \sum_{i=1}^{n-1}(1 - U_i^k)$, o que permite concluir que a probabilidade *a posteriori* para B poderia ser estimada pela seguinte

expressão

$$P(B = b|X_{(n)}) = \frac{\sum_k 1_{\{B^k=b\}}}{T}, \quad \forall\, b \in I, \tag{4.41}$$

onde T é o tamanho da amostra de vetores $(U_1, \ldots, U_{n-1})$ utilizada e 1_A representa a função indicador para o evento A.

Com isso, nos falta apenas descrever o procedimento para gerar as seqüências $(U_1^k, \ldots, U_{n-1}^k)$. A idéia é gerar estes vetores via amostrador de Gibbs o qual exige o conhecimento prévio de todas as distribuições condicionais $P(U_r = u_r|U_1 = u_1, \ldots, U_{r-1} = u_{r-1}, U_{r+1} = u_{r+1}, \ldots, U_{n-1} = u_{n-1}, X_{(n)})$, onde $u_j \in \{0, 1\}$ para todo $j = 1, \ldots, n-1$.

Como cada uma dessas distribuições é uma Bernoulli, é suficiente para o processo de simulação determinar-se as razões dadas na seguinte expressão, para cada $r = 1, \ldots, n-1$

$$R_r = \frac{P(U_r = 1|U_1^k = u_1, \ldots, U_{r-1}^k = u_{r-1}, U_{r+1}^{k-1} = u_{r+1}, \ldots, U_{n-1}^{k-1} = u_{n-1}, X_{(n)})}{P(U_r = 0|U_1^k = u_1, \ldots, U_{r-1}^k = u_{r-1}, U_{r+1}^{k-1} = u_{r+1}, \ldots, U_{n-1}^{k-1} = u_{n-1}, X_{(n)})}, \tag{4.42}$$

as quais, pode-se verificar, são dadas por

$$R_r = \frac{c_{xy}^*}{c_{xr}^* c_{ry}^*}, \tag{4.43}$$

onde c_{ij}^* é a coesão *a posteriori* para o bloco X_{ij},

$$x = \begin{cases} \text{Máx}_i\{0 < i < r, U_i^k = 0\} & \text{se } U_i^k = 0 \text{ para algum } i \in \{1, \ldots, r-1\} \\ 0 & \text{caso contrário} \end{cases} \tag{4.44}$$

e o índice y é dado por

$$y = \begin{cases} \text{Min}_i\{r < i < n, U_i^{k-1} = 0\} & \text{se } U_i^{k-1} = 0 \text{ para algum } i \in \{r+1, \ldots, n-1\} \\ n & \text{caso contrário.} \end{cases} \tag{4.45}$$

Sendo assim, o critério para escolha dos vetores $(U_1^k, \ldots, U_{n-1}^k)$ torna-se

$$U_r^k = \begin{cases} 1 & \text{se } \frac{c_{xy}^*}{c_{xr}^* c_{ry}^*} \geq \frac{1-s}{s} \\ 0 & \text{caso contrário,} \end{cases} \tag{4.46}$$

onde s é um número escolhido aleatoriamente no intervalo $(0, 1)$.

Para escolher a amostra de vetores $(U_1, \ldots, U_{n-1})$ a ser utilizada, podemos utilizar algum dos seguintes procedimentos sugeridos na literatura. Podemos gerar várias amostras de vetores e utilizar o último vetor que aparece em cada uma delas para fazer parte da amostra final, ou podemos gerar uma única amostra de vetores e selecionar alguns dos vetores gerados de tal forma que a possível correlação existente entre eles seja eliminada ou ainda, se a correlação existente entre os vetores é pequena, podemos considerar como amostra final a parte da amostra original a partir do ponto em que se verifica a estacionaridade da cadeia gerada.

Com isso fica completamente determinado o procedimento para determinar as distribuições *a posteriori* para a partição aleatória ρ e para o número de blocos pertencentes à partição B.

Na seção 5.5 o leitor encontrará uma aplicação dos resultados que apresentamos neste capítulo a dados do mercado acionário chileno para verificar a ocorrência de pontos de mudanças tanto no retorno diário esperado quanto na volatilidade e também encontrará, no capítulo 5, justificativas para as distribuições *a priori* aqui adotadas dentro da perspectiva da teoria de econômica de Mercados Eficientes.

Capítulo 5

Ilustrando com Dados Reais

5.1 Introdução

Ao longo de todo este trabalho ilustramos através de exemplos simples - e às vezes pouco reais - situações onde o uso do Condicionamento Bayesiano é desprovido de sentido, como o é, às vezes, na incorporação de informações imprevistas para a obtenção da distribuição *a posteriori*. Apresentamos nos capítulos 1 e 2 algumas maneiras alternativas para obter-se a distribuição *a posteriori* em algumas destas situações, destacando dentre estas a Regra de Jeffrey para a qual apresentamos uma versão preditivista (seção 2.3).

Também discutimos no capítulo 4 um procedimento que permite inferir sobre pontos de mudança já ocorridos (em sequências de quantidades aleatórias), os quais muitas vezes tiveram a sua ocorrência determinada pelo acontecimento ou observação de alguma coisa não considerada previamente. A localização destes pontos de mudança são muitas vezes esclarecedoras e informativas tanto sobre o fenômeno em si como sobre as possíveis causas de sua ocorrência, o que torna relevante a busca de procedimentos eficientes para inferir sobre tais pontos.

Devemos ao leitor uma situação real que ilustre melhor a importância dos procedimentos estudados neste trabalho. Este capítulo vem cumprir, pelo menos parcialmente, esta tarefa,

aplicando os procedimentos sugeridos neste trabalho a dados do mercado acionário chileno.

Na seção 5.3 faremos uma breve descrição dos dados que usaremos para análises, os quais correspondem à série formada por preços de negociação das ações da empresa chilena ENDESA (Empresa Nacional de Electricidad S.A.) atingidos no fechamento do mercado em cada dia, no período compreendido entre 1987 e 1994. Apresentaremos alguns gráficos e algumas medidas que permitem compreender o comportamento de tais dados ao longo do tempo bem como uma suscinta descrição dos fatos políticos e econômicos que antecederam às mudanças ocorridas no comportamento destes dados.

Na seção 5.4 apresentamos um algoritmo para fazer previsões de retornos futuros utilizando a Regra de Jeffrey para incorporar informações imprevistas e utilizando-nos de distribuições *a priori* J-conjugadas naturais as quais terão seu uso justificado na próxima seção com argumentos da teoria econômica de mercado eficiente. Compararemos os resultados obtidos por este procedimento com as previsões feitas utilizando apenas o Condicionamento Bayesiano, sem usar qualquer método para incorporar imprevistos. Aplicaremos o algoritmo obtido a dados ENDESA.

Na seção 5.5 aplicaremos os resultados obtidos no capítulo 4 à série ENDESA, fazendo inferências sobre pontos de mudanças tanto no retorno esperado como na volatilidade do mercado. Uma das importâncias deste estudo está em verificar que a volatilidade (medida como variância) no mercado chileno é estocástica como comprovado preliminarmente por Maeda(1996) e Correa(1998)- que aplica a estrutura do MPP supondo que os retornos da série ENDESA são distribuídos segundo uma distribuição t centralizada. Também forneceremos uma estimativa para a distribuição *a posteriori* do número de blocos e para a probabilidade da partição cuja ocorrência é a mais provável.

O modelo t adotado também terá seu uso justificado na próxima seção juntamente com o não uso do Condicionamento Bayesiano, em certas circunstâncias, partindo de argumentos da teoria econômica de mercados eficientes (Ver Fama(1970), Mandelbrot(1963) e Maeda(1996)).

5.2 Mercado Eficiente e o Modelo Adotado

Segundo Fama(1970) o papel primário do mercado de capitais é a alocação ótima dos recursos das reservas de capital e idealmente os preços deveriam assinalar precisamente as melhores fontes para a alocação destes recursos. Um mercado que considera como hipótese que os preços sempre refletem completamente a informação relevante para as decisões sobre tais alocações e que se baseia no fato de que as informações são escassas e caras é dito eficiente.

Basicamente, a teoria de mercados eficientes supõe que uma série de retornos não tem memória, isto é, retornos passados não podem ser usados para prever retornos futuros. Fama (1965) traduz esta hipótese estatisticamente através da seguinte frase:

" *In statistical terms the theory says that successive price changes are independent, identicaly distributed random variables.*"

Segundo ele um dos motivos que provoca esta percepção de " independência" é o grau de conhecimento do investidor e a sua perspicácia em avaliar os efeitos das novas informações sobre o comportamento do mercado.

" *...one of the forces which helps to produce independence of successive price changes may be the existence of sophisticated traders... that has a special talent in detecting dependencies in series of price changes for individual segurities, or that the trader has special talent for predicting the appearence of new information and evaluating its effects on intrinsic values.*" (Fama (1965)).

Essas afirmações nos levam a pensar que a teoria de mercado eficiente parte, não com uma suposição de independência entre os retornos, mas sim por considerar que os investidores percebem alguma semelhança entre estes retornos. Esta semelhança pode ser matematicamente traduzida através de alguma condição de invariância, refletindo apenas que a existência de qualquer relação de dependência entre os retornos é meramente fruto do desconhecimento do investidor sobre a série. Por exemplo, o investidor pode achar que a sequência de retornos é

independente e identicamente distribuída, condicional a alguma característica intríseca que resume efeitos da informação (como são , por exemplo, a volatilidade e o retorno esperado) - o que reflete apenas que, para o investidor, a sequência de retornos segue alguma rota que ele (investidor) não consegue identificar precisamente. Ou seja, uma vez que o retorno total é conhecido, qualquer caminho seguido pelos retornos que tenham este mesmo total tem, aos olhos do investidor, igual chance ocorrência. Como vimos no capítulo 3, o que ocorre aqui é que o investidor julga que a sequência de retornos é invariante sob transformações ortogonais que preservam vetores de uns e, uma vez que este julgamento é feito, o teorema 3.2.2 diz que

$$P(X_1 \leq x_1, \ldots X_n \leq x_n) = \int \Pi_{i=1}^{n} N(\mu, \sigma^2) dQ(\mu, \sigma^2) \tag{5.1}$$

para qualquer $n \geq 1$, onde μ e σ^2 podem ser interpretados como o retorno médio esperado e a volatilidade do mercado, respectivamente.

Uma segunda consideração importante da teoria de mercado eficiente foi estabelecida por Mandelbrot(1963) o qual, além de encontrar evidências de não-normalidade dos retornos, obtém a seguinte relação entre estes, fornecendo indicativos de modificações na volatilidade no decurso do tempo.

" Large daily price changes tend to be followed by large daily changes and small daily price changes tend to be followed by small daily changes."

Essa suposição sugere uma relação de linearidade entre o retorno a ser obtido futuramente e o retorno obtido no instante atual, o que justifica as condições de linearidade nas esperanças condicionais feitas na proposição 3.3.1 - expressões (3.47) e (3.48). Mais ainda, seguidores desta teoria econômica acreditam que toda mudança intrínseca nos valores de uma certa ação são provocadas por alguma informação não programada ou pela ocorrência de eventos imprevistos - ou seja, pela ocorrência de algo não previamente considerado pelo investidor. A "informação pública" (por exemplo, taxa de inflação, índices industriais, etc) incorpora-se imediatamente aos preços, enquanto que as mudanças que ocorrem na volatilidade são provocadas por "informações privadas" (informações não antecipadas) de eventos significativos. Uma boa explicação sobre este assunto pode ser encontrada em Maeda(1996) o qual afirma

" *La variación de la volatilidad transada para um activo* **estará explicada solamente por lo desconocido o lo no antecipado**[1] *y por lo tanto la volatilidad estará asociada ... a la llegada de nuevas notícias (notícias no antecipadas).*"

Desta forma, sob as suposições da teoria de mercado eficiente fica justificado o uso de um modelo t para descrever o comportamento dos retornos diários em situações onde as únicas informações recebidas são aquelas já previamente consideradas (ver proposição 3.3.1 para os detalhes matemáticos) e além disto justifica-se o uso da priori conjugada natural dada na expressão (3.50) para a volatilidade do mercado σ^2 e para o retorno médio diário μ.

Uma vez que as mudanças intrísecas no comportamento dos preços são provocadas pela ocorrência de informações não previamente consideradas, o uso do Condicionamento Bayesiano para fazer previsões pode ser inapropriado (ver cap.1) sugerindo-nos o uso de algum procedimento alternativo para este fim. Como as hipóteses da teoria de mercados eficientes descritas acima tendem a se manterem válidas mesmo depois do recebimento de alguma notícia inesperada - o que é exigido pela Regra Preditivista de Jeffrey (ver proposição 2.3.1, por exemplo) - e como segundo Fama(1965) a ocorrência de eventos imprevistos provocam mudanças na opinião do investidor apenas sobre alguns indicadores do mercado (como volatilidade, p.ex.), justifica-se usarmos a Regra de Jeffrey para fazer previsões de retornos futuros quando imprevistos ocorrem . Este fato também justifica o uso de distribuições *a priori* J-conjugadas naturais (seções 2.4 e 2.5) no algoritmo de previsão que apresentaremos na seção 5.4.

5.3 Os Dados

Para nossas considerações ao longo deste capítulo, as quantidades aleatórias X_i e D_i denotarão, respectivamente, o retorno (ou rentabilidade) e o preço diferenciado alcançado por uma certa ação no dia i, para $i \geq 1$, os quais são definidos da seguinte forma

$$D_i = P_i - P_{i-1} \tag{5.2}$$

[1]grifo do autor

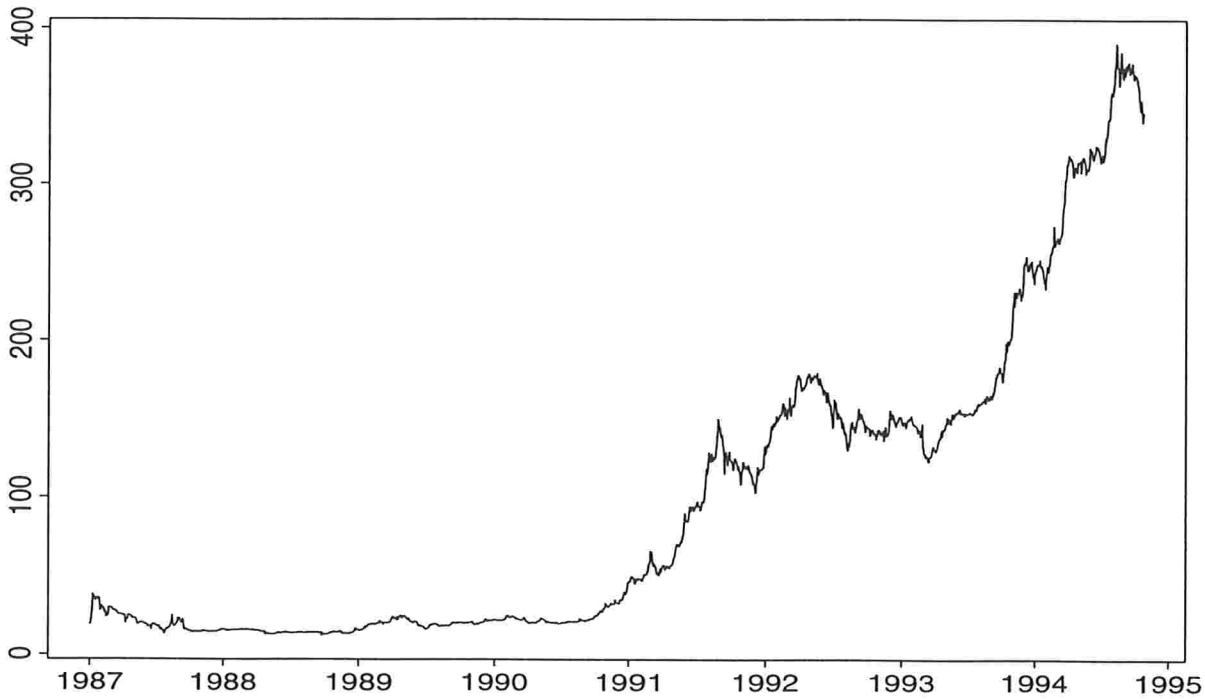
$$X_i = \frac{D_i}{P_{i-1}} \tag{5.3}$$

onde P_i é o preço atingido pela ação de interesse no fechamento do mercado no dia i.

Nosso principal objetivo nesta seção é apresentar uma breve descrição dos dados da série ENDESA, explorando o comportamento das séries ENDESA diferenciada e Rentabilidade ENDESA que são derivadas da série original usando as expressões (5.2) e (5.3), respectivamente. As séries Rentabilidade ENDESA mensal e ENDESA Diferenciada mensal, que também serão consideradas, são originadas da série de preços ENDESA da mesma forma tomando-se o preço atingido pela ação no fechamento do mercado no último dia de cada mês.

O gráfico apresentado na figura 5.1 apresenta os dados da série "ENDESA" formada pelos preços diários atingido pela venda da ação da empresa chilena ENDESA no fechamento do mercado, desde janeiro de 1987 até dezembro de 1994.

Figura 5.1: Preços Diários no Fechamento do Mercado - ENDESA(1987/1994)

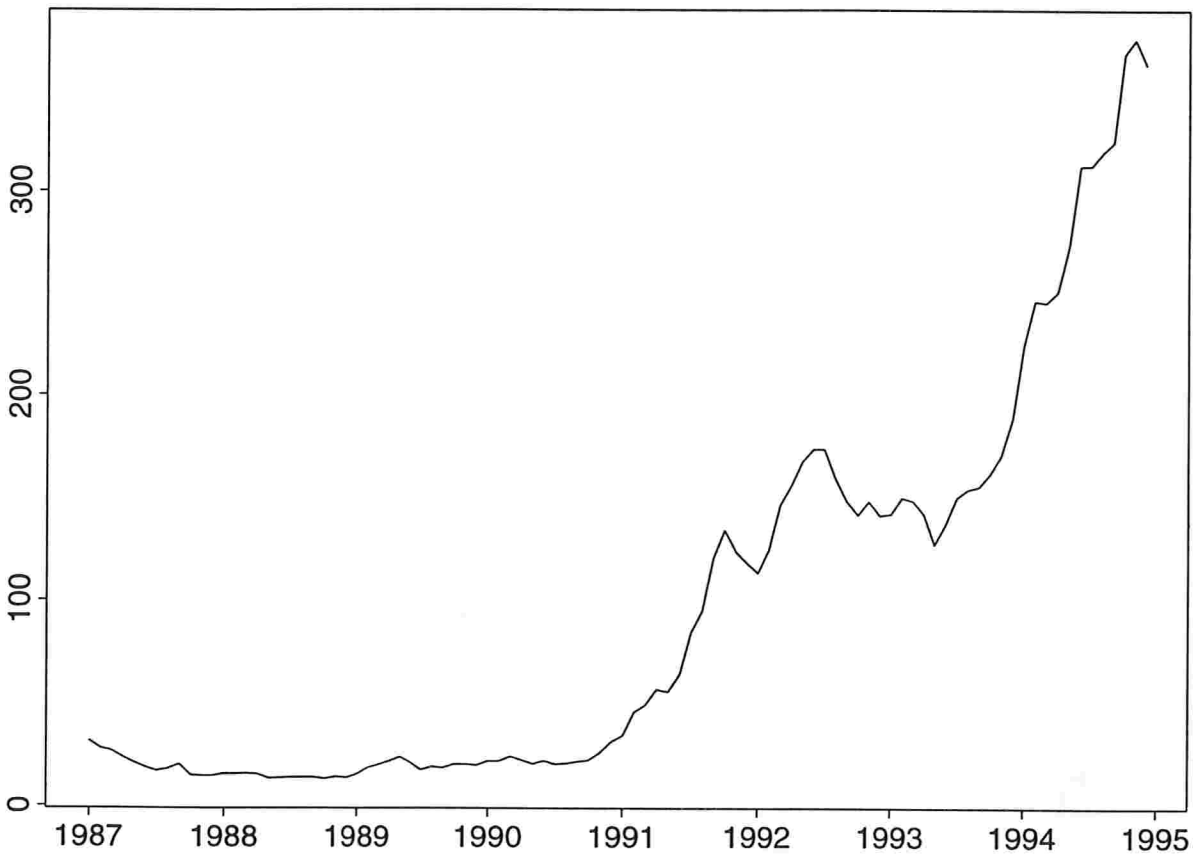


Perceba que série ENDESA apresenta um comportamento bastante estável entre 1987

e o final de 1990, com pequenas oscilações em torno de um valor médio próximo de zero e mudanças bastante acentuadas tanto no preço médio quando na variabilidade no final de 1990 e no final de 1993. Nota-se também que a mudança ocorrida no final de 1993 parece mais acentuada do que aquela ocorrida no final de 1990 e também observa-se que entre estas duas mudanças acentuadas ocorre um período de estabilização dos preços, entre aproximadamente a metade de 1992 e o final do terceiro trimestre de 1993, havendo pouca variabilidade em torno de um valor médio próximo de 150 pesos.

Nos dois próximos gráficos pode-se observar, respectivamente, o comportamento da média e variância mensais calculadas utilizando-se os dados da série ENDESA. Pode-se perceber que há um crescimento quase exponencial da média ao longo do período observado.

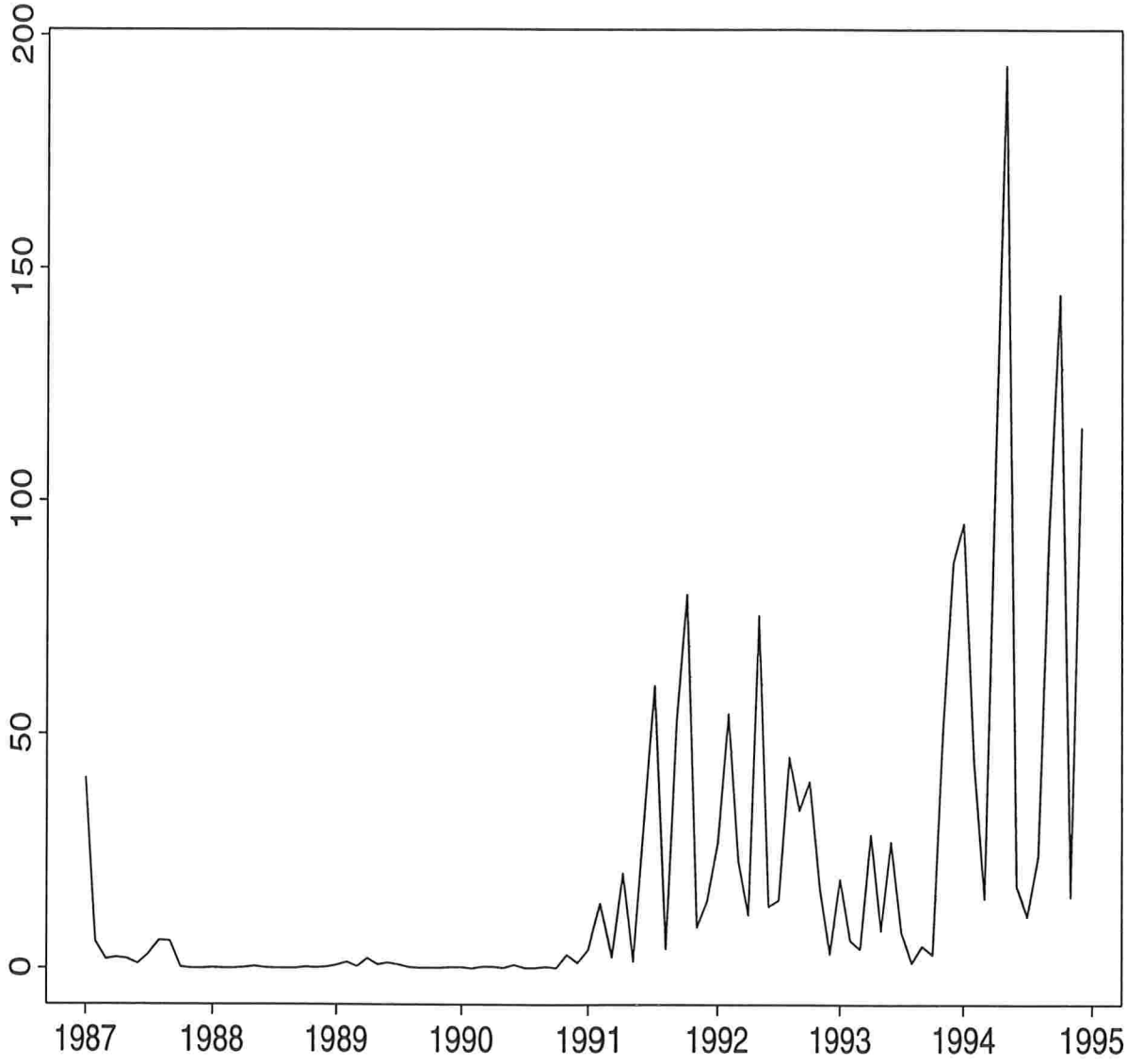
Figura 5.2: Preços Médios Mensais - ENDESA(1987/1994)



Para a variância nota-se um comportamento bem estável até o final de 1990 e entre o

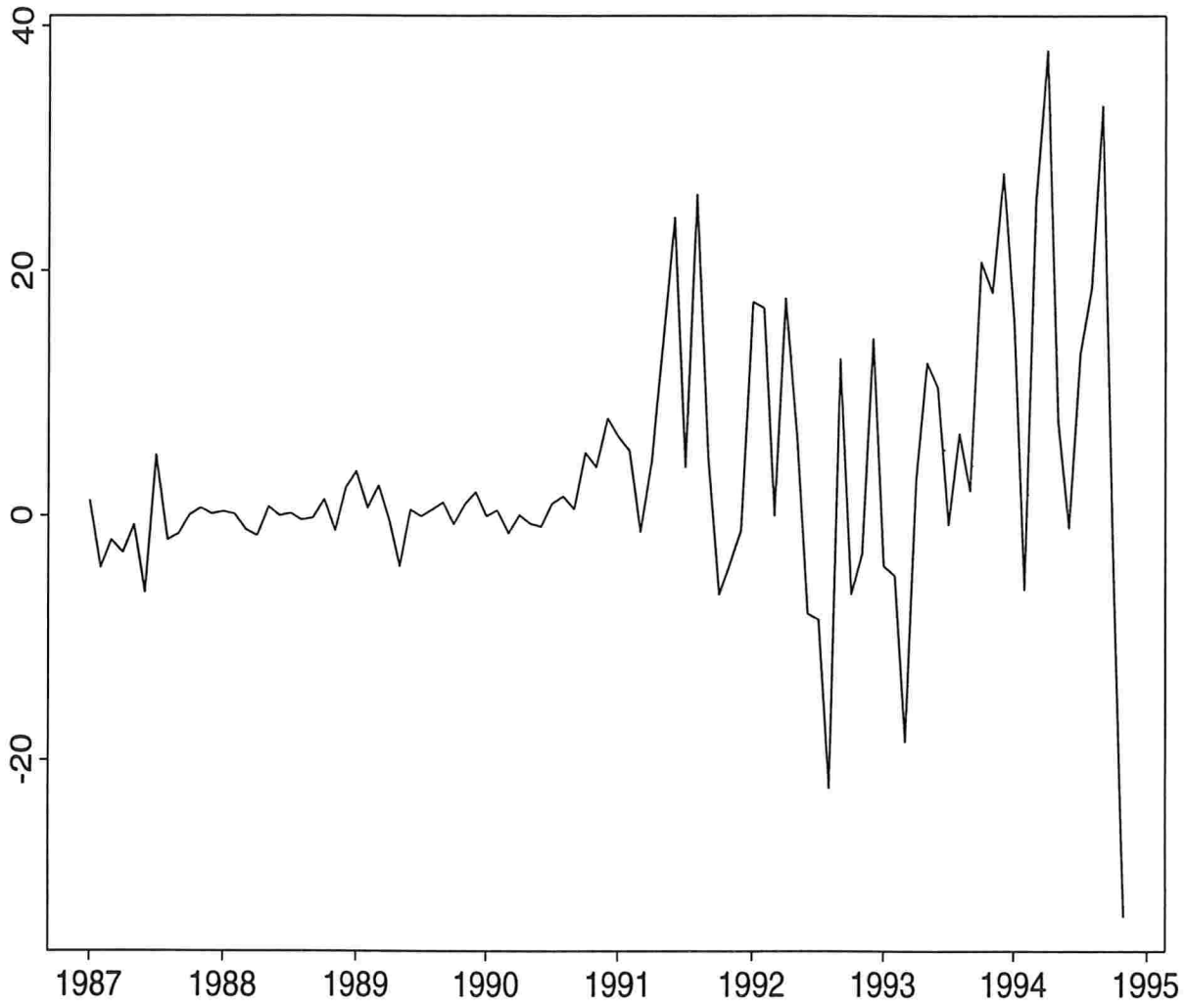
final de 1992 e o final de 1993. Em ambos períodos a variância mensal oscila em torno de valores pequenos. Nos outros dois períodos ocorre uma oscilação maior no valor mensal da variância, sendo esta mais intensa a partir do fim de 1993.

Figura 5.3: Variância Mensal - ENDESA(1987/1994)



O gráfico na figura 5.4 mostra a série contendo os Preços Diferenciados Mensais.

Figura 5.4: ENDESA Diferenciada Mensal(1987/1994)



Como antes, o comportamento dos preços diferenciados é bem estável entre 1987 e 1990, apresentando pouca variabilidade em torno de uma média próxima de zero. Também é possível notar que ocorre maior oscilação dos preços diferenciados a partir do final de 1990 e esta oscilação é ainda mais acentuada a partir do último trimestre de 1993. O preço diferenciado médio também têm seu comportamento alterado a partir do final de 1991, ocorrendo maior oscilação em seus valores a partir deste período e novas mudanças na metade de 1992 e no final de 1993. Estes períodos também marcam mudanças na variabilidade,

ocorrendo um nítido aumento na variância depois do final de 1991.

O comportamento descrito acima pode ser melhor percebido quando observamos os seguintes gráficos que mostram o comportamento mensal do preço diferenciado médio e de sua variabilidade.

Figura 5.5: Média Mensal - ENDESA Diferenciada (1987/1994)

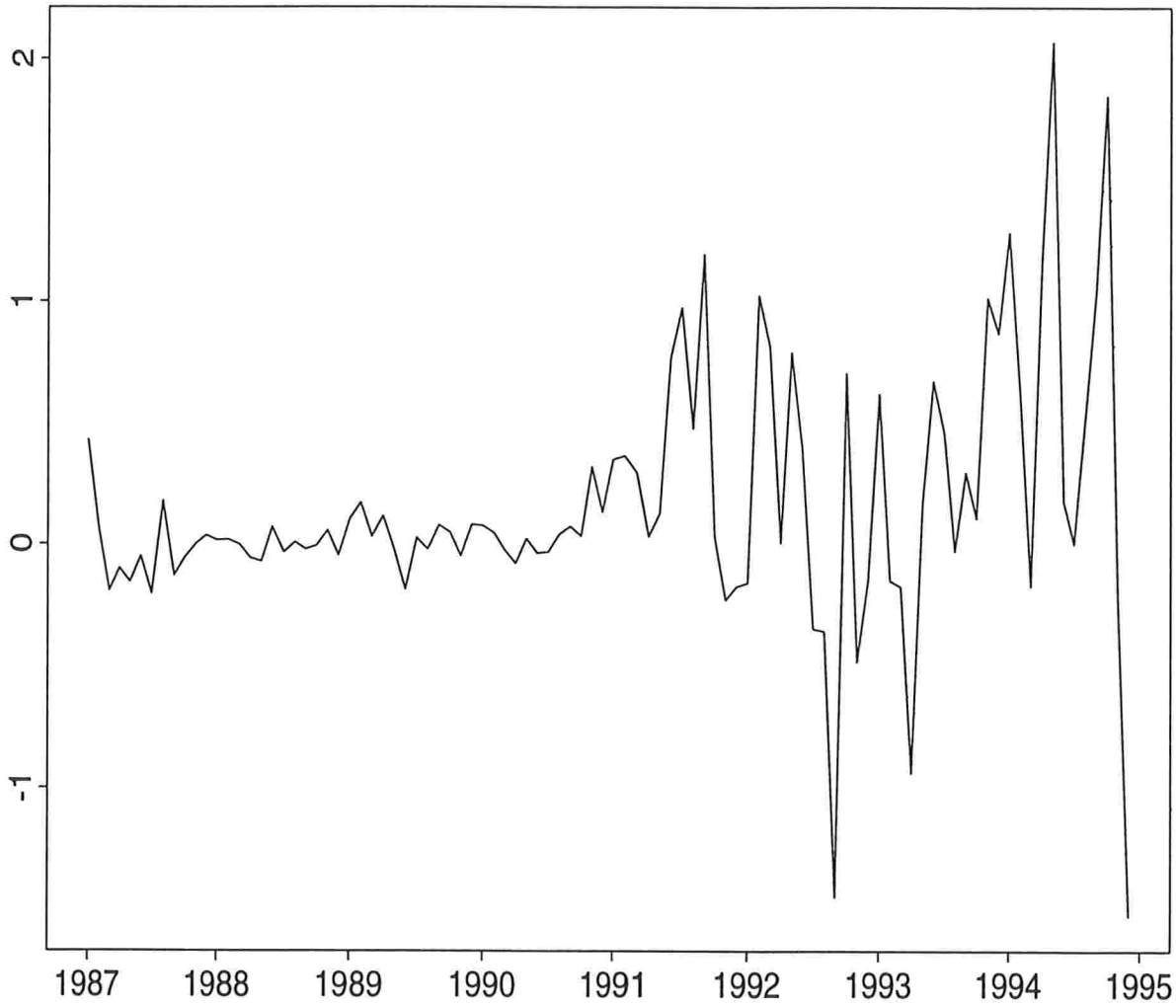
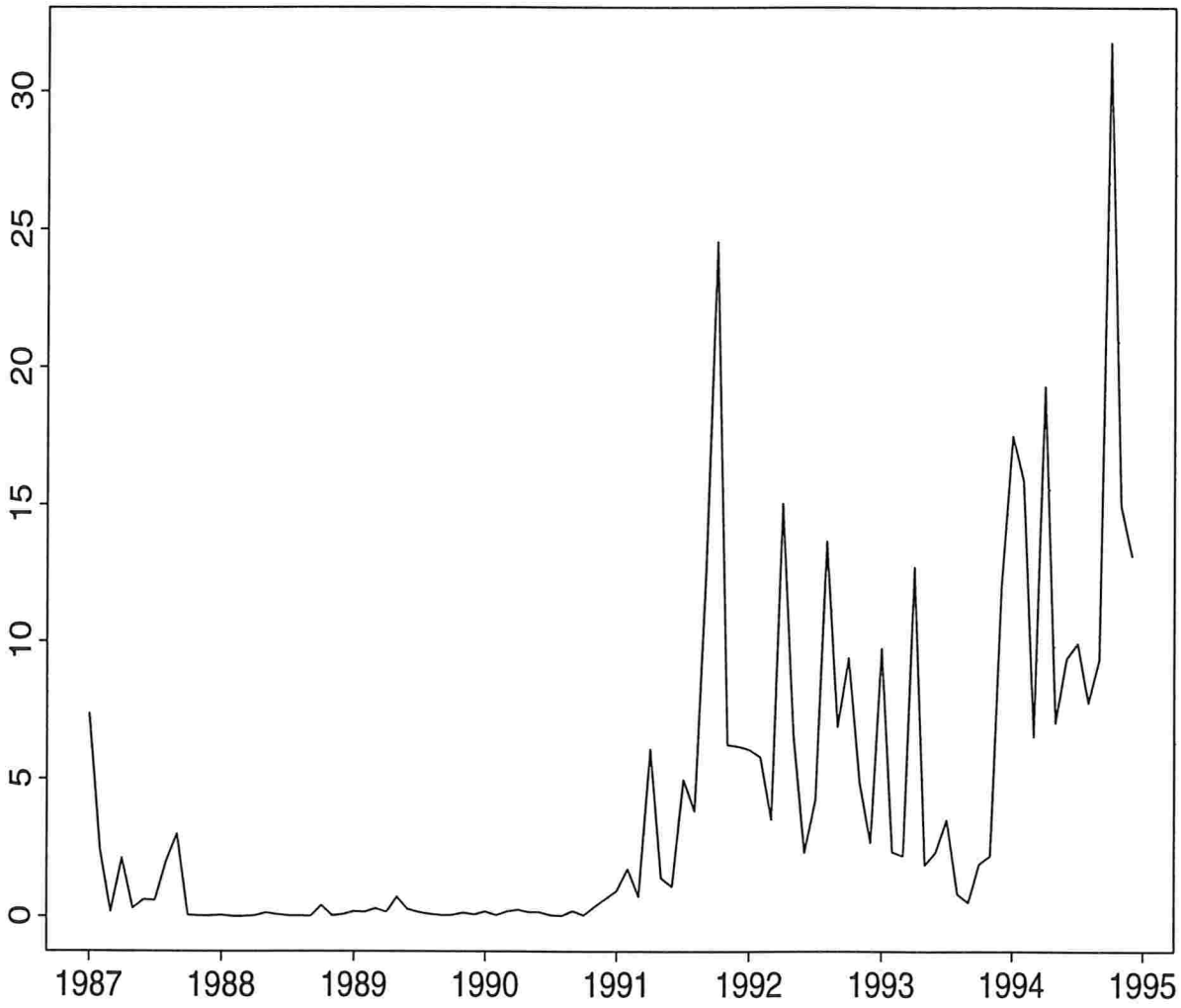


Figura 5.6: Variância Mensal - ENDESA Diferenciada (1987/1994)



Perceba que a tendência de crescimento exponencial antes observada para o comportamento das médias mensais não é mais observada aqui e que o comportamento para a variância e muito similar àquele apresentado pela variância mensal dos preços diários.

Observe também os seguintes histogramas exibindo o comportamento da série diferenciada nos períodos mencionados acima e também o seu comportamento ano–a–ano.

Figura 5.7: Histogramas para os Períodos em que Parecem Haver Mudanças - ENDESA Diferenciada Diária

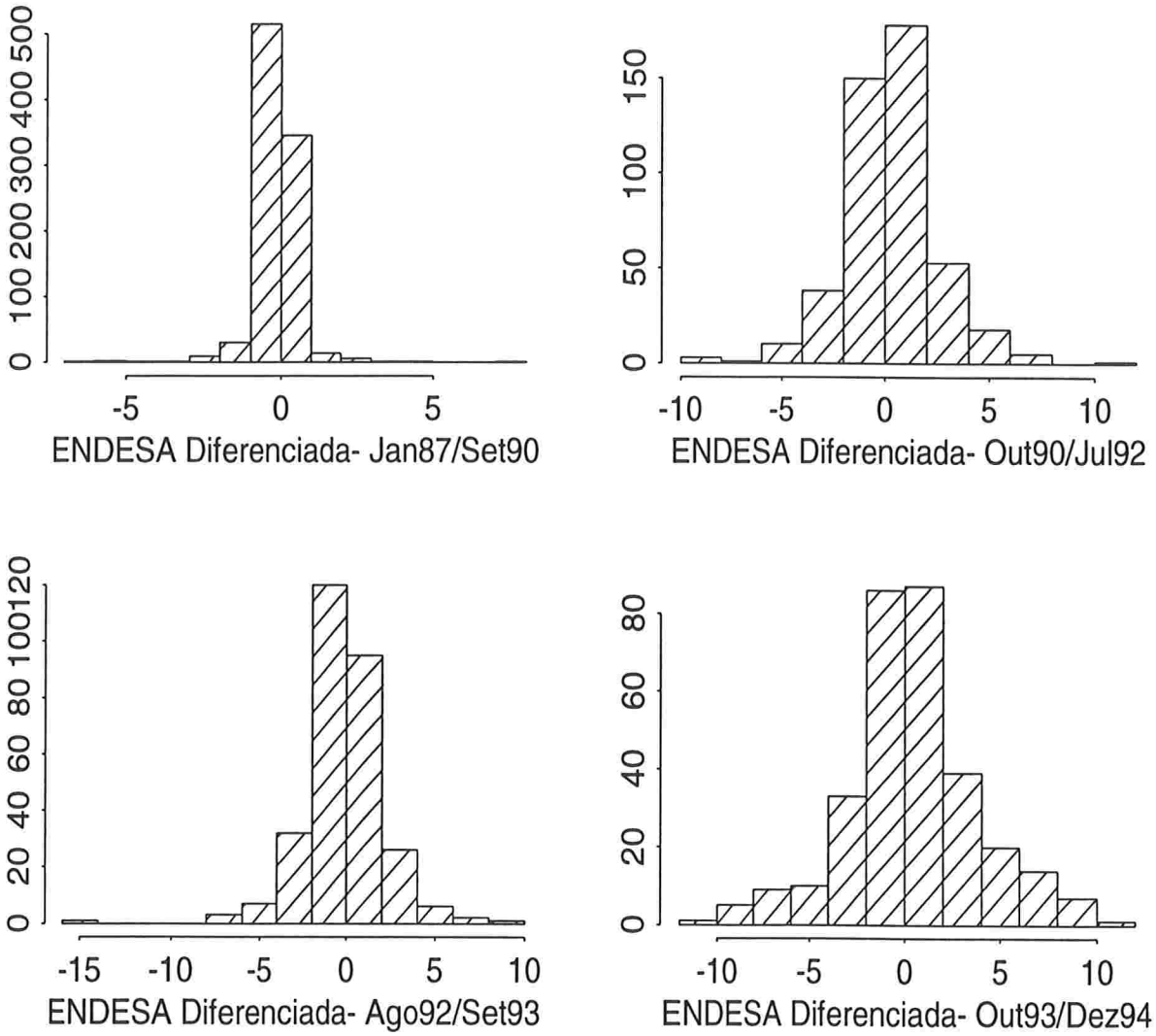
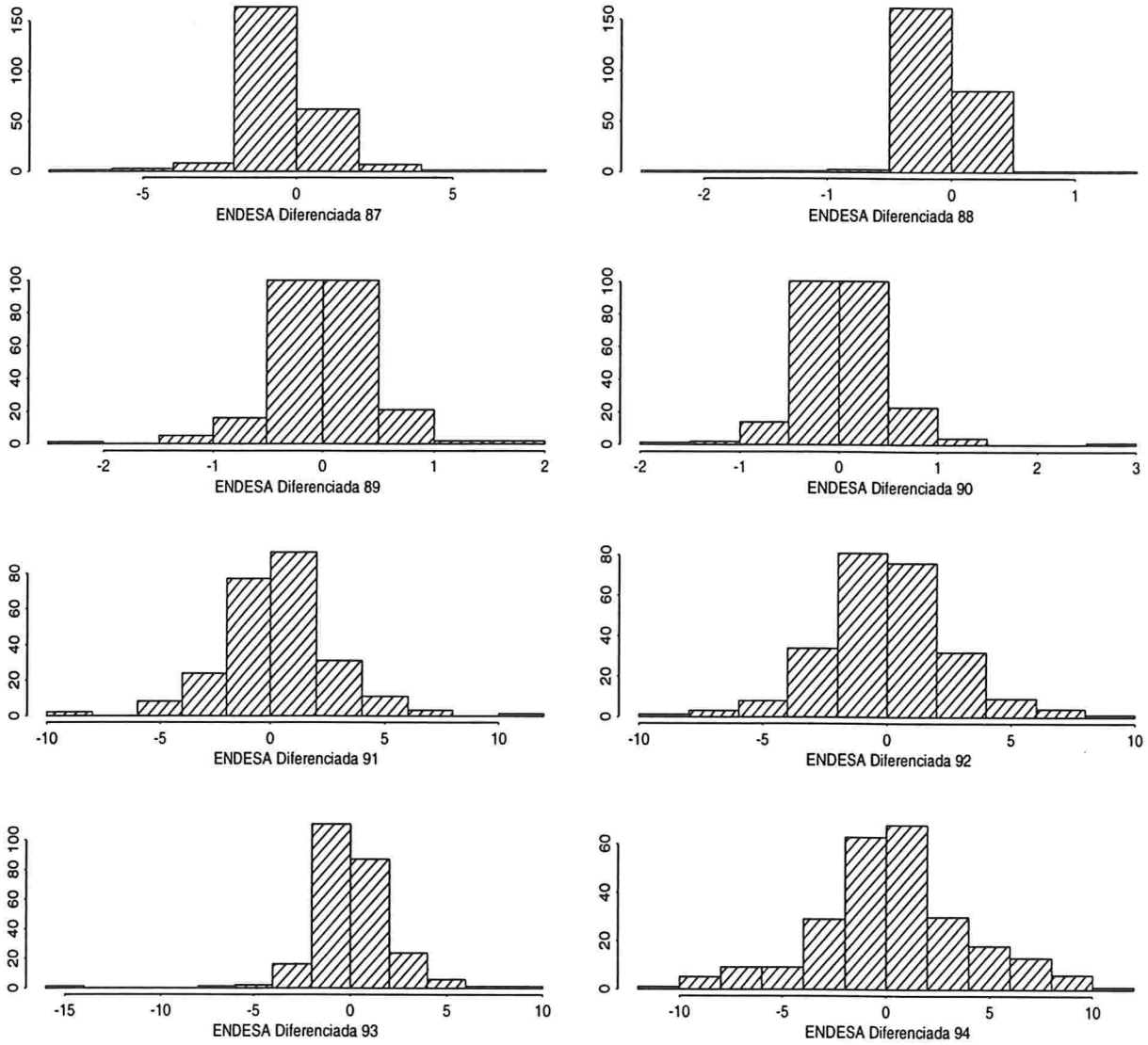


Figura 5.8: Histogramas ano–a–ano - ENDESA Diferenciada Diária



As tabelas 5.1 e 5.2 apresentam preços diferenciados médios e a variabilidade calculadas a partir dos preços diferenciados diários nos períodos que pareciam haver mudanças no comportamento da série e dentro de cada ano, respectivamente. Pode-se notar que tanto os histogramas apresentados anteriormente quanto os dados apresentados abaixo confirmam a análise anteriormente feita.

Tabela 5.1: Média e Variância nos Períodos em que Parecem Haver Mudanças - ENDESA Diferenciada Diária

PERÍODO	MÉDIA	VARIÂNCIA
Jan 87 - Set 90	0.0033	0.4959
Out 90 - Jul 92	0.3194	5.1983
Ago 92 - Set 93	−0.0307	5.2712
Out 93 - Dez 94	0.5913	12.0945

Tabela 5.2: Média e Variância ano-a-ano - ENDESA Diferenciada Diária

ANO	MÉDIA	VARIÂNCIA
1987	$-0,0158$	1.5114
1988	−0.0060	0.0650
1989	0.0273	0.1917
1990	0.0468	0.1878
1991	0.3453	5.8827
1992	0.0884	6.7612
1993	0.2460	4.4481
1994	0.5635	13.6851

Comparando estes resultados com aqueles apresentados nas tabelas 5.3 e 5.4 abaixo as quais exibem os preços diferenciados médios e a variabilidade partindo de dados da série ENDESA diferenciada Mensal, percebemos que há um comportamento similar destas duas variáveis.

Tabela 5.3: Média e Variância nos Períodos em que Parecem Haver Mudanças - ENDESA Diferenciada Mensal

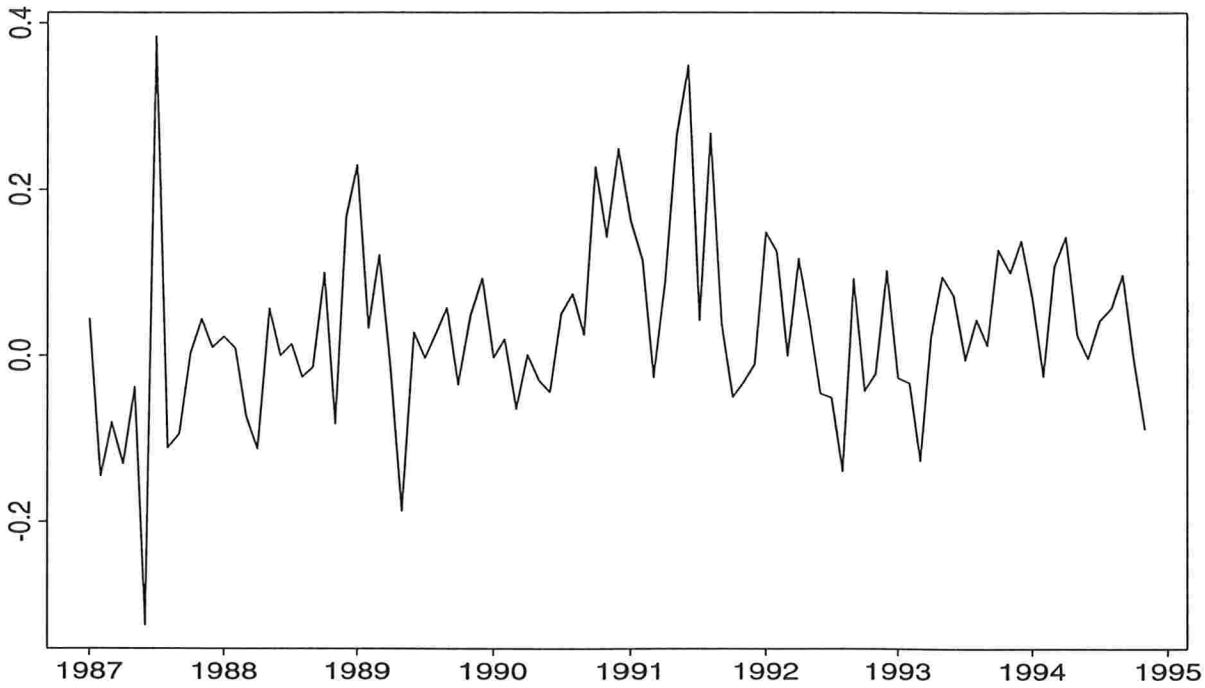
PERÍODO	MÉDIA	VARIÂNCIA
Jan 87 - Set 90	−0.1153	3.8169
Out 90 - Jul 92	6.2359	96.7396
Ago 92 - Set 93	0.1429	125.4123
Out 93 - Dez 94	12.7143	343.9698

Tabela 5.4: Média e Variância ano-a-ano - ENDESA Diferenciada Mensal

ANO	MÉDIA	VARIÂNCIA
1987	−1.0542	8.2374
1988	0.0417	1.2187
1989	0.5333	3.6841
1990	1.4792	7.9920
1991	6.4583	107.8499
1992	3.1667	172.2588
1993	6.1250	164.1939
1994	10.0909	406.7034

Para a série Rentabilidade mensal, exibida na figura 5.9, também não se observa mais a tendência de crescimento exponencial, assim como tínhamos para os preços diferenciados.

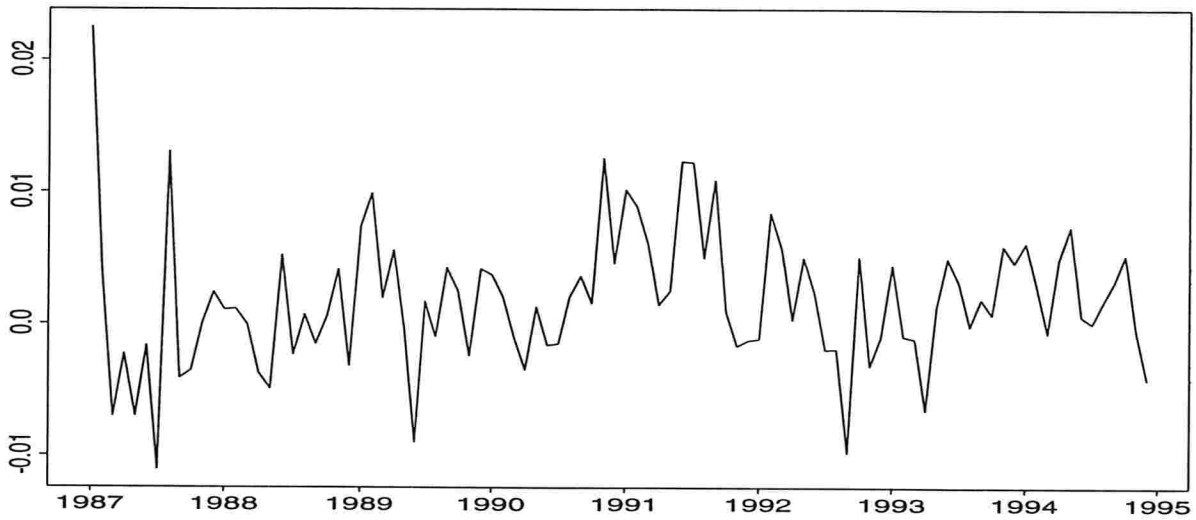
Figura 5.9: Rentabilidade Mensal -ENDESA(1987/1994)



Aqui também parece razoável considerar mudanças nos períodos anteriormente citados, só que neste caso percebe-se uma tendência de decrescimento da variância, se considerados estes períodos. O retorno esperado parece ter sua maior alta entre final de 90 e metade de 92, onde parece também haver volatilidade alta. Se o interesse maior está na detecção de pontos de mudanças mais do que em fazer previsões, o final de 87 também parece marcar uma mudança tanto na variância quanto no retorno médio.

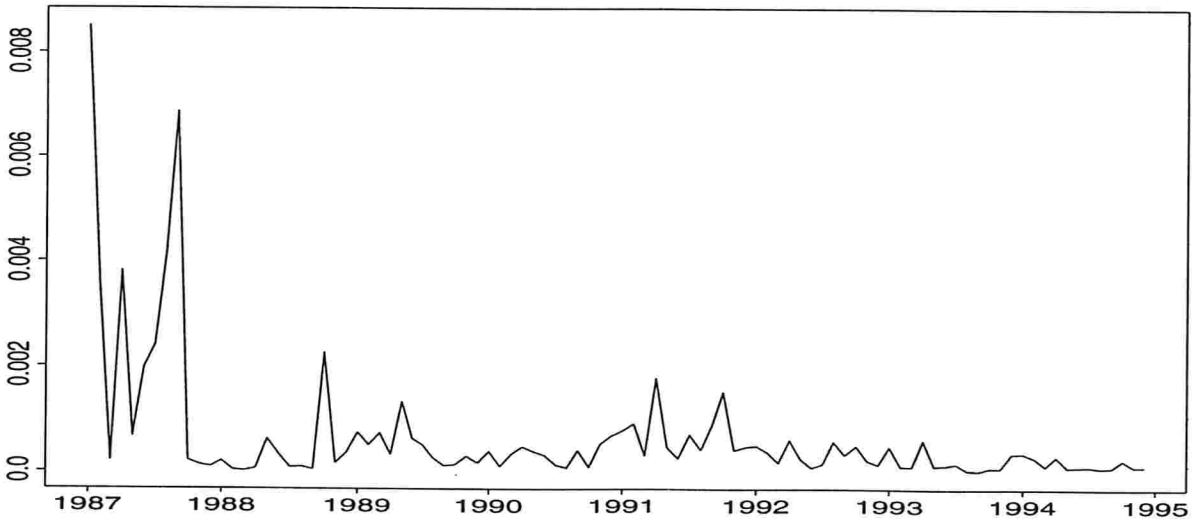
O comportamento do retorno médio mensal pode ser observado no gráfico que segue.

Figura 5.10: Rentabilidade Média Mensal -ENDESA(1987/1994)



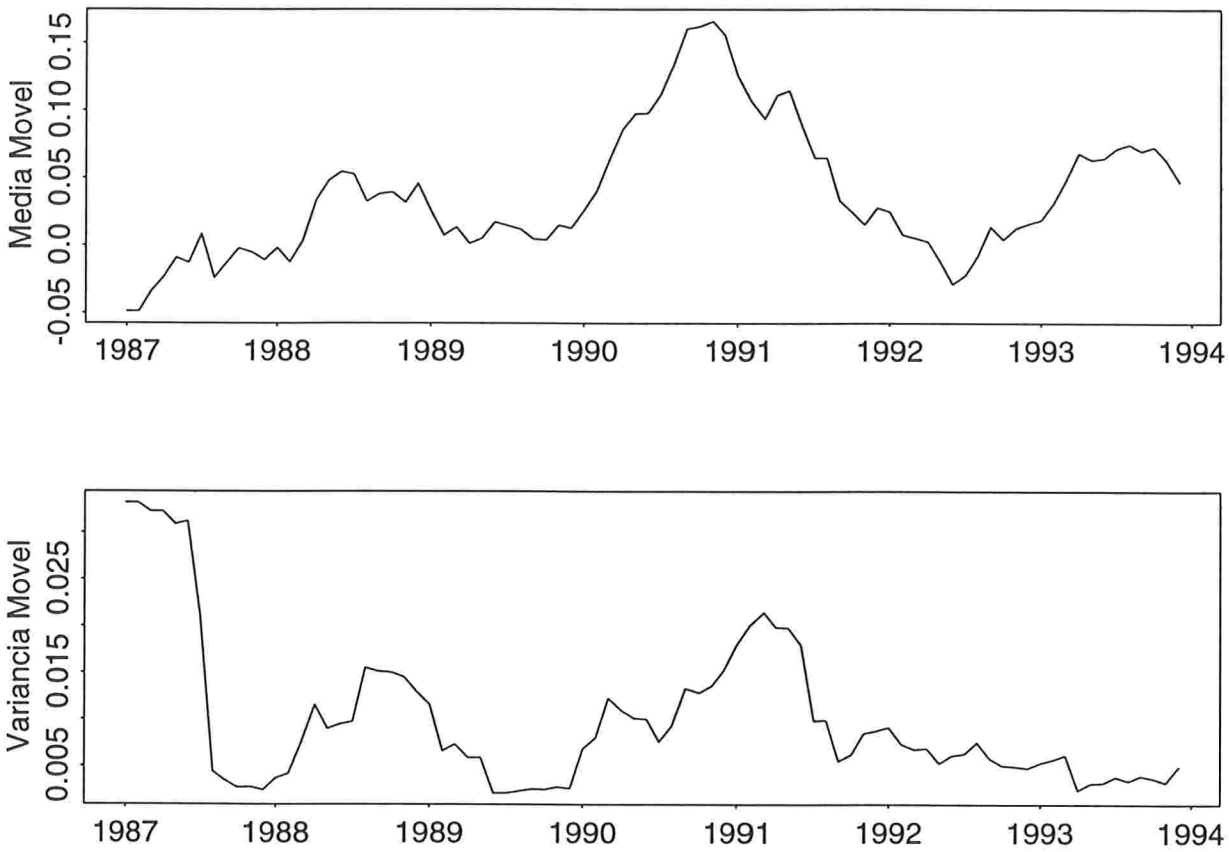
O próximo gráfico mostra o comportamento ao longo do tempo da variabilidade da série rentabilidade ENDESA.

Figura 5.11: Variância Mensal da Rentabilidade -ENDESA(1987/1994)



O comportamento anteriormente descrito pode ser melhor percebido na próxima figura, onde consideramos as médias e variâncias móveis de ordem dez.

Figura 5.12: Médias e Variâncias Móveis de ordem 10 -Rentabilidade ENDESA(1987/1994)



Perceba que, do ponto-de-vista descritivo, é evidente um pequeno número de mudanças ao longo do tempo, tanto no retorno esperado quanto na volatilidade.

Analogamente ao estudo feito para os preços diferenciados exibiremos a seguir tabelas e histogramas mostrando o comportamento dos Retornos mensais e diários nos períodos de interesse e ano–a–ano.

Tabela 5.5: Média e Variância nos Períodos em que Parecem Haver Mudanças - Rentabilidade ENDESA Diária

PERÍODO	MÉDIA	VARIÂNCIA
Jan 87 - Set 90	0.0007	0.00098
Out 90 - Jul 92	0.0047	0.00056
Ago 92 - Set 93	$-6.34\text{x}10^{-5}$	0.00025
Out 93 - Dez 94	0.0025	0.00016

Tabela 5.6: Média e Variância ano–a–ano - Rentabilidade ENDESA Diária

ANO	MÉDIA	VARIÂNCIA
1987	0.0004	0.0027
1988	−0.0002	0.0003
1989	0.0019	0.0005
1990	0.0020	0.0003
1991	0.0056	0.0008
1992	0.0009	0.0003
1993	0.0016	0.0002
1994	0.0022	0.0002

Tabela 5.7: Média e Variância nos Períodos em que Parecem Haver Mudanças - Rentabilidade ENDESA Mensal

PERÍODO	MÉDIA	VARIÂNCIA
Jan 87 - Set 90	0.0013	0.0120
Out 90 - Jul 92	0.0986	0.0142
Ago 92 - Set 93	0.0036	0.0058
Out 93 - Dez 94	0.0565	0.0048

Tabela 5.8: Média e Variância ano-a-ano - Rentabilidade Mensal- ENDESA

ANO	MÉDIA	VARIÂNCIA
1987	-0.0363	0.0281
1988	0.0054	0.0061
1989	0.0332	0.0099
1990	0.0544	0.0105
1991	0.1011	0.0179
1992	0.0273	0.0082
1993	0.0355	0.0059
1994	0.0386	0.0045

Figura 5.13: Histogramas nos períodos em que parecem haver mudanças - Rentabilidade ENDESA Diária

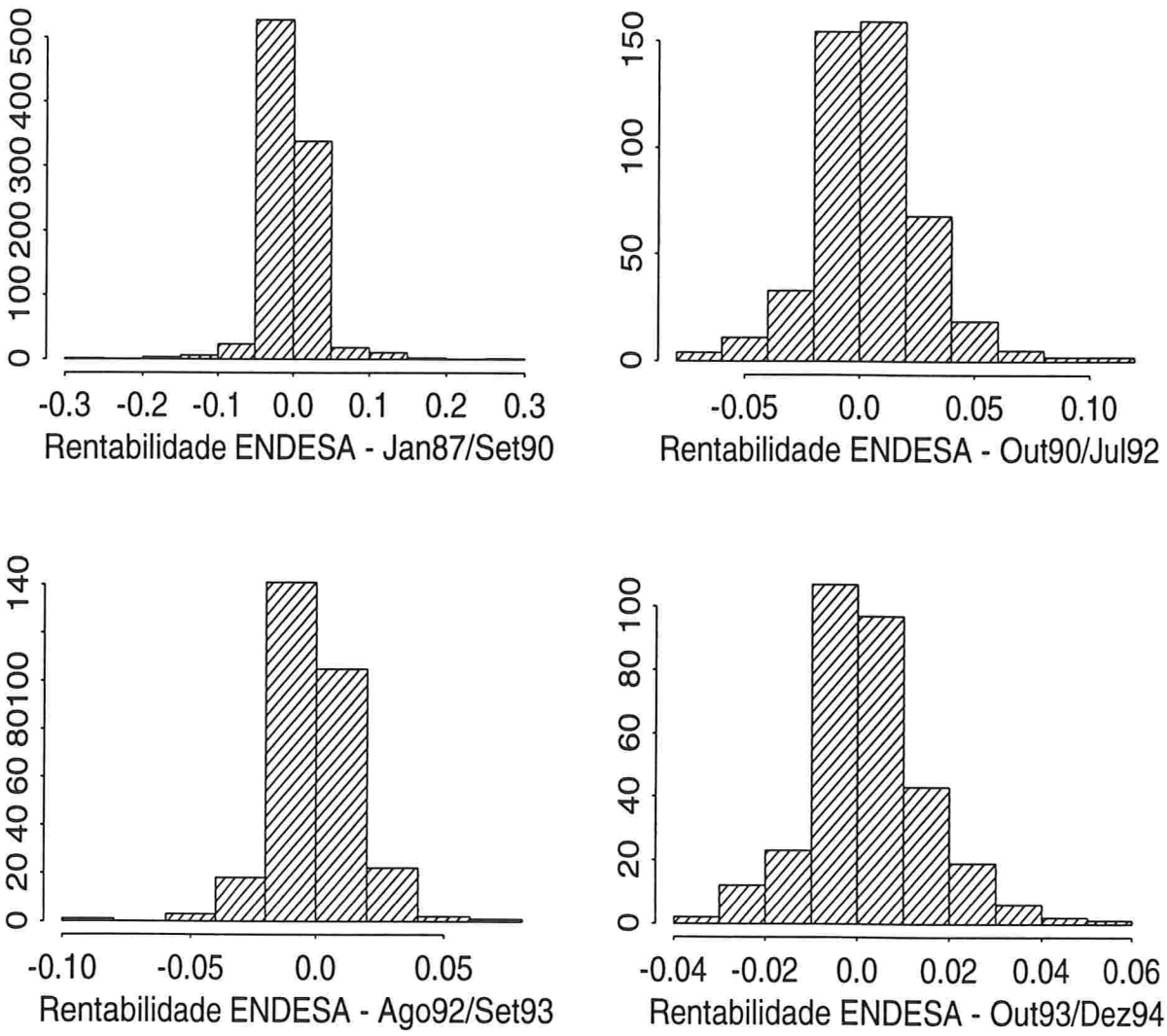
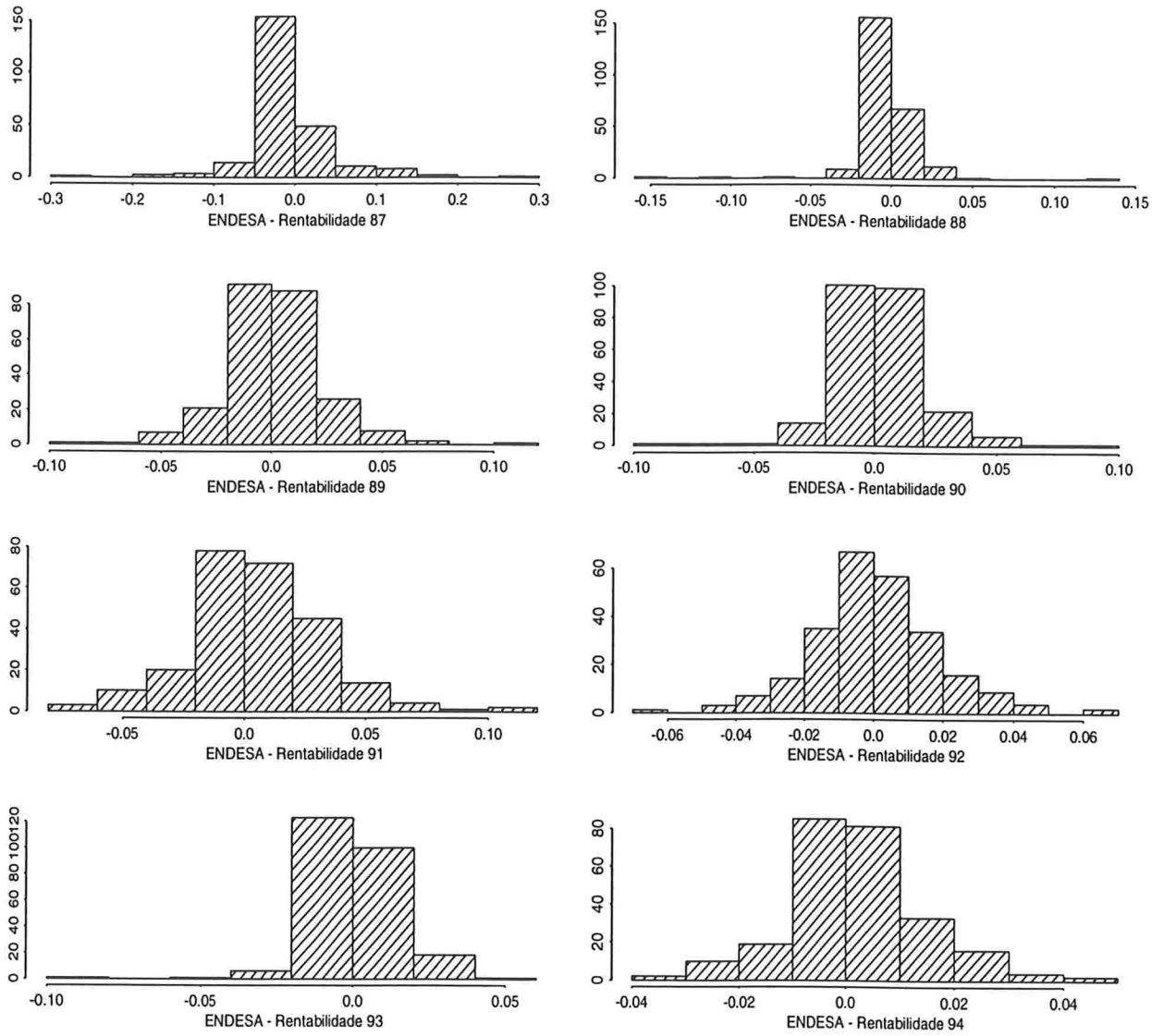


Figura 5.14: Histogramas ano–a–ano - Rentabilidade ENDESA Diária



Maeda(1996) fornece o seguinte panorama vivido pelo Chile no período estudado, apontando alguns eventos possíveis causadores da instabilidade observada a partir de 1991 no mercado chileno. Para Maeda(1996) a instabilidade observada a partir de 1991 pelo mercado chileno deve-se não só às mudanças ocorridas na política interna chilena como também a alguns acontecimentos internacionais importantes. O ano de 1988 foi um marco importante na História do Chile, pois realizou-se em 5 de outubro o plebiscito para saber sobre a continuidade ou não do governo militar. Neste dia, registrou-se uma queda de 13 por cento na rentabilidade da bolsa. Também em junho deste ano é implementado o sistema de

telepregão na bolsa de valores e ocorre uma queda nos impostos sobre combustíveis devido a baixa mundial sofrida pelo preço do petróleo. Em 1989, ocorre a proíbição da entrada de algumas espécies de frutas chilenas no mercado norte-americano, provocando um ajuste monetário pelas autoridades da área econômica. Em dezembro deste mesmo ano, realiza-se a eleição para presidente. A guerra no Golfo Pérsico, iniciada em 1990, provoca neste ano uma alta no preço do petróleo e também é responsável pela baixa generalizada deste preço a partir de 1991. Internamente, o Chile passou por um período de ajuste em 90, quando foram aprovadas as reformas tributária e trabalhista e amplia-se, nesta época, na Bolsa Acionária, o sistema de telepregão permitindo negociações também no período da tarde. Neste ano começa a gestão do primeiro presidente eleito (Patricio Alwyn) após um longo período de ditadura militar. Em 91, ocorre o aumento do volume negociado na Bolsa Acionária e a Bolsa Eletrônica de Santiago começa a operar provocando uma competição entre as duas Bolsas. O governo de Alwyn estrutura-se. Em 92, há uma baixa nas taxas de interesse do Banco Central provocada pela baixa mundial do dólar e a bolsa amplia suas possibilidades de negócios com o ingresso de 26 novas ações. Em 93 a recessão nos países desenvolvidos provoca uma baixa nas exportações chilenas causando desequilíbrio na balança comercial. Os efeitos deste desequilíbrio são pelo menos parcialmente compensados pelo aumento do dinamismo do mercado acionário. Finalmente em 94, destaca-se a queda da inflação e o início das negociações com o NAFTA e com o MERCOSUL.

Na próxima seção apresentamos um algoritmo para fazer previsões de retornos futuros utilizando a Regra Preditivista de Jeffrey para incorporar eventos imprevistos e partindo do pressuposto que o mercado acionário chileno satisfaz as exigências da Teoria de Mercado Eficiente. O algoritmo obtido será aplicado à série Rentabilidade Mensal - ENDESA.

5.4 A Regra de Jeffrey e o Mercado Chileno

Sustentados pelos argumentos fornecidos nas seções 5.2 e 5.3 desenvolveremos um algoritmo para previsões de retornos futuros com base em toda a informação disponível (mesmo aquelas não previamente antecipadas) adotando o Condicionamento Bayesiano usual para

incorporar informações previamente consideradas e usando a Regra Preditivista de Jeffrey para incorporar aquelas informações que modificam em algum sentido a sua opinião sobre o comportamento dos dados. Estes argumentos também justificam o uso do modelo t dado na expressão (3.49) para descrever o comportamento dos retornos nos períodos em que o mercado recebe apenas as informações previamente antecipadas.

Para derivar a distribuição preditiva para esse processo de previsão considere o resultado apresentado no lema 3.3.1 estabelecido por Arellano-Valle e Bolfarine (1994). Usando este resultado facilmente derivamos a prova da seguinte proposição a qual fornece o mecanismo de previsões de retornos futuros quando estamos em situações em que o mercado financeiro está estável.

Proposição 5.4.1 *Seja $X_1, X_2, \ldots$ uma sequência de q.a reais, $O(\mathbf{1})$-invariantes. Se*

$$\begin{aligned} E(X_3|X_1,X_2) &= a(X_1+X_2)+b_1 \qquad \text{e} \\ E(X_3^2|X_1,X_2) &= a(X_1^2+X_2^2)+b_2 \end{aligned} \tag{5.4}$$

onde $a \in (0,1/2)$, $b_1 \in \mathcal{R}$ e $b_2 \in \mathcal{R}^+$ de tal forma que $b_2 > \frac{b_1^2}{1-2a}$, então considerando $X_{(k)} = (X_1,\ldots,X_k)^t$ e $X_{(n-k)} = (X_{k+1},\ldots,X_n)$, $k \geq 1$, temos:

$$X_{(n-k)}|X_{(k)} = \mathbf{x}_k \sim t_{n-k}(\mu_{n-k}(\mathbf{x}_k); \Sigma_{n-k\cdot k}; \lambda_{q(\mathbf{x}_k)}, \nu_{n-k}) \tag{5.5}$$

onde,

$\mu_{n-k}(\mathbf{x}_k) = \{(a\sum_{i=1}^k x_i + b_1)/(1+a(k-2))\}\mathbf{1}_{n-k\times 1}$;

$\Sigma_{n-k\cdot k} = I_{n-k} + \frac{a}{1+a(k-2)}\mathbf{1}_{n-k\times n-k}$;

$q(x_k) = \sum_{i=1}^k x_i^2 + \frac{b_1^2 k}{(1-2a)(1+a(k-2))} - \frac{a}{1+a(k-2)}(\sum_{i=1}^k x_i)^2 - \frac{2b_1}{1+a(k-2)}\sum_{i=1}^k x_i$;

$\lambda_{q(x_k)} = \frac{b_2}{a} - \frac{b_1^2}{a(1+a(k-2))} + \sum_{i=1}^k x_i^2 - \frac{a}{1+a(k-2)}(\sum_{i=1}^k x_i)^2 - \frac{2b_1}{1+a(k-2)}\sum_{i=1}^k x_i$; e

$\nu_{n-k} = \frac{1+a(k+1)}{a}$

Prova: Se consideramos a seguinte partição

$$X_{(n)} = \begin{pmatrix} X_{(k)} \\ X_{(n-k)} \end{pmatrix}; \quad \mu_n = \frac{b_1}{1-2a} \begin{pmatrix} \mathbf{1}_{k\times 1} \\ \mathbf{1}_{n-k\times 1} \end{pmatrix} \quad \text{e}$$

$$\Sigma_n = I_n + \frac{1}{1-2a}\mathbf{1}_{n\times n} = \begin{bmatrix} \Sigma_{kk} & \Sigma_{k(n-k)} \\ \Sigma_{(n-k)k} & \Sigma_{(n-k)(n-k)} \end{bmatrix}$$

onde

$$\Sigma_{jj} = I_j + \tfrac{a}{1-2a}\mathbf{1}_{j\times j}$$

$$\Sigma_{ji} = \tfrac{a}{1-2a}\mathbf{1}_{j\times i};\ j \neq i;\ j,i \in \{k, n-k\}$$

e o resultado apresentado em Searle (1982, pp 132), o qual fornece que

$$\Sigma_{kk}^{-1} = I_k - \frac{a}{1+a(k-2)}\mathbf{1}_{k\times k},$$

a prova do resultado proposto segue imediatamente do lema 3.3.1.

□

Do resultado acima, obtém-se

$$E(X_{k+1}|X_{(k)}) = \tfrac{a\sum_{i=1}^{k} X_i + b_1}{1+a(k-2)} \quad \text{e}$$

$$\mathrm{Var}(X_{k+1}|X_{(k)}) = \tfrac{a}{1+a(k-2)}\left\{\tfrac{b_2}{a} - \tfrac{b_1^2}{a(1+a(k-2))} + \textstyle\sum_{i=1}^{k} X_i^2 - \tfrac{a}{1+a(k-2)}(\textstyle\sum_{i=1}^{k} X_i)^2 - \tfrac{2b_1\sum_{i=1}^{k} X_i}{1+a(k-2)}\right\}, \tag{5.6}$$

o que nos fornece o retorno esperado *a posteriori* para o dia $k+1$ e o erro associado à estimativa feita, respectivamente. Note que como $0 < a < 1/2$ e como $k \geq 1$,

$$\nu_{k+1} \geq 1/a + 2 \ \geq 4$$

o que garante a existência das estimativas acima para todo k.

Um outro resultado que nos interessa é a determinação da distribuição *a posteriori* para o retorno esperado (medido através da média) e para a volatilidade (medida através

da variância) do mercado - isto é, interessa-nos fazer inferências sobre (μ, σ^2) - quando as hipóteses da proposição 3.3.1 são válidas. Usando o teorema de Smith(1981) (teorema 3.2.2) e a proposição 3.3.1, verificamos que a quantidade aleatória (μ, σ^2) tem como distribuição *a posteriori* a distribuição Normal-Gama-Inversa especificada abaixo

$$\begin{aligned} \mu|\sigma^2, x_{(k)} &\sim \mathcal{N}\left(\tfrac{a\sum_{i=1}^{k} x_i + b_1}{1+a(k-2)}; \tfrac{a\sigma^2}{1+a(k-2)}\right) \quad \text{e} \\ \sigma^2|x_{(k)} &\sim \mathcal{GI}\left(\tfrac{g}{2}; \tfrac{a(k+1)+1}{a}\right), \end{aligned} \tag{5.7}$$

onde

$$g = \frac{b_2}{a} + \sum_{i=1}^{k} x_i^2 - \frac{(a\sum_{i=1}^{k} x_i + b_1)^2}{a + a^2(k-2)}.$$

Se adotamos como penalidade a função de perda quadrática obtemos como estimativas *a posteriori* para o retorno esperado e para a volatilidade do mercado os seguintes valores, respectivamente,

$$\begin{aligned} E(\mu|x_{(k)}) &= E\{E(\mu|\sigma^2, x_{(k)})\} = \frac{a\sum_{i=1}^{k} x_i + b_1}{1+a(k-2)} \\ &= \frac{ak}{1+a(k-2)}\bar{x}_k + \frac{1-2a}{1+a(k-2)}\frac{b_1}{1-2a} \quad \text{e} \qquad (5.8) \\ E(\sigma^2|x_{(k)}) &= \frac{ag}{a(k-1)+1}. \qquad (5.9) \end{aligned}$$

Ocorrendo algum imprevisto no tempo k de tal forma que o julgamento de $O(\mathbf{1})$-invariância seja mantido, mas as condições apontadas na expressão (5.4) sejam modificadas para

$$\begin{aligned} E^*(X_{k+3}|X_{k+1}, X_{k+2}) &= a^*(X_{k+1} + X_{k+2}) \quad + b_1^* \\ E^*(X_{k+3}^2|X_{k+1}, X_{k+2}) &= a^*(X_{k+1}^2 + X_{k+2}^2) \quad + b_2^*, \end{aligned} \tag{5.10}$$

onde $a^* \in (0, 1/2)$, $b_1^* \in \mathcal{R}$, $b_2^* > \frac{b_1^{*2}}{1-2a^*}$, $a^* \neq a$, $b_1^* \neq b_1$, $b_2^* \neq b_2$ e E^* denota o valor esperado obtido após o recebimento da informação imprevista, a distribuição *a posteriori* fornecida pela Regra Preditivista de Jeffrey é a seguinte distribuição t:

$$X_{k+1} \sim t\left(\frac{b_1^*}{1-2a^*}\ ;\ \frac{1-a^*}{1-2a^*}\ ;\ \frac{1}{a^*}\left(b_2^* - \frac{b_1^{*2}}{1-2a^*}\right); \frac{1+a^*}{a^*}\right) \tag{5.11}$$

e a distribuição a *a posteriori* para (μ, σ^2) é a seguinte distribuição Normal-Gama-Inversa

$$\begin{aligned} \mu|\sigma^2 &\sim \mathcal{N}^* \left(\tfrac{b_1^*}{1-2a^*} \; ; \; \tfrac{a^*\sigma^2}{2(1-2a^*)}\right) \\ \sigma^2 &\sim \mathcal{GI}^* \left(\tfrac{1}{2a^*}\left(b_2^* - \tfrac{(b_1^*)^2}{1-2a^*}\right) \; ; \; \tfrac{1+a^*}{2a^*}\right). \end{aligned} \tag{5.12}$$

Com isso podemos agora fazer previsões para retornos futuros após o recebimento da informação não previamente antecipada através do retorno esperado e determinamos o erro associado a tal estimativa, respectivamente, da seguinte forma

$$\begin{aligned} E^*(X_{k+1}) &= \frac{b_1^*}{1-2a^*} \\ \mathrm{Var}^*(X_{k+1}) &= \frac{b_2^*(1-2a^*) - b_1^{*2}}{(1-2a^*)^2}. \end{aligned} \tag{5.13}$$

onde Var* denota a variância obtida após a incorporação da nova informação.

Uma estimativa para o retorno esperado e para a volatilidade após o conhecimento de algum evento não previamente considerado é dada por

$$E^*(\mu) = \frac{b_1^*}{1-2a^*} \tag{5.14}$$

$$E^*(\sigma^2) = \frac{b_2^*(1-2a^*) - b_1^{*2}}{(1-a^*)(1-2a^*)}. \tag{5.15}$$

No apêndice Um apresentamos algumas tabelas ilustrando a influência dos valores de a, b_1 e b_2 no comportamento das variáveis aleatórias X_i's.

A seguir exibiremos um exemplo ilustrando o uso do algoritmo proposto, o qual foi implementado em FORTRAN e que é apresentado no Apêndice Dois, ao final deste trabalho.

Exemplo 5.1: A seguinte tabela mostra um exemplo de aplicação do algoritmo de previsão desenvolvido nesta seção, onde consideramos como Valor Real como sendo o valor da variável sobre a qual nos interessa fazer previsões e que será lida passo-a-passo pelo algoritmo. Na apresentação da tabela assumimos como zero valores previstos e esperados de ordem inferior a 10^{-4}.

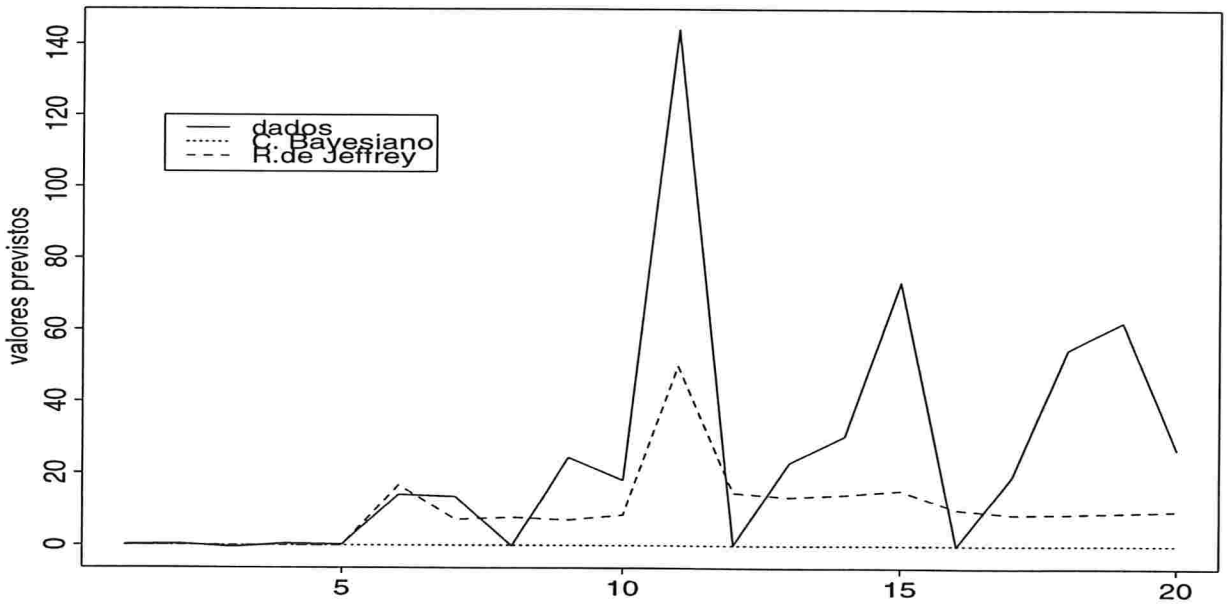
Tabela 5.9: Previsões e Erros obtidos utilizando apenas o Condicionamento Bayesiano e utilizando a Regra Preditivista de Jeffrey

	Valor Previsto		*Variância a posteriori*	
Valor Real	Bayes	Jeffrey	Bayes	Jeffrey
0.0937532	0	0	0.1000	0.1000
0.2348605	0	0	0.1000	0.1000
−0.5086271	0	0	0.1000	0.1000
0.4470131	0	0	0.1000	0.1000
0.2110936	0	0	0.1000	0.1000
14.0922356	0	16.666	0.1000	0.555
13.4479058	1.456×10^{-3}	7.121	0.1198	64.1297
−0.2593757	2.800×10^{-3}	7.754	0.1379	61.3189
24.4903995	2.774×10^{-3}	7.026	0.1379	61.0515
18.1194724	5.221×10^{-3}	8.481	0.1978	79.2642
144.0349004	7.03×10^{-3}	50.000	0.2306	4.9999
0.1140547	2.142×10^{-2}	14.6987	2.3029	1696.8662
23.0120705	2.143×10^{-2}	13.532	2.3027	1576.7725
30.5514515	2.373×10^{-2}	14.234	2.3552	1466.1387
73.4094281	2.678×10^{-2}	15.360	2.4481	1382.1216
−0.1692146	3.411×10^{-2}	10.204	2.9856	919.3461
19.4263359	3.408×10^{-2}	8.770	2.9853	802.9041
54.6328834	3.602×10^{-2}	8.863	3.0225	796.9011
62.3209464	4.147×10^{-2}	9.258	3.3198	807.935
26.7705502	4.769×10^{-2}	9.711	3.7066	824.8896

A amostra apresentada na primeira coluna da tabela anterior foi obtida encadeando-se sucessivamente quatro amostras de tamanho cinco geradas respectivamente das seguintes distribuições $t(0.0001, 0, 0.1)$, $t(0.2, 10, 167)$, $t(0.4, 10, 501)$ e $t(0.01, 10, 1003)$, onde $t(a, b_1, b_2)$ e a distribuição fornecida na expressão (5.5).

Perceba que as previsões feitas usando a Regra Preditivista de Jeffrey são mais próximas do valor real do que aquelas fornecidas usando o Condicionamento Bayesiano apenas, ou seja, sem considerar qualquer método alternativo para incorporar mudanças não previamente consideradas. O Condicionamento Bayesiano além de fornecer previsões em geral ruins, também fornece variância *a posteriori* muito pequenas (ver tabela 5.9) indicando que as previsões feitas utilizando este procedimento são muito precisas. Note também que o Condicionamento Bayesiano não incorpora as mudanças bruscas ocorridas nos parâmetros (tabela 5.10), o que também pode ser observados no seguinte gráfico:

Figura 5.15: Estudo Comparativo entre as Previsões feitas pela Regra de Jeffrey e Condicionamento Bayesiano



Note na próxima tabela que o Condicionamento Bayesiano não capta as modificações sofridas por μ e σ^2.

Tabela 5.10: Comportamento *a Posteriori* de μ e σ^2

Valor Esperado para μ		*Valor Esperado para* σ^2	
Bayes	Jeffrey	Bayes	Jeffrey
0	0	0.1000	0.1000
0	0	0.1000	0.1000
0	0	0.1000	0.1000
0	0	0.1000	0.1000
0	0	0.1000	0.1000
0	16.666	0.1000	0.555
1.456×10^{-3}	7.121	0.1198	64.1297
2.800×10^{-3}	7.754	0.1379	61.3189
2.774×10^{-3}	7.026	0.1379	61.0515
5.221×10^{-3}	8.481	0.1978	79.2642
7.03×10^{-3}	50.000	0.2306	4.9999
2.142×10^{-2}	14.6987	2.3029	1696.8662
2.143×10^{-2}	13.532	2.3027	1576.7725
2.373×10^{-2}	14.234	2.3552	1466.1387
2.678×10^{-2}	15.360	2.4481	1382.1216
3.411×10^{-2}	10.204	2.9856	919.3461
3.408×10^{-2}	8.770	2.9853	802.9041
3.602×10^{-2}	8.863	3.0225	796.9011
4.147×10^{-2}	9.258	3.3198	807.935
4.769×10^{-2}	9.711	3.7066	824.8896

Nosso próximo passo consiste em aplicar esse algoritmo de previsão a dados reais. Escolhemos para isso a série Rentabilidade ENDESA Mensal para a qual a modelagem sugerida anteriormente fica justificada pelas suposições da teoria sobre Mercados Eficientes. Apresentamos a seguir os resultados obtidos utilizando várias distribuições *a priori* e supondo haver mudanças em outubro de 1990, agosto de 1992 e setembro de 1993.

Figura 5.16: Previões - Rentabilidade ENDESA

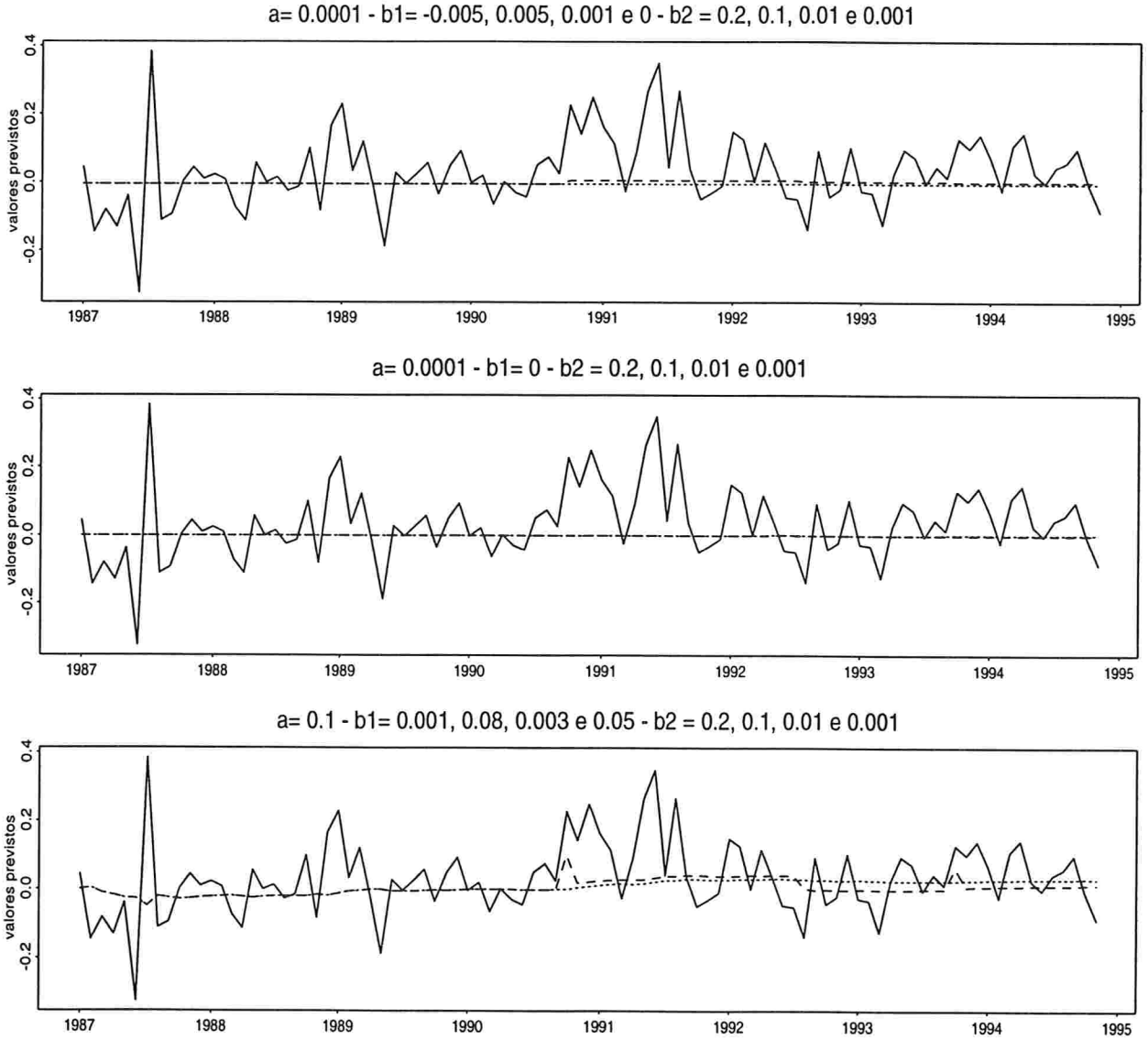
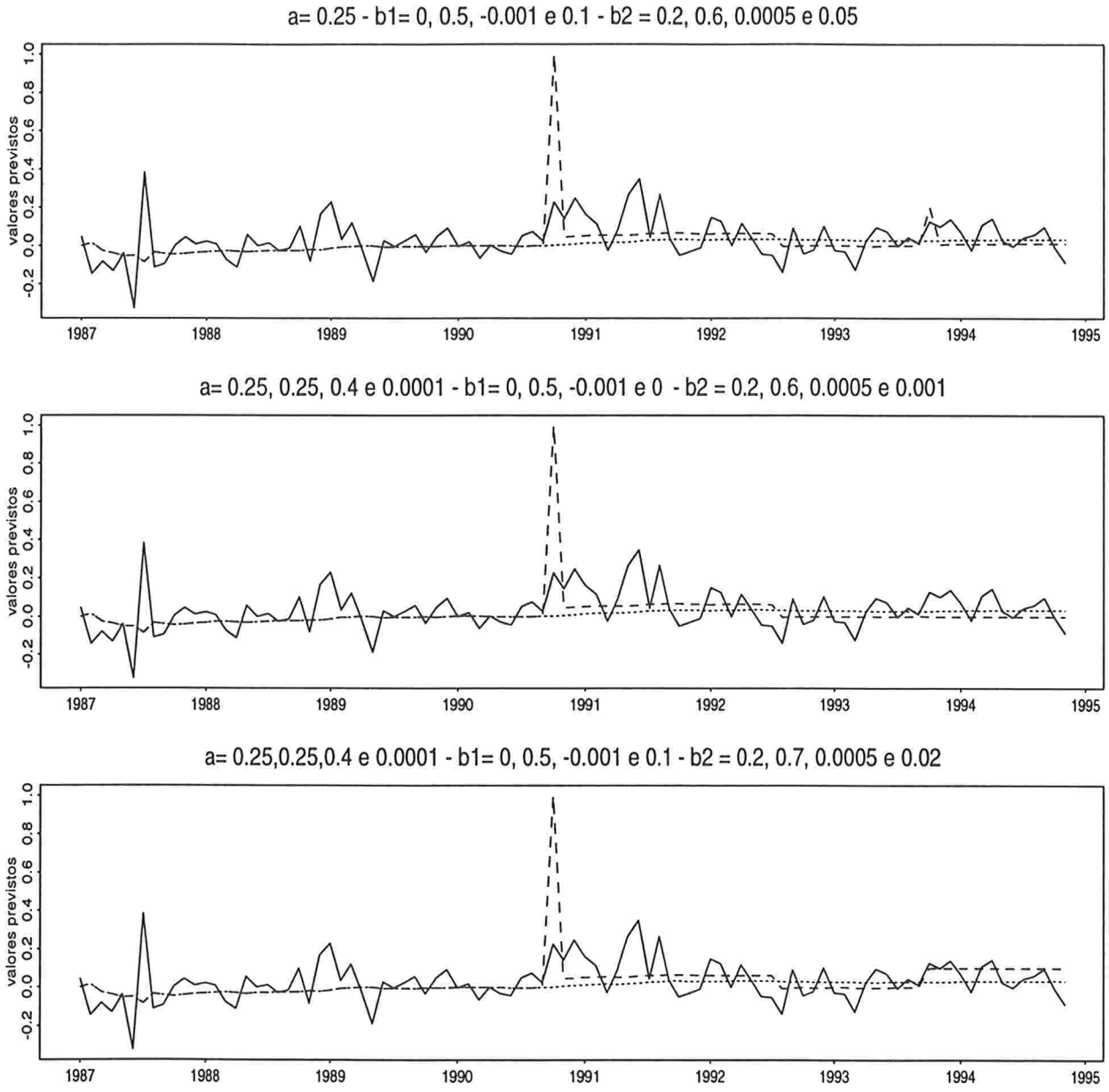


Figura 5.17: Previões - Rentabilidade ENDESA



OBS: a linha cheia representa os dados reais, a tracejada representa as previsões com a Regra de Jeffrey e a pontilhada as previsões com o Condicionamento Bayesiano. No que segue a linha cheia representa o resultado obtido quando o Condicionamento Bayesiano é utilizado e a linha tracejada simboliza os resultados obtidos quando a Regra de Jeffrey é considerada.

A seguir apresentamos o comportamento dos erros, do retorno esperado e volatilidade *a posteriori* para cada procedimento de previsão sugerido. As distribuições *a priori* utilizadas são respectivamente as mesmas adotadas para as previsões, exibidas nas figuras 5.16 e 5.17.

Figura 5.18: Erro *a posteriori* - Rentabilidade ENDESA

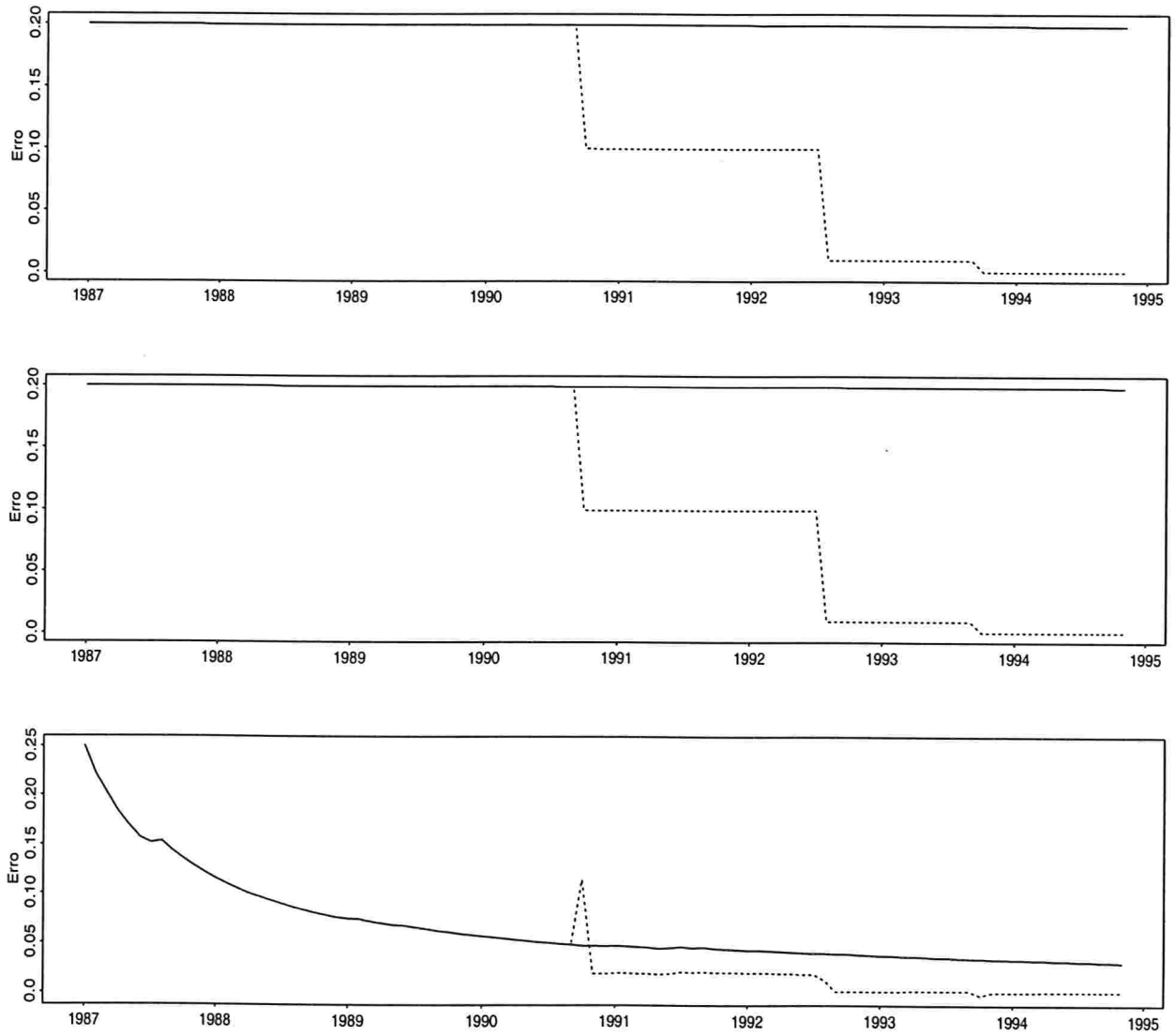
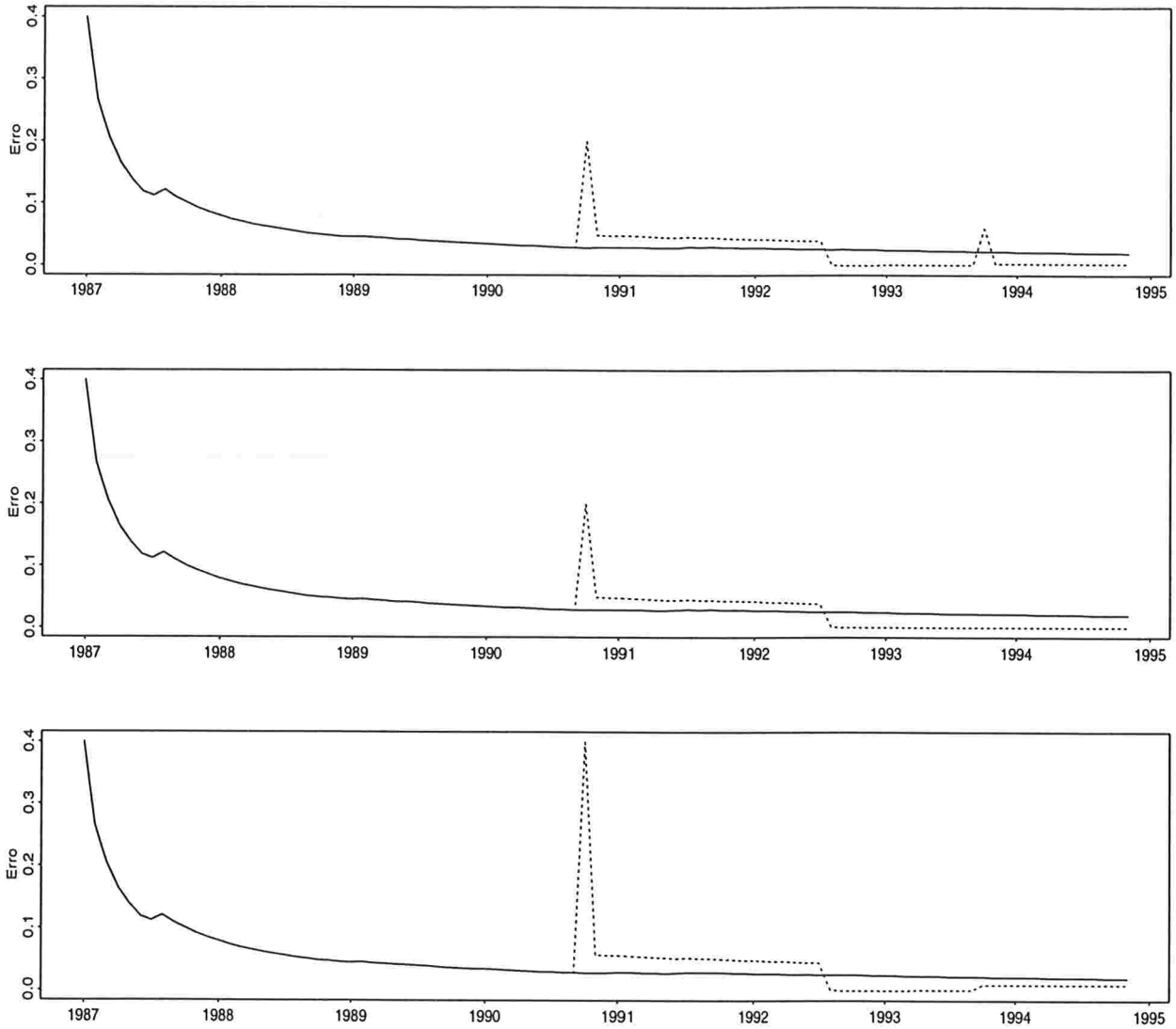


Figura 5.19: Erro *a posteriori* - Rentabilidade ENDESA



Os próximos gráficos mostram o comportamento do retorno esperado e da volatilidade *a posteriori*. Perceba a eficiência da Regra de Jeffrey na incorporação de mudanças no comportamento destes dois objetos.

Figura 5.20: Retorno Médio *a posteriori* - Rentabilidade ENDESA

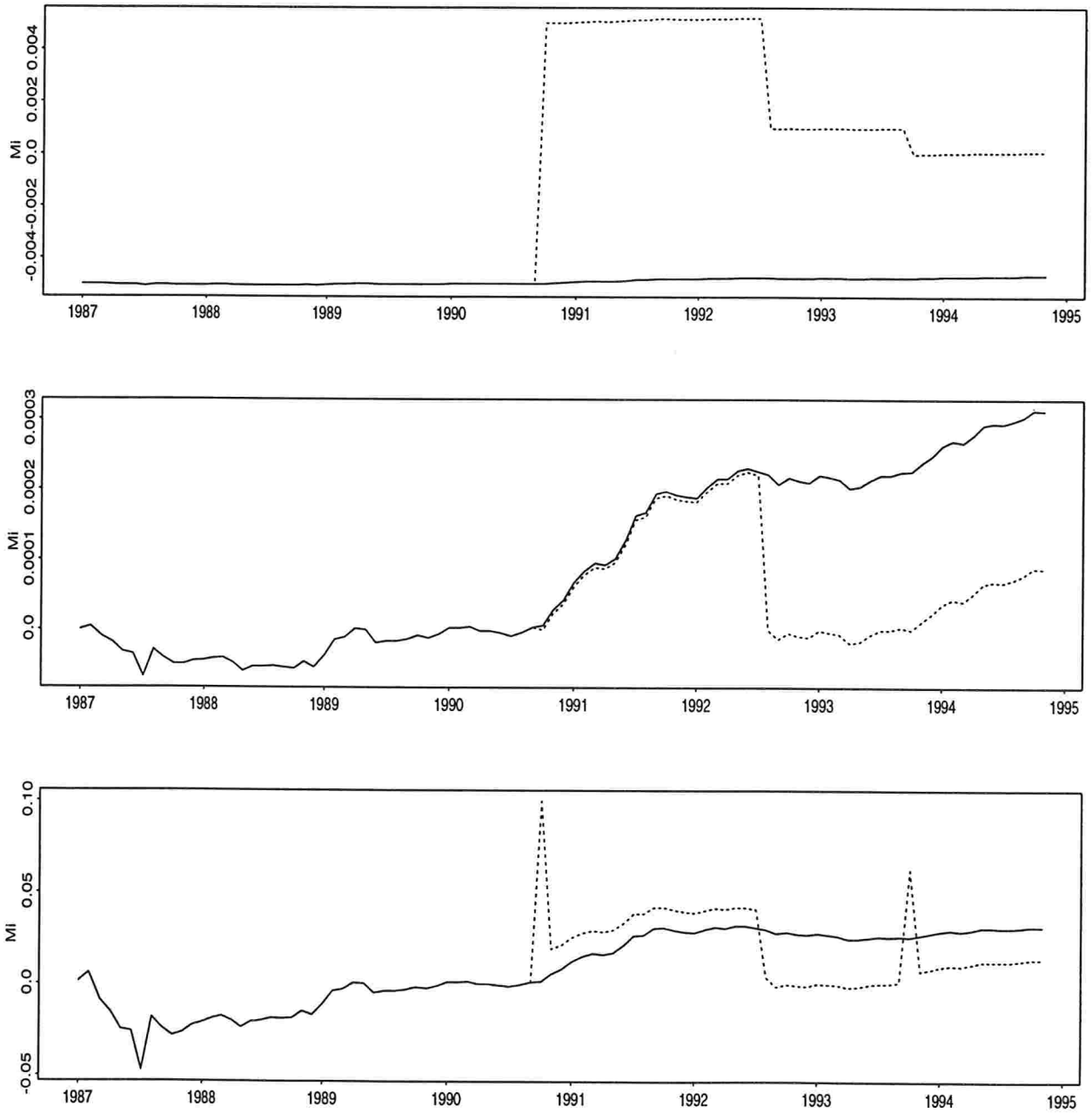


Figura 5.21: Retorno Médio *a posteriori* - Rentabilidade ENDESA

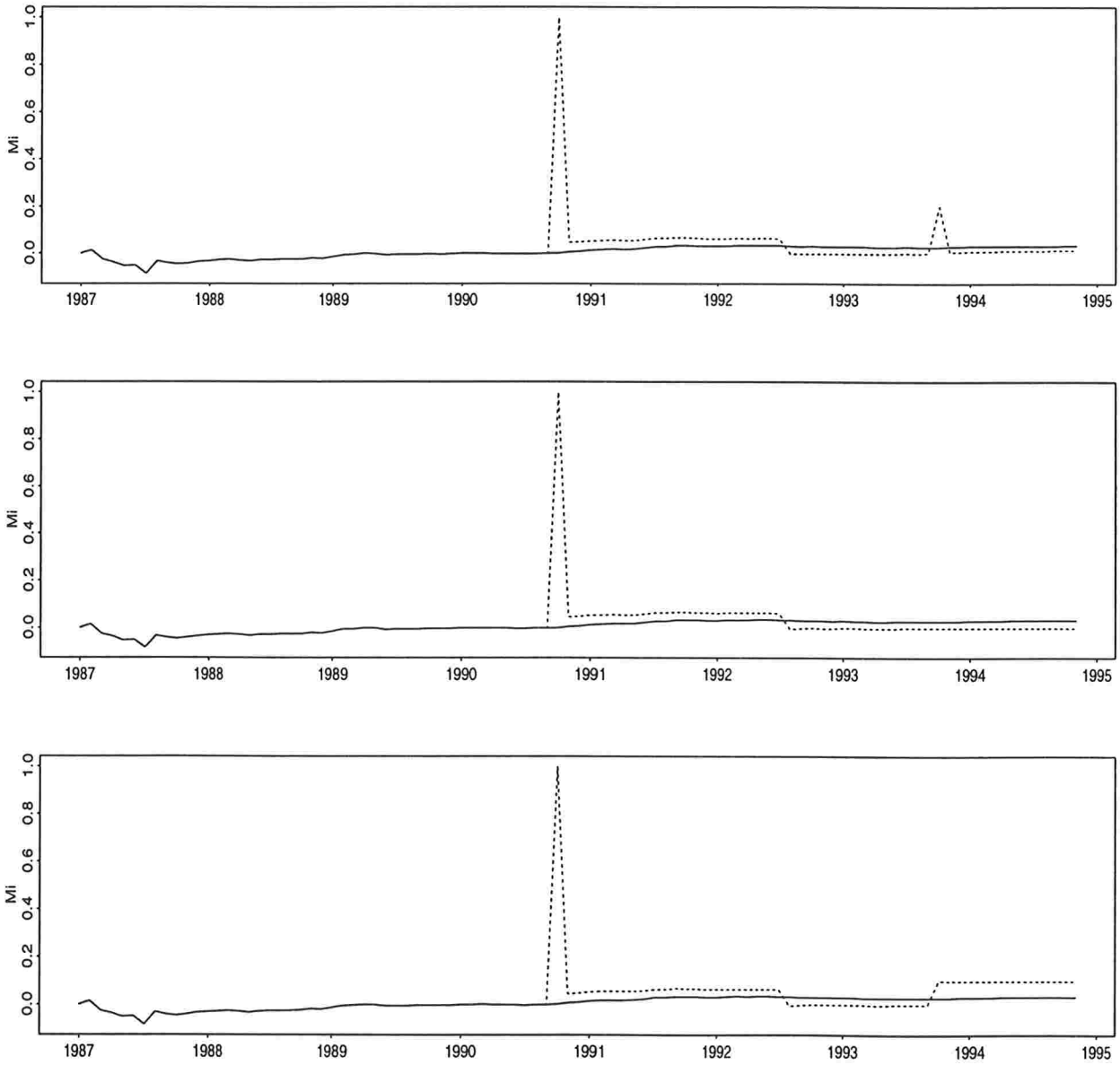


Figura 5.22: Volatilidade *a posteriori* - Rentabilidade ENDESA

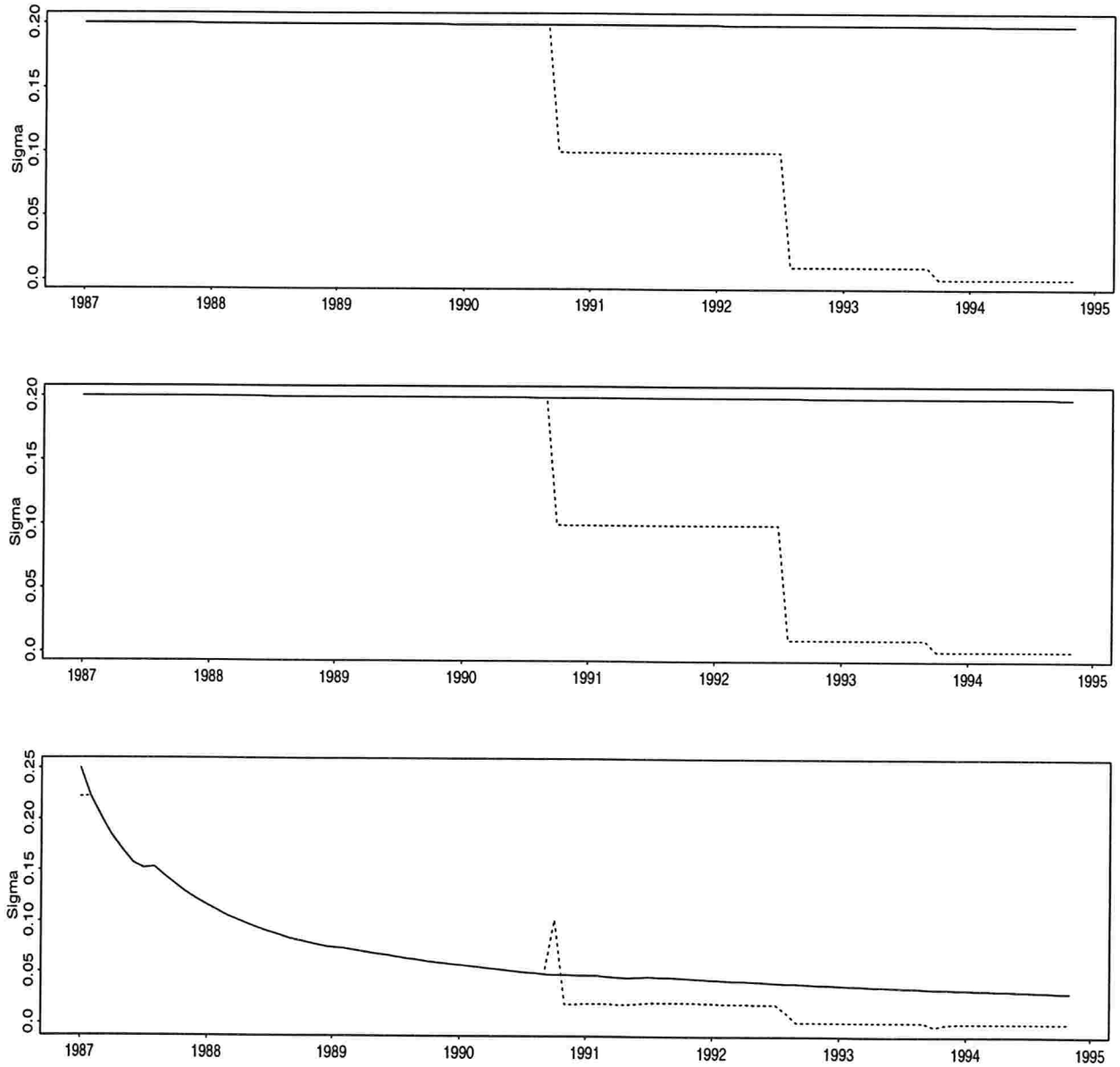
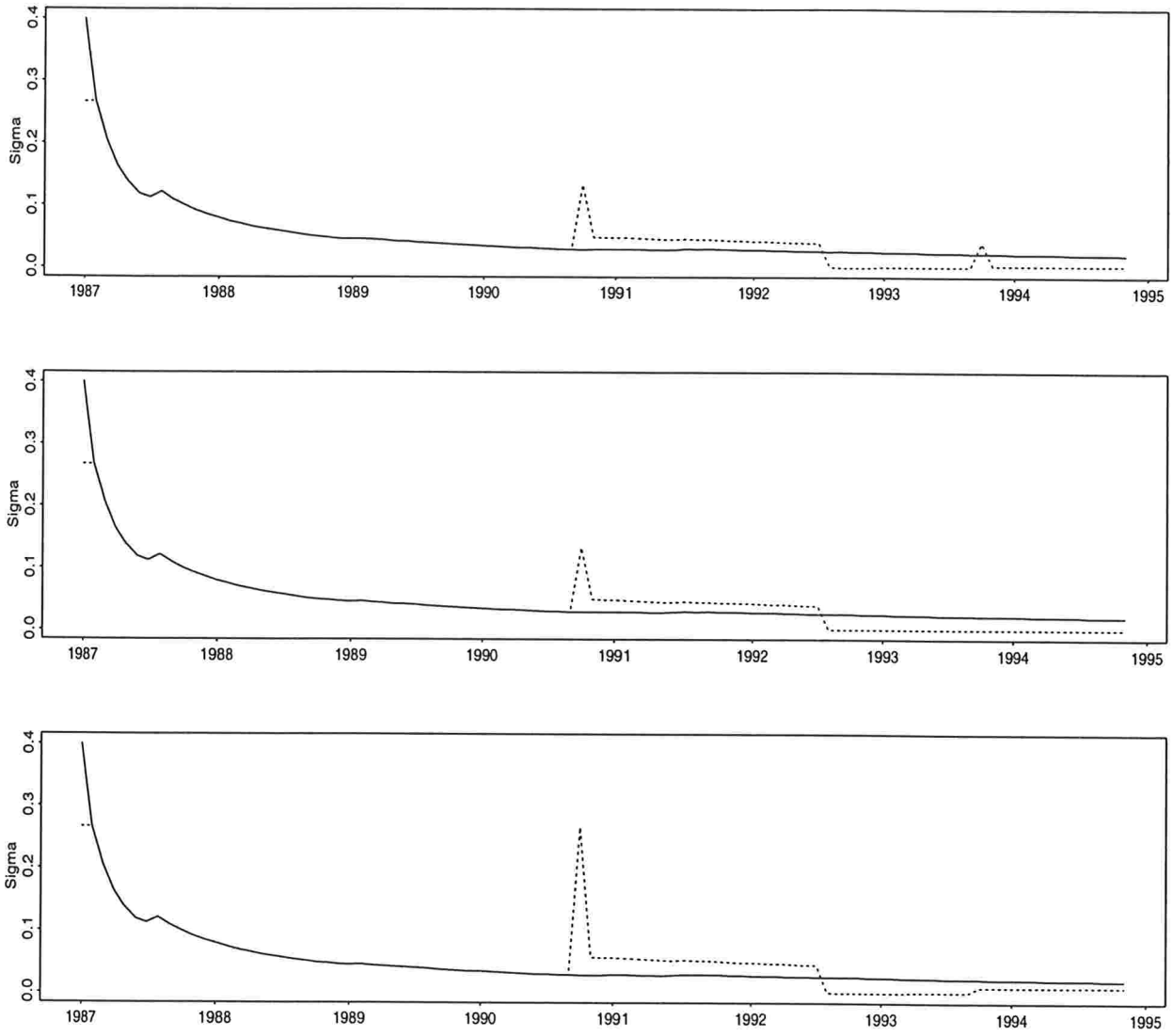


Figura 5.23: Volatilidade *a posteriori* - Rentabilidade ENDESA



Pode-se perceber que o mecanismo de previsões que envolve a Regra de Jeffrey é mais eficiente em incorporar modificações bruscas na série do que demonstra ser o procedimento que só utiliza o Condicionamento Bayesiano. Ver figuras 5.20, 5.21, 5.22 e 5.23, as quais fornecem as estimativas *a posteriori* para os parâmetros intrínsecos envolvidos no teorema de Smith(1981) - ou seja, no contexto visto aqui, para a volatilidade e para o retorno esperado.

As previsões feitas utilizando a Regra de Jeffrey em geral são melhores do que aquelas feitas com o Condicionamento Bayesiano (ver figura 5.17) - apesar de ambos apresentarem

precisão geralmente baixa - mas como estas previsões dependem fortemente da sua opinião sobre a série, elas podem ser piores do que as previsões feitas com o Condicionamento Bayesiano se a sua opinião não refletir de maneira suficientemente boa as consequências da nova informação sobre o mercado (ver terceiro gráfico, figura 5.16).

Na próxima seção veremos uma aplicação dos procedimentos desenvolvidos no capítulo 4 a esta mesma série de dados.

5.5 O MPP e o Mercado Chileno

Nesta seção aplicaremos os procedimentos desenvolvidos no capítulo 4 à série Rentabilidade Mensal ENDESA. Nosso objetivo aqui é apresentar o comportamento *a posteriori* para o retorno esperado (μ) e para a volatilidade (σ^2) e fazer inferências sobre o número de pontos de mudança que ocorrem nesta série. A partição mais provável será apresentada com sua respectiva probabilidade.

De acordo como as justificativas dadas na seção 5.2, dentro de cada bloco $[ij]$ as observações são julgadas condicionalmente iid's com distribuição $\mathcal{N}(\mu^{ij}, (\sigma^{ij})^2)$ e também pelos argumentos dados na seção 5.2, adotaremos uma distribuição *a priori* Normal-Gama-Inversa para descrever a incerteza sobre o parâmetro $(\mu^{ij}, (\sigma^{ij})^2)$ do bloco $[ij]$. Isto é vamos considerar que *a priori*

$$\mu^{ij} \sim \mathcal{N}(m, \tau(\sigma^{ij})^2), \tag{5.16}$$

onde m é um número real qualquer e τ é um número real positivo, e por esperar-se que, quanto maior a volatilidade, menor o bloco de observações formados, assumiremos inicialmente que

$$(\sigma^{ij})^2 \sim \mathcal{GI}\left(\frac{(j-i)a}{2}; \frac{b(j-i)^2}{2} + 1\right), \tag{5.17}$$

onde a e b são números reais positivos. Perceba que com esta escolha e considerando $a = b$ a expectativa inicial sobre a relação entre a volatilidade e o tamanho do bloco é satisfeita pois, neste caso,

$$E((\sigma^{ij})^2) = \frac{a}{b(j-i)} = \frac{1}{j-i}.$$

Como coesões *a priori* assumiremos aquelas fornecidas na expressão (4.33), pois segundo a teoria de mercados eficientes pontos de mudanças ocorrem devido ao conhecimento de informações não previamente consideradas e, portanto, pontos de mudanças já ocorridos não são informativos sobre pontos de mudanças futuros (Ver seção 4.4). Como antes mencionado, valores pequenos de p favorecem a formação de poucos blocos, o que parece ser razoável para o mercado chileno. Apesar disto, a análise feita aqui contemplará vários valores de p. Também assumiremos que $\tau = 1$.

Nas figuras 5.24 a 5.29 apresentamos as estimativas *a posteriori* para μ (primeira coluna) e σ^2 (segunda coluna), assumindo como regra de penalidade a função de perda quadrática. Para termos uma idéias sobre a sensibilidade do modelo, apresentamos estes e os demais resultados a ser visto nesta seção considerando distintos parâmetros para a distribuição *a priori* de $(\mu^{ij}, (\sigma^{ij})^2)$.

As médias *a posteriori* para μ^{ij} e $(\sigma^{ij})^2$ serão contrastados, respectivamente, com as médias e variâncias móveis de ordem 10, para termos uma idéia se o comportamento das estimativas é similar àquele observado na análise descritiva.

A seguir, apresentamos as médias *a posteriori* para μ e σ^2 considerando $a = b = 0.5$ e diferentes valores para m. As médias *a posteriori* são apresentadas em linhas cheias e as médias e variâncias móveis são apresentadas em linhas pontilhadas.

Figura 5.24: Médias *a posteriori* para μ e σ^2 - $m = 0$, $a = 0.5$ e $b = 0.5$

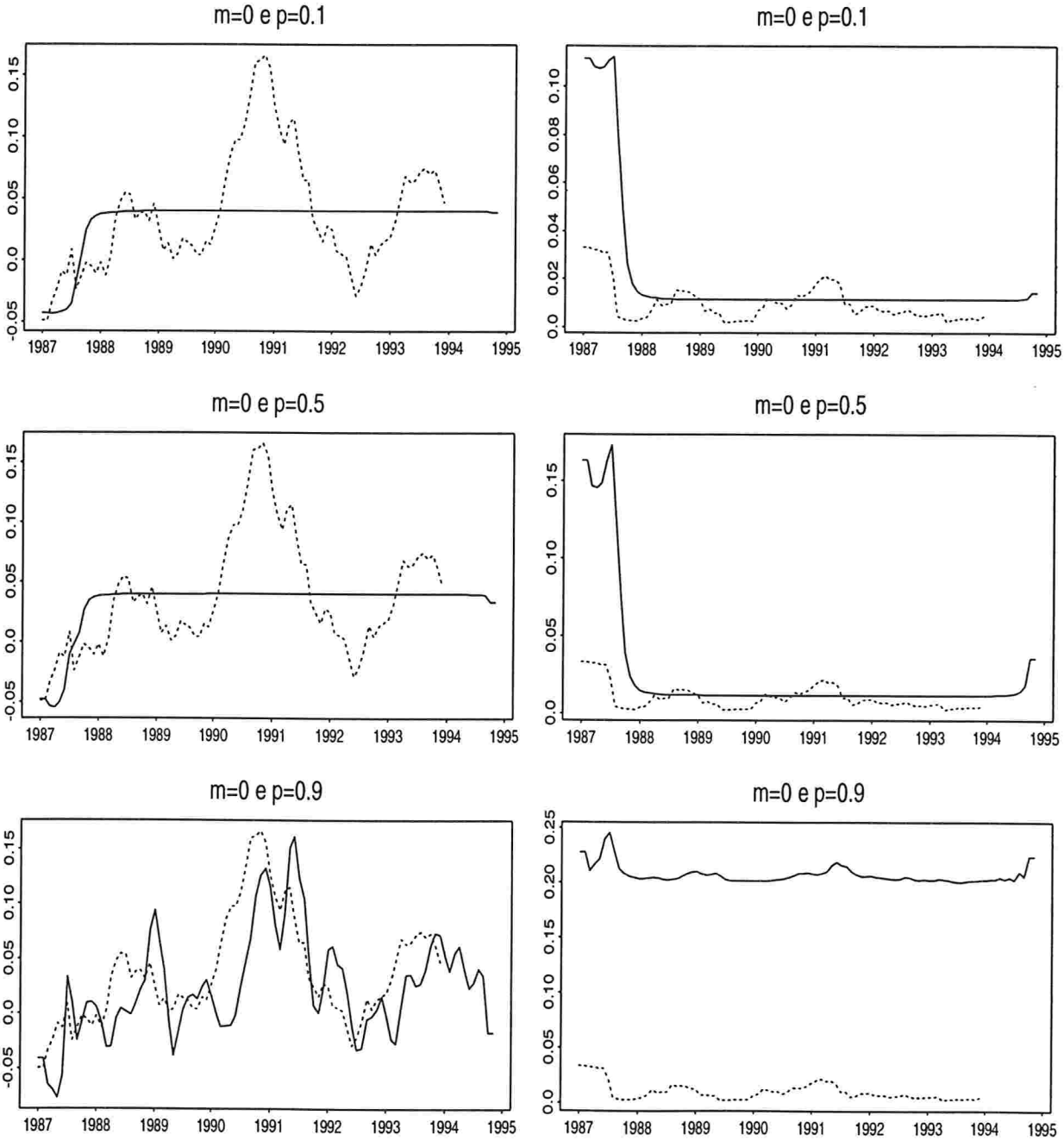
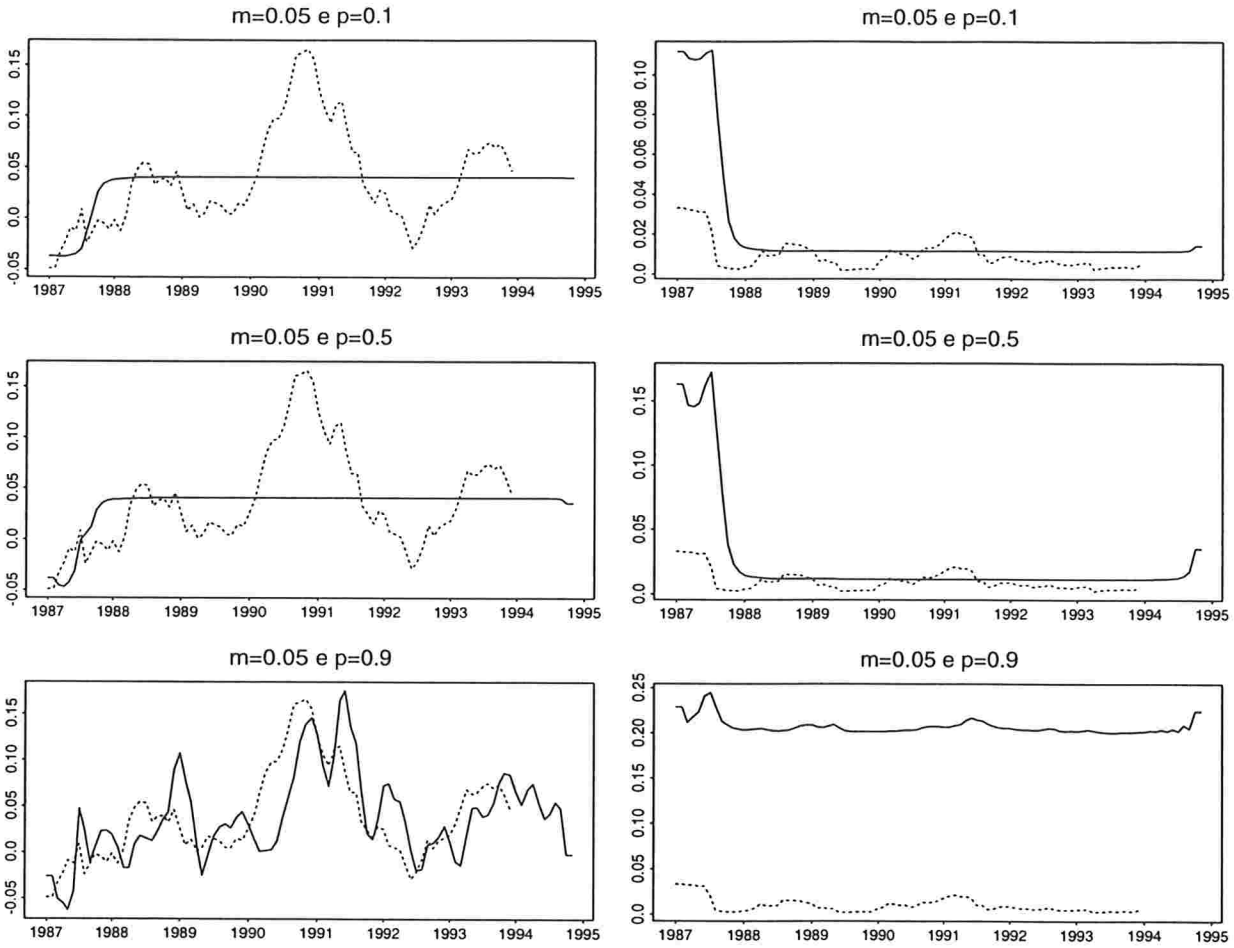


Figura 5.25: Médias *a posteriori* para μ e σ^2 - $m = 0.05$, $a = 0.5$ e $b = 0.5$

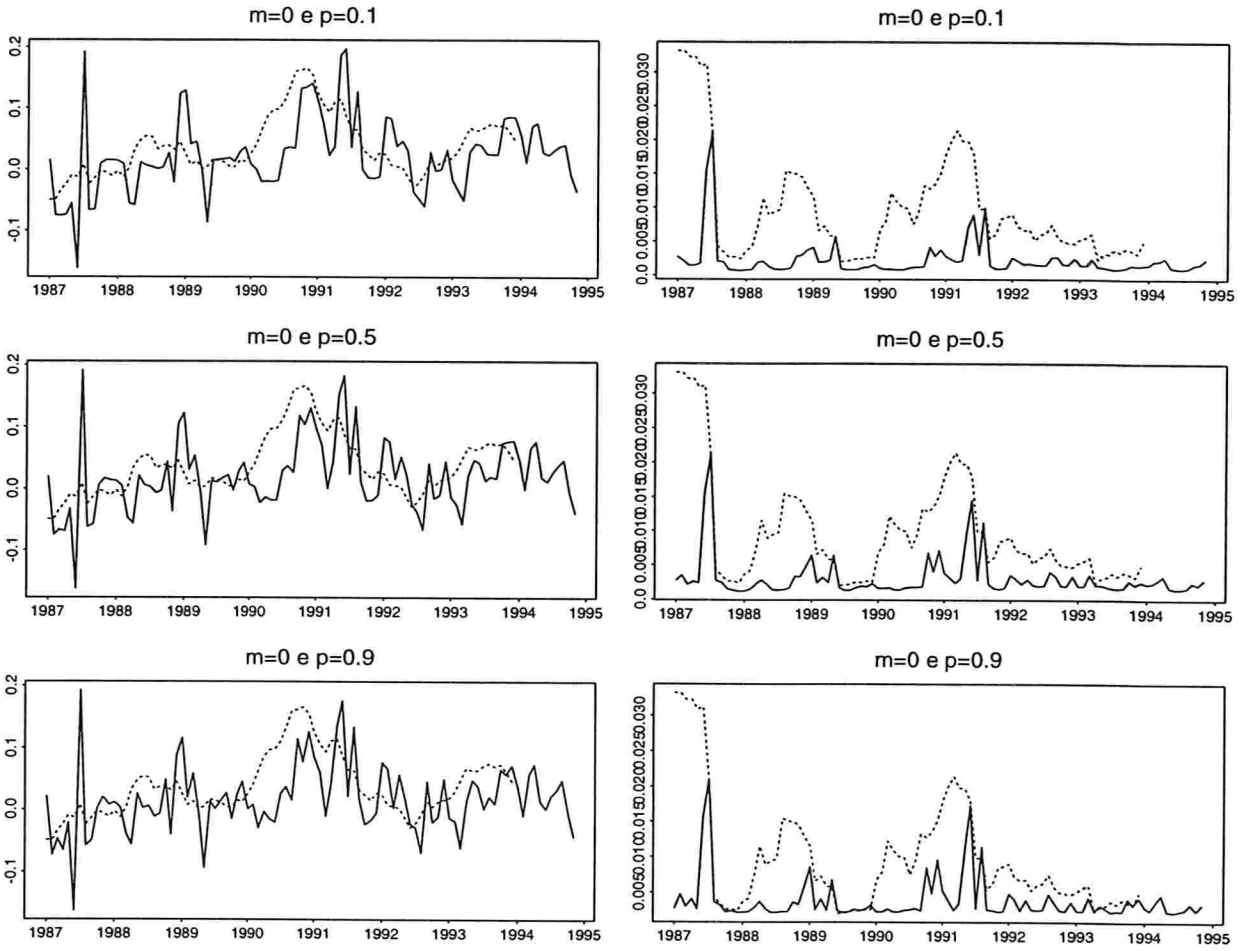


Perceba que as estimativas tanto para μ quanto para σ^2 dependem do valor adotado para p. Nos casos vistos nas figuras 5.24 e 5.25, as estimativas para μ estão próximas dos valores atingidos pelas médias móveis quando $p = 0.9$, onde pode-se perceber um grande número de pontos de mudanças. Para $p = 0.1$ e $p = 0.5$ o número de pontos de mudanças detectados é pequeno, o que vem de encontro com o que se esperava para a série analisada. Qualquer que seja o valor de m, as estimativas para σ^2 estão em geral afastadas dos valores observados para as variâncias móveis quando $p = 0.9$ e para o ano de 1987, quando $p = 0.1$ e $p = 0.5$. Também observa-se que com valores pequenos de p o número de pontos de mudanças na variância é pequeno, os quais ocorrem durante o ano de 1987 e no final de 1994 e além disto, as médias *a posteriori* obtidas são mais suaves, tanto para μ quanto para σ^2, quando

comparadas com as médias e variâncias móveis.

Considerando que o comportamento de $(\sigma^{ij})^2$ é descrito pela distribuição dada na expressão (5.17) com $a = 0.01$ e $b = 3$ e considerando $m = 0$, obtém-se

Figura 5.26: Médias *a posteriori* para μ e σ^2 - $m = 0$, $a = 0.01$ e $b = 3$

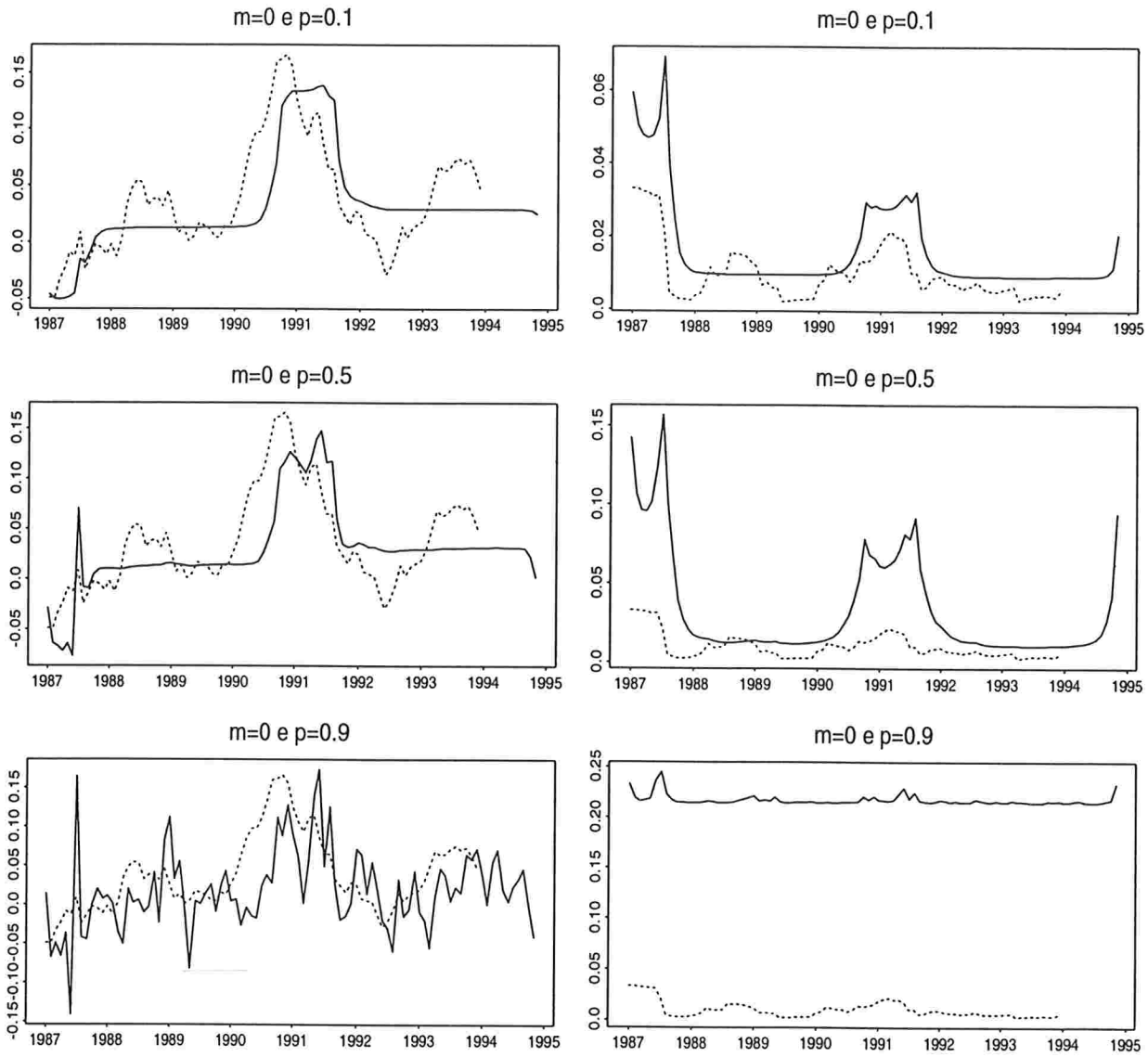


Perceba que, na figura 5.26, as médias *a posteriori* para σ^2 têm comportamento mais próximo do comportamento das variâncias móveis que o observado anteriormente nas figuras 5.24 e 5.25, independentemente do valor escolhido para p. Também, independentemente do valor de p o número de pontos de mudança tanto no retorno médio quanto na volatilidade são muito altos.

Considerando, novamente, $m = 0$ e admitindo agora que $a = 1$ e $b = 3$ as seguintes

estimativas *a posteriori* foram observadas

Figura 5.27: Médias *a posteriori* para μ e σ^2 - $m = 0$, $a = 1$ e $b = 3$



Neste caso, a escolha de $a = 1$ e $b = 3$ parece não ser uma boa escolha para descrever a incerteza sobre a volatilidade, fornecendo novamente estimativas para σ^2 cujo comportamento destoa do comportamento das variâncias móveis. Por outro lado, para valores pequenos de p, percebe-se que as mudanças no retorno médio e volatilidade ocorrem basicamente nos períodos previamente mencionados na análise descritiva.

Assumindo que volatilidade e tamanho dos blocos não se relacionam diretamente, consideraremos uma nova distribuição *a priori* para $(\sigma^{ij})^2$ a qual não depende do tamanho do bloco. Nos dois graficos que seguem, assumimos que $(\sigma^{ij})^2$ tem seu comportamento descrito pela distribuição $\mathcal{GI}(a/2, d/2)$ para todo i e j, onde $a > 0$ e $d > 0$.

Figura 5.28: Médias *a posteriori* para μ e σ^2 - $m = 0$, $a = 0.01$ e $d = 3$

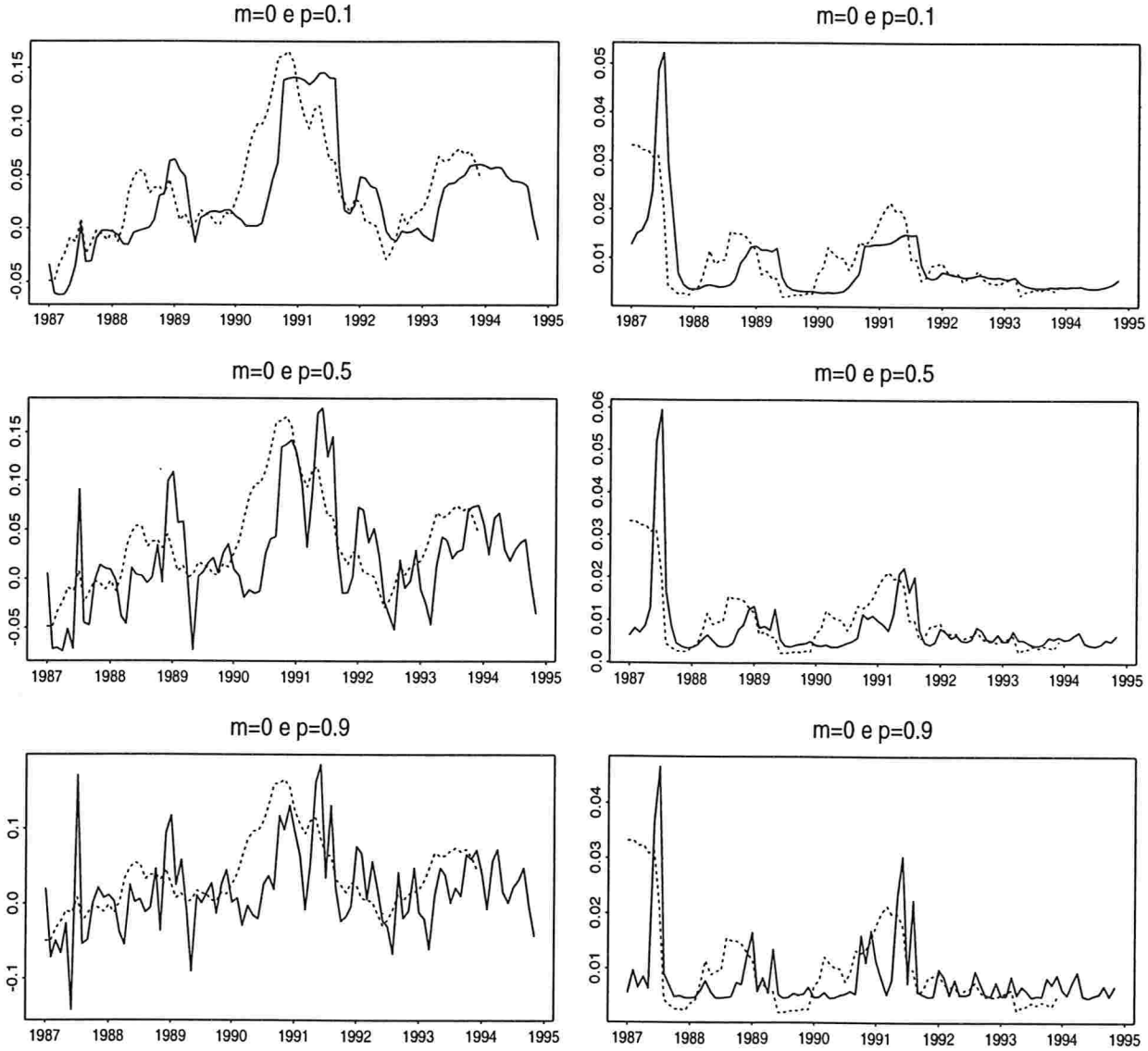
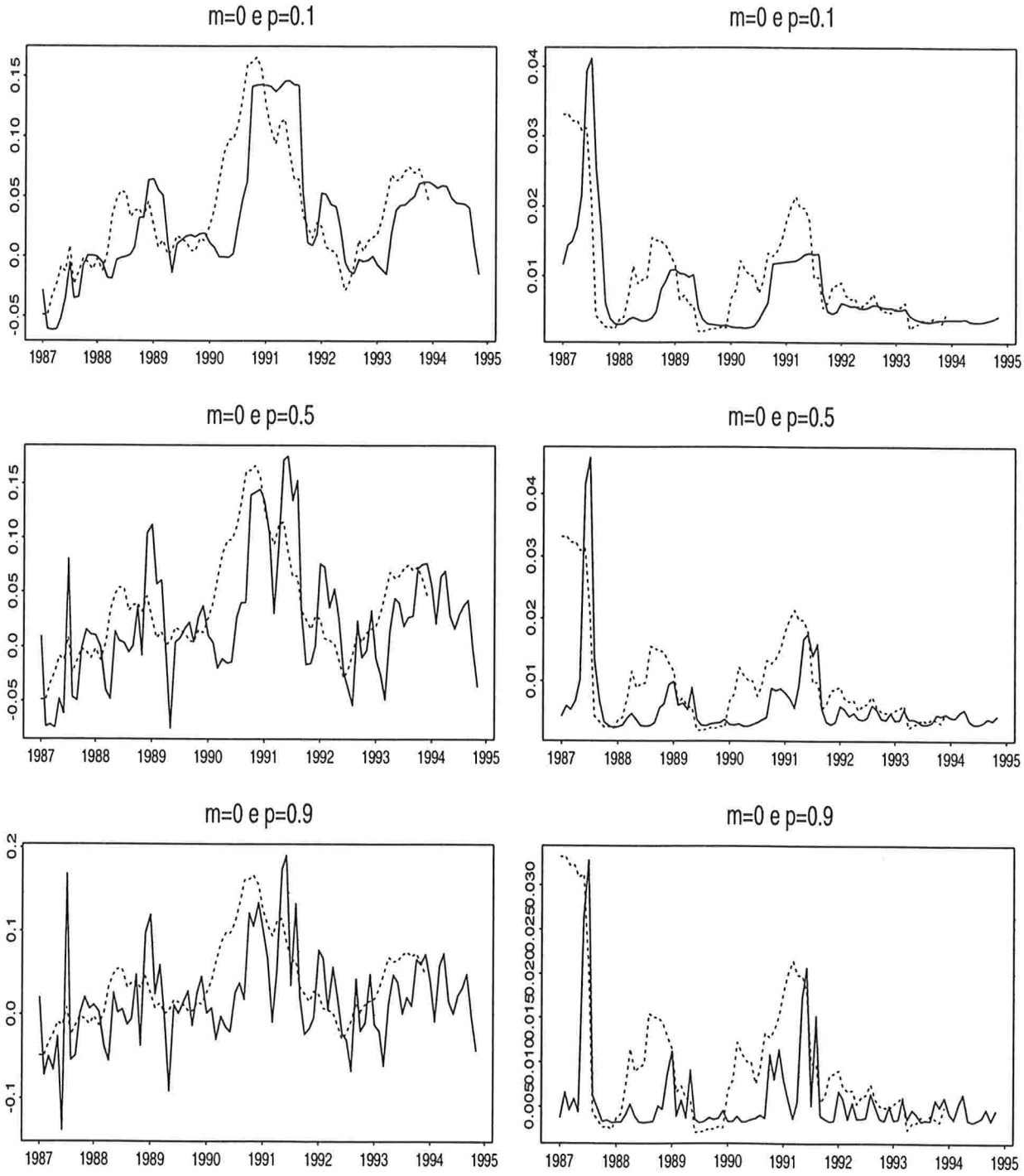


Figura 5.29: Médias *a posteriori* para μ e σ^2 - $m = 0$, $a = 0.01$ e $d = 4$



Nos dois casos, as estimativas para o retorno médio são muito parecidas quando as comparamos para o mesmo valor de p e também apresentam comportamento dentro do esperado. Já para a volatilidade, obtivemos valores mais próximos do esperado quando $d = 4$. Perceba que ao assumirmos uma distribuição *a priori* para $(\sigma^{ij})^2$ que não depende do tamanho do bloco, obtivemos estimativas para a volatilidade com comportamento muito similar ao observado para as variâncias móveis. Isto pode ter ocorrido por não termos feito escolhas adequadas para os valores de a e b, nos casos anteriores. Como antes, observamos que, também para estas distribuições *a priori*, o número de pontos de mudanças detectados aumenta com o valor de p.

A seguir apresentaremos a estimativa para o número de blocos referentes a cada uma das distribuições *a priori* consideradas nas figuras 5.24 a 5.29 e gráficos mostrando a relação entre μ e σ^2.

De acordo com a metodologia apresentada na seção 4.5, para obtermos a estimativa *a posteriori* para o número de blocos B e para a partição ρ geramos uma amostra de 10 mil seqüências ou vetores de zeros e uns, com tamanho 95, partindo de uma seqüência inicial de zeros. (Sementes diferentes foram consideradas, as quais geraram amostras com comportamentos similares aos obtidos partindo da semente nula, por isso optamos por apresentar os resultados obtidos partindo da semente nula). Observando o comportamento das médias ergódicas, percebemos que a amostra gerada atinge estabilidade após três mil iterações. Desprezamos, então, as primeiras seqüências geradas e consideramos como amostra final todas as seqüências a partir deste ponto de estabilização. Apesar de encontrarmos auto-correlação não muito próxima de zero para alguns valores de p, ainda assim a amostra considerada é adequada pois estamos interessados apenas nas estimativas de B e ρ e não na variância destes estimadores (ver Gamerman(1997) para detalhes). Estudos sobre a auto-correlação da amostra e sobre a convergência para a distribuição *a posteriori* para B são apresentados no Apêndice Quatro.

Figura 5.30: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão para μ e σ^2- $m = 0$, $a = 0.5$ e $b = 0.5$

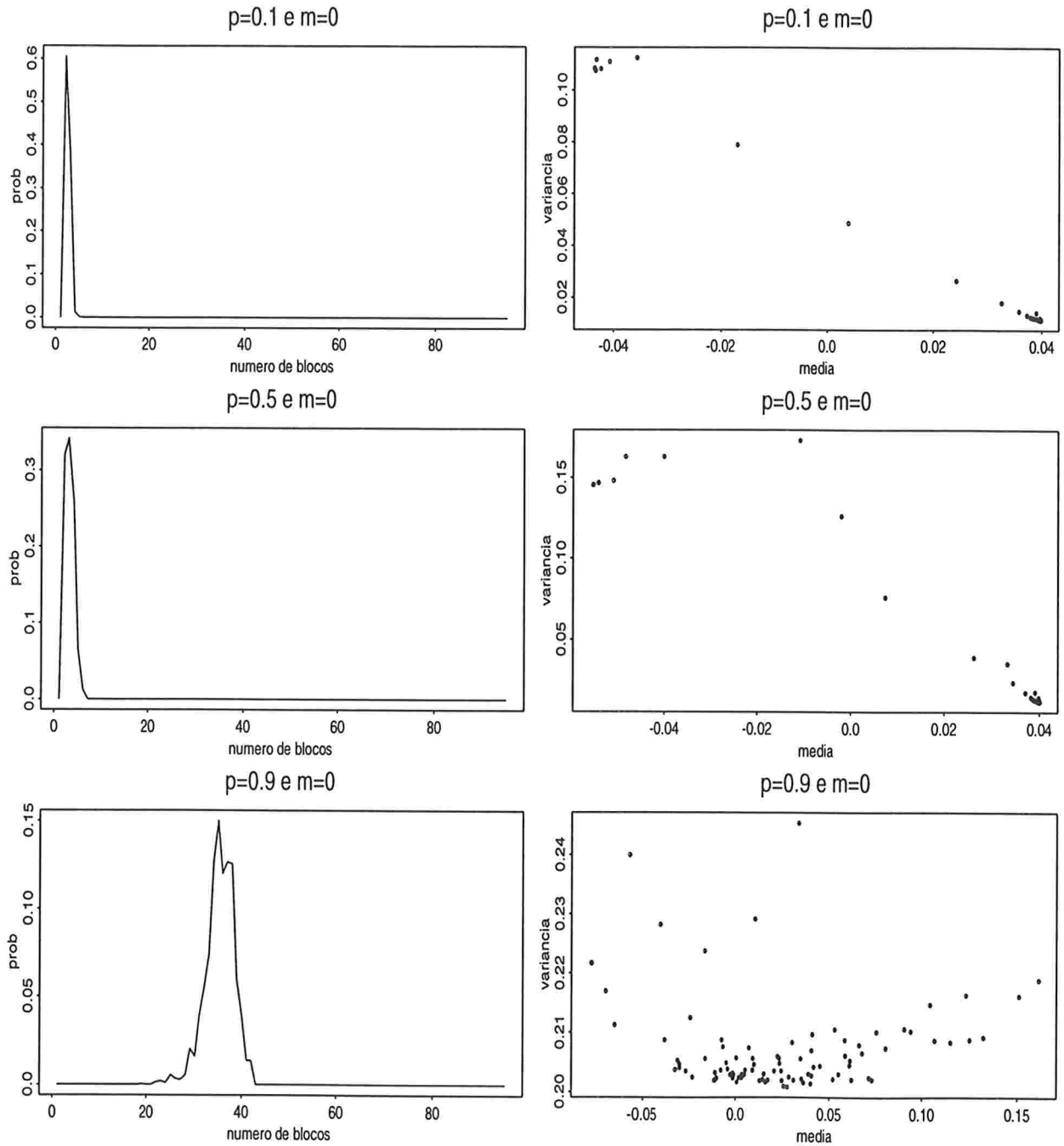


Figura 5.31: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão para μ e σ^2- $m = 0.05$, $a = 0.5$ e $b = 0.5$

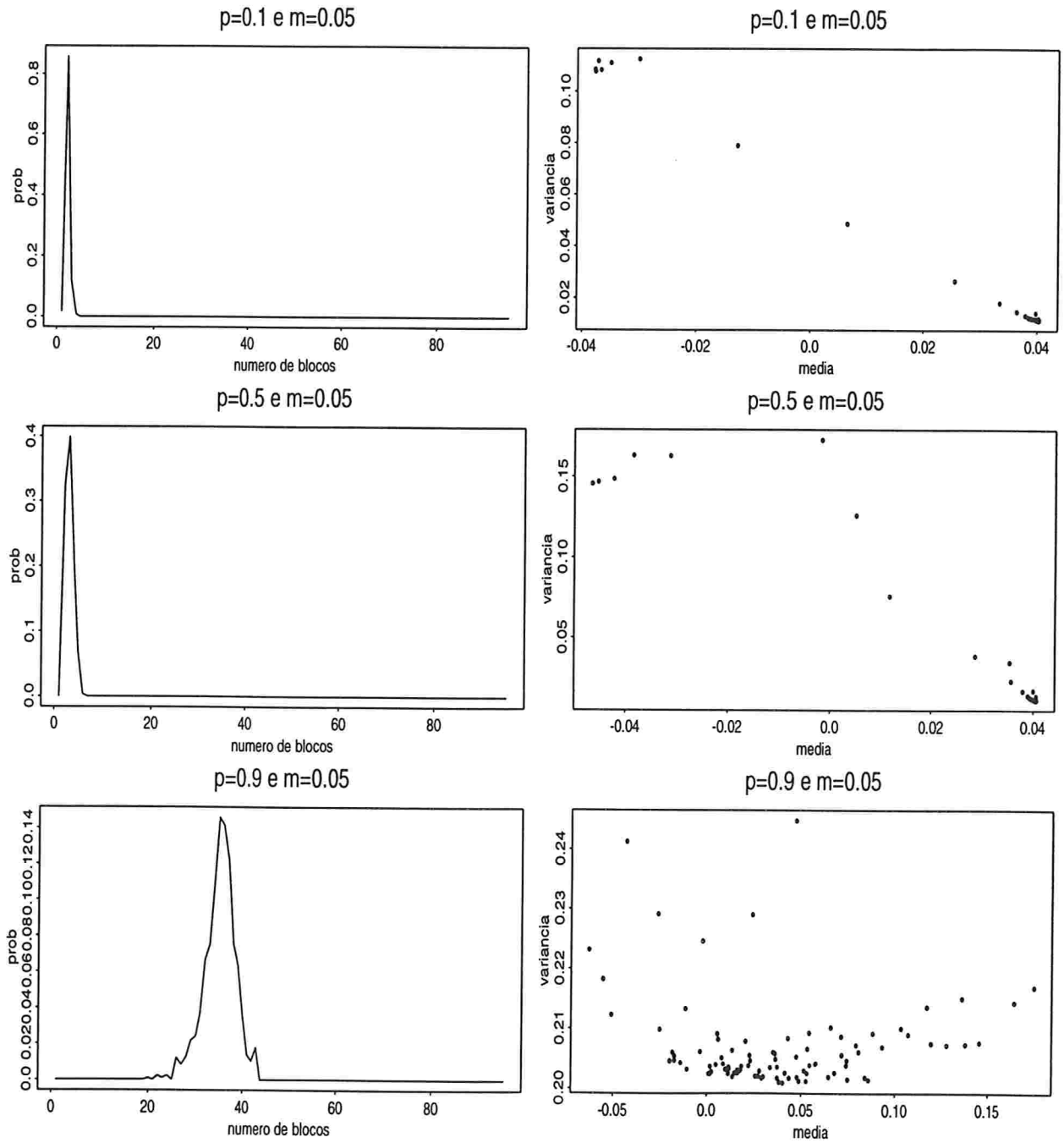


Figura 5.32: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão para μ e σ^2- $m = 0$, $a = 0.01$ e $b = 3$

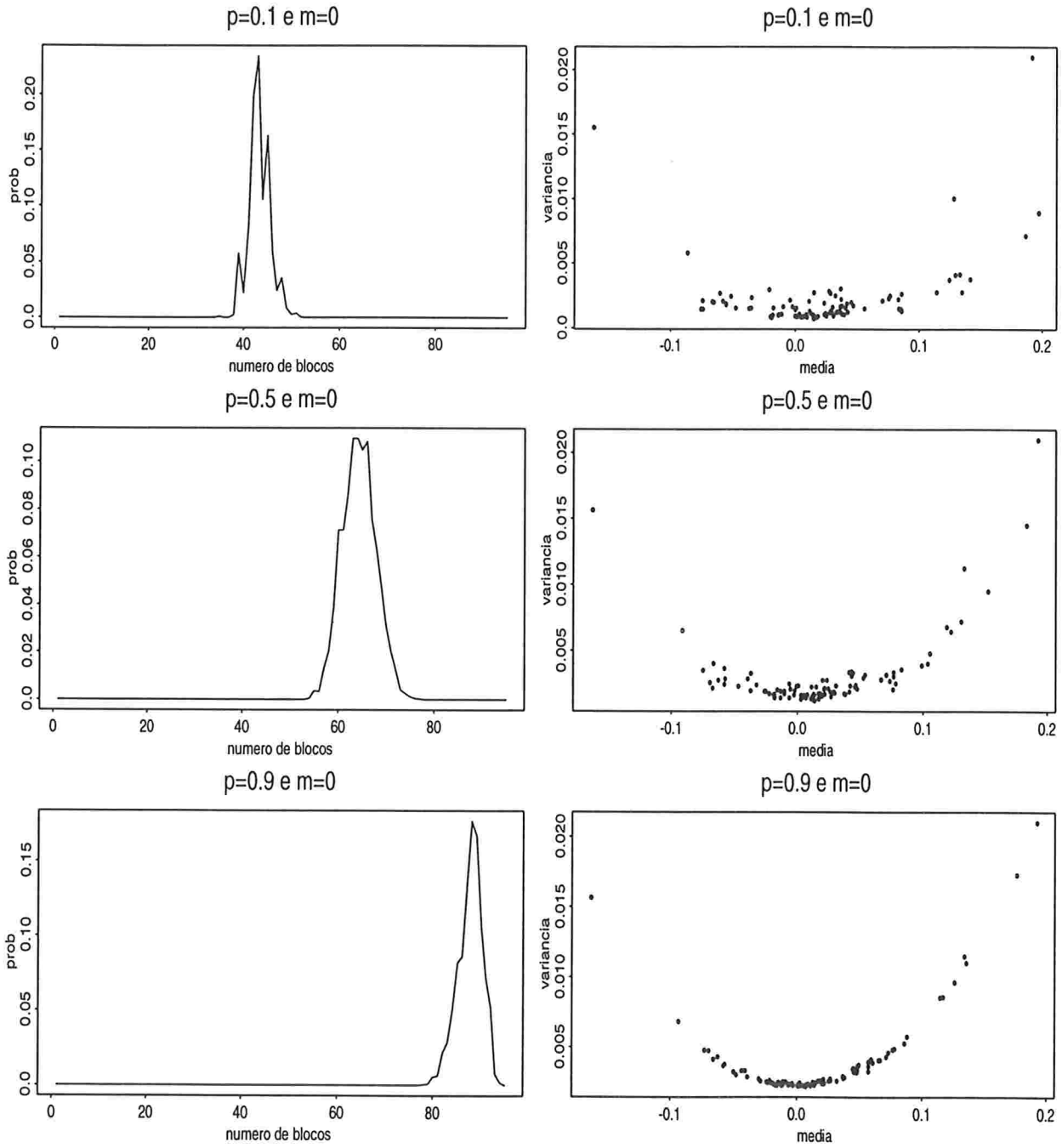
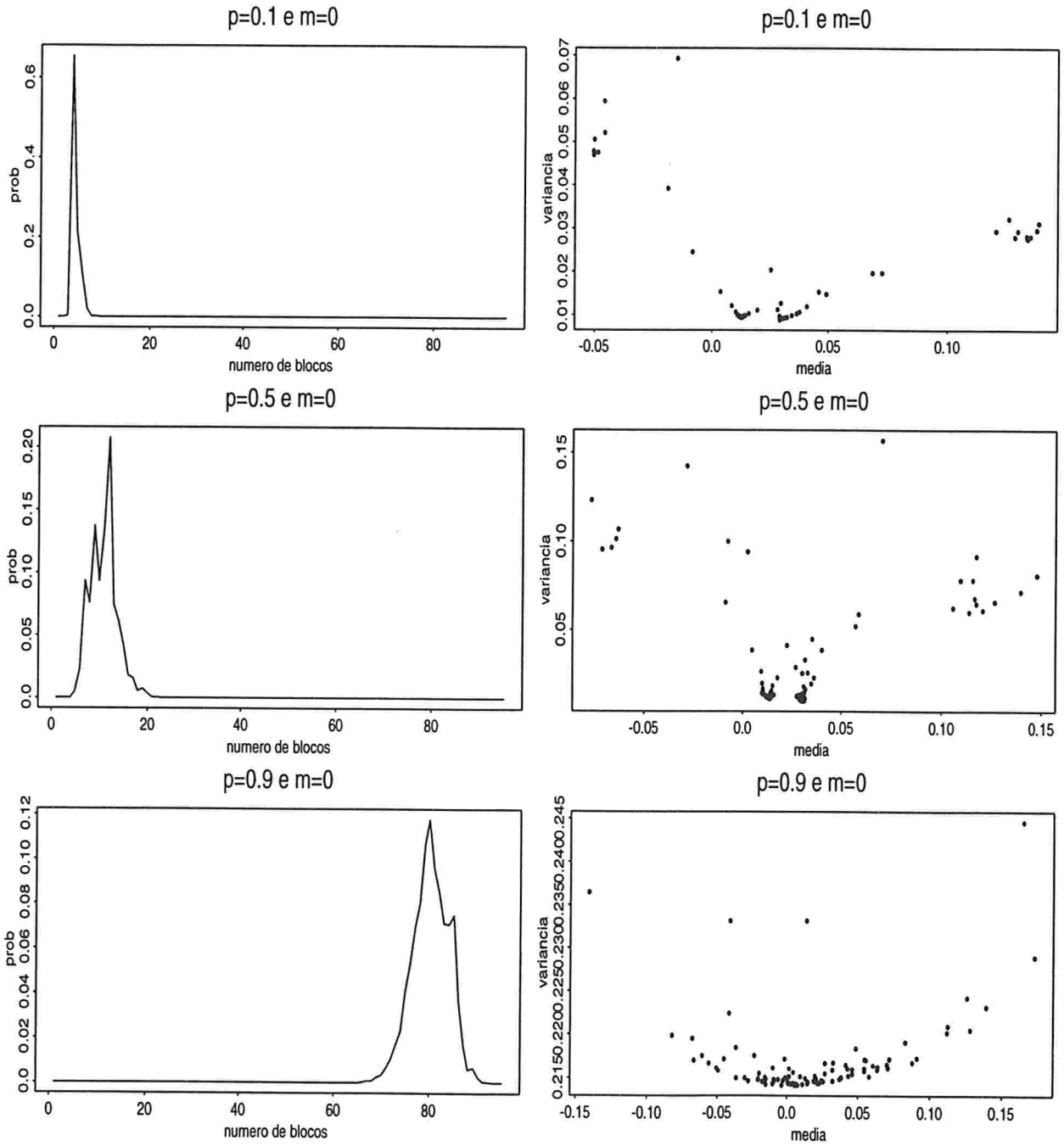


Figura 5.33: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão para μ e σ^2- $m = 0$, $a = 1$ e $b = 3$



Assumindo que $(\sigma^{ij})^2$ tem seu comportamento descrito pela distribuição $\mathcal{GI}(a/2, d/2)$ para todo i e j obtemos os seguintes resultados.

Figura 5.34: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão μ e σ^2- $m = 0$, $a = 0.01$ e $d = 3$

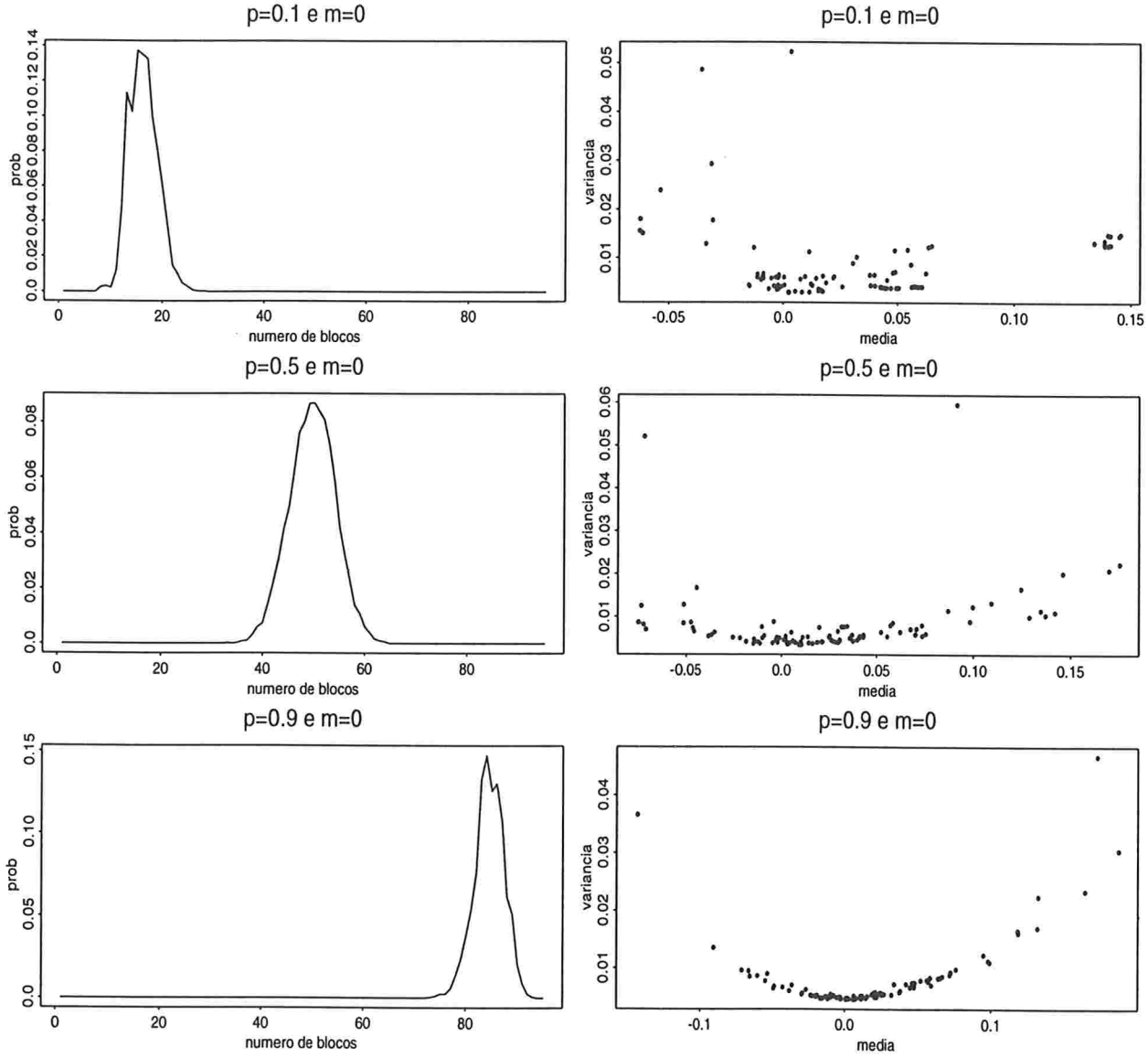
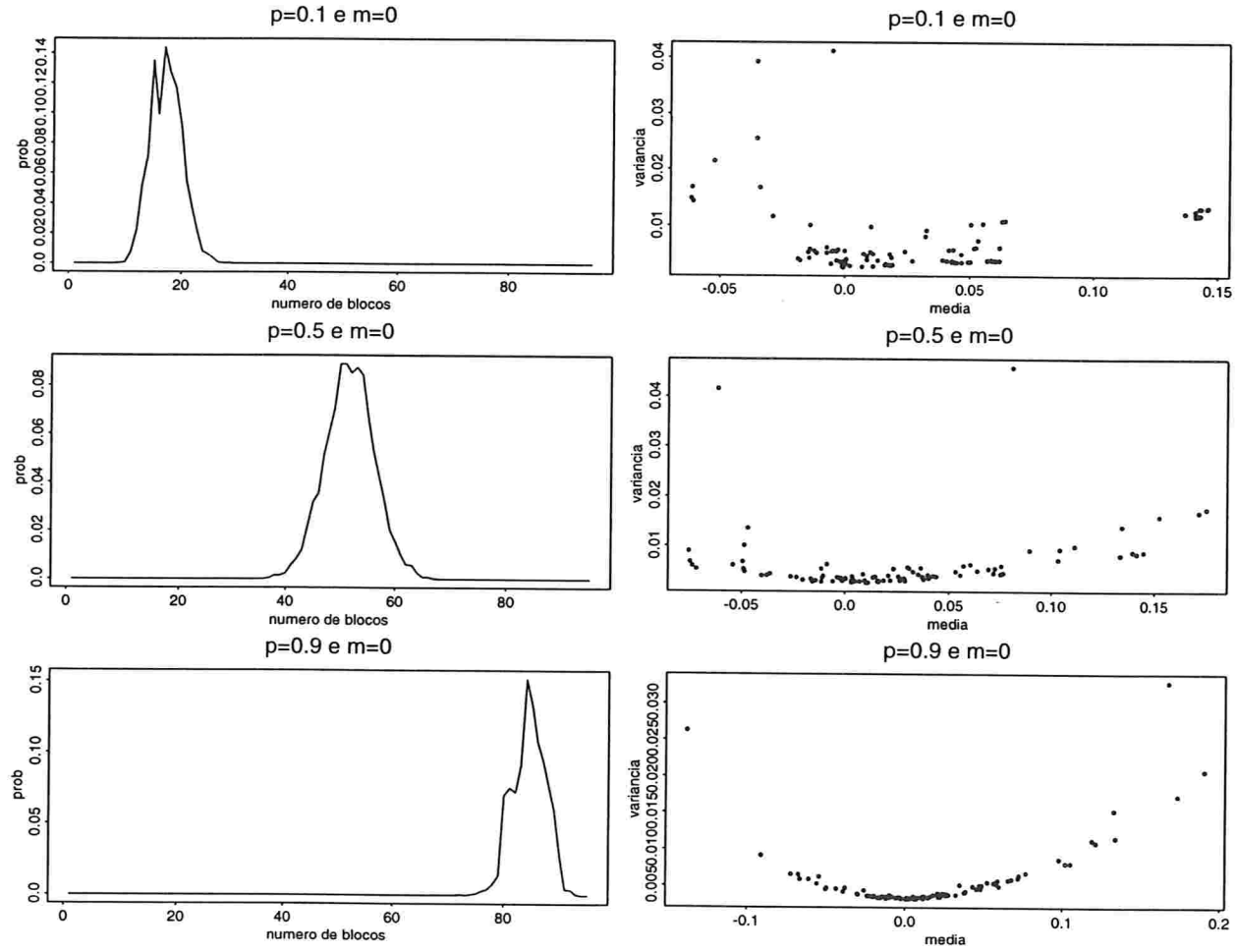


Figura 5.35: Distribuição *a posteriori* estimada para o número de blocos B e diagrama de dispersão μ e σ^2- $m = 0$, $a = 0.01$ e $d = 4$



Percebe-se, dos gráficos anteriores, que independentemente das distribuições *a priori* adotadas para descrescrever o comportamento de μ e σ^2, o número de blocos presentes na partição do conjunto de dados cresce a medida que o valor de p cresce. Também parece ser comum ocorrerem mudanças simultâneas no retorno esperado e na volatilidade. Isto pode ser percebido, por exemplo, no diagrama de dispersão da figura 5.30, quando $p = 0.1$ e $m = 0$. Neste exemplo, ocorrem poucas mudanças tanto no retorno médio quanto na volatilidade mas estas, em sua maioria, ocorrem simultaneamente, concordando com o que foi observado no respectivo gráfico das estimativas vistos na figura 5.24. Se consideramos o diagrama de dispersão da figura 5.32, referente a $p = 0.9$, notamos que, em geral, quando ocorrem mudanças na volatilidade também ocorrem mudanças no retorno esperado. Apesar

disto, existe um grande número de mudanças no retorno esperado que não é acompanhada por mudanças na volatilidade.

A seguir, apresentaremos a partição cuja ocorrência *a posteriori* é de maior probabilidade para cada um dos casos estudados aqui. Os gráficos a seguir contrastam a série de retornos com a seqüência de zeros e uns correspondente à partição gerada (ver expressão (4.38)). Um ponto de mudança ocorre quando um zero ocorre. Desta forma, o primeiro gráfico da figura abaixo corresponte à ocorrência da partição $\rho = \{0,\ 8,\ 95\}$.

Figura 5.36: Partição cuja ocorrência é a mais provável- $m = 0$, $a = 0.5$ e $b = 0.5$

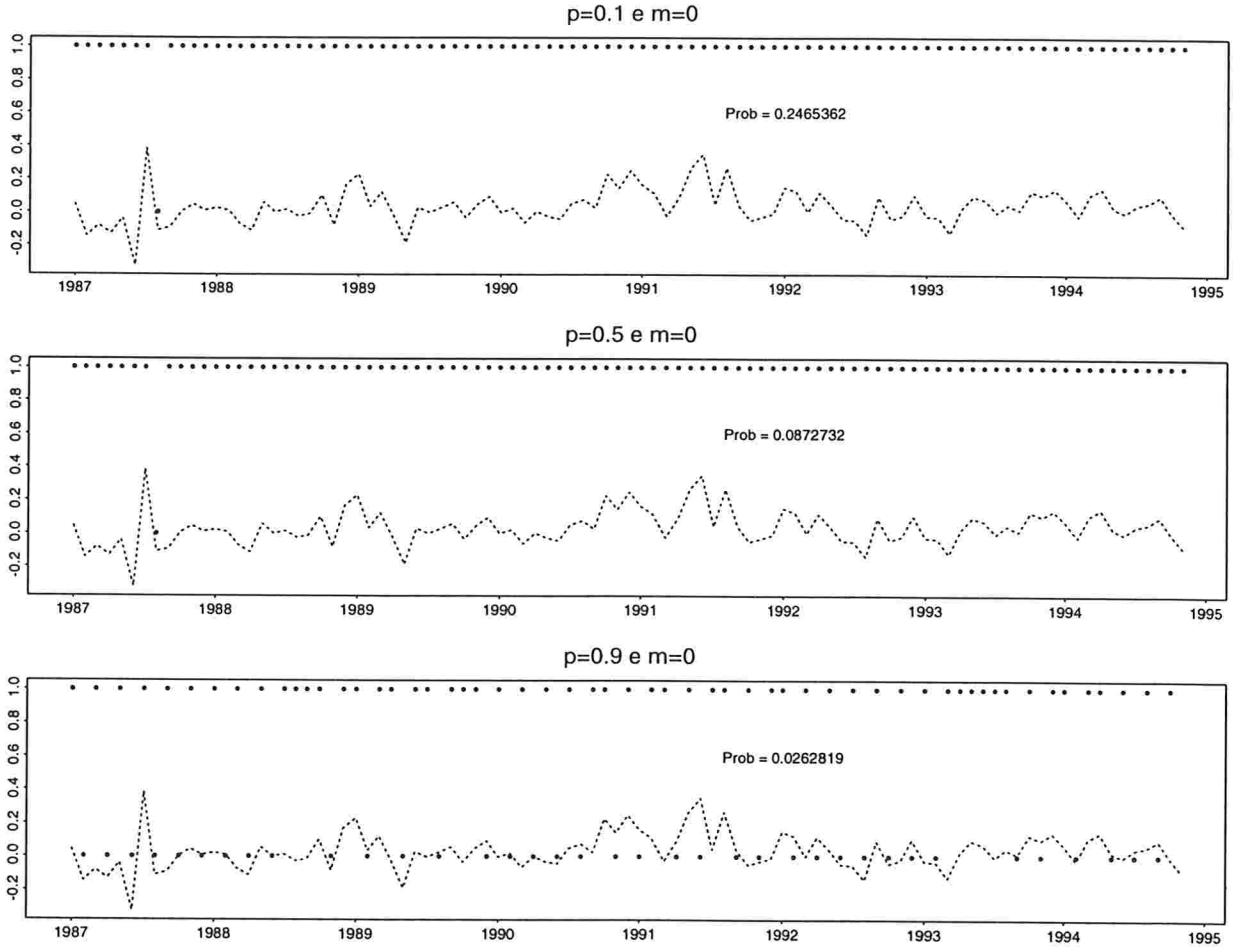


Figura 5.37: Partição cuja ocorrência é a mais provável- $m = 0.05$, $a = 0.5$ e $b = 0.5$

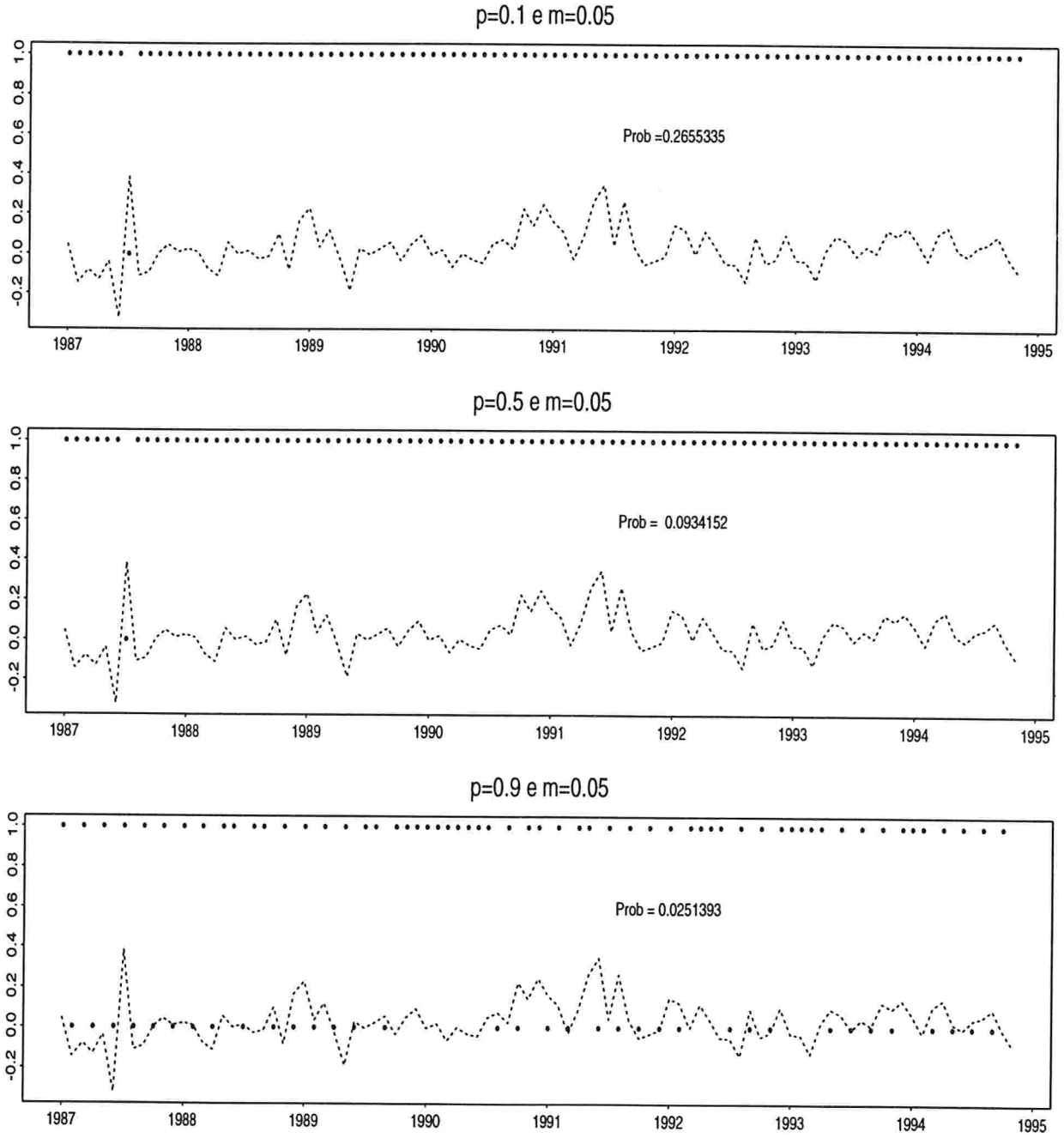


Figura 5.38: Partição cuja ocorrência é a mais provável- $m = 0$, $a = 0.01$ e $b = 3$

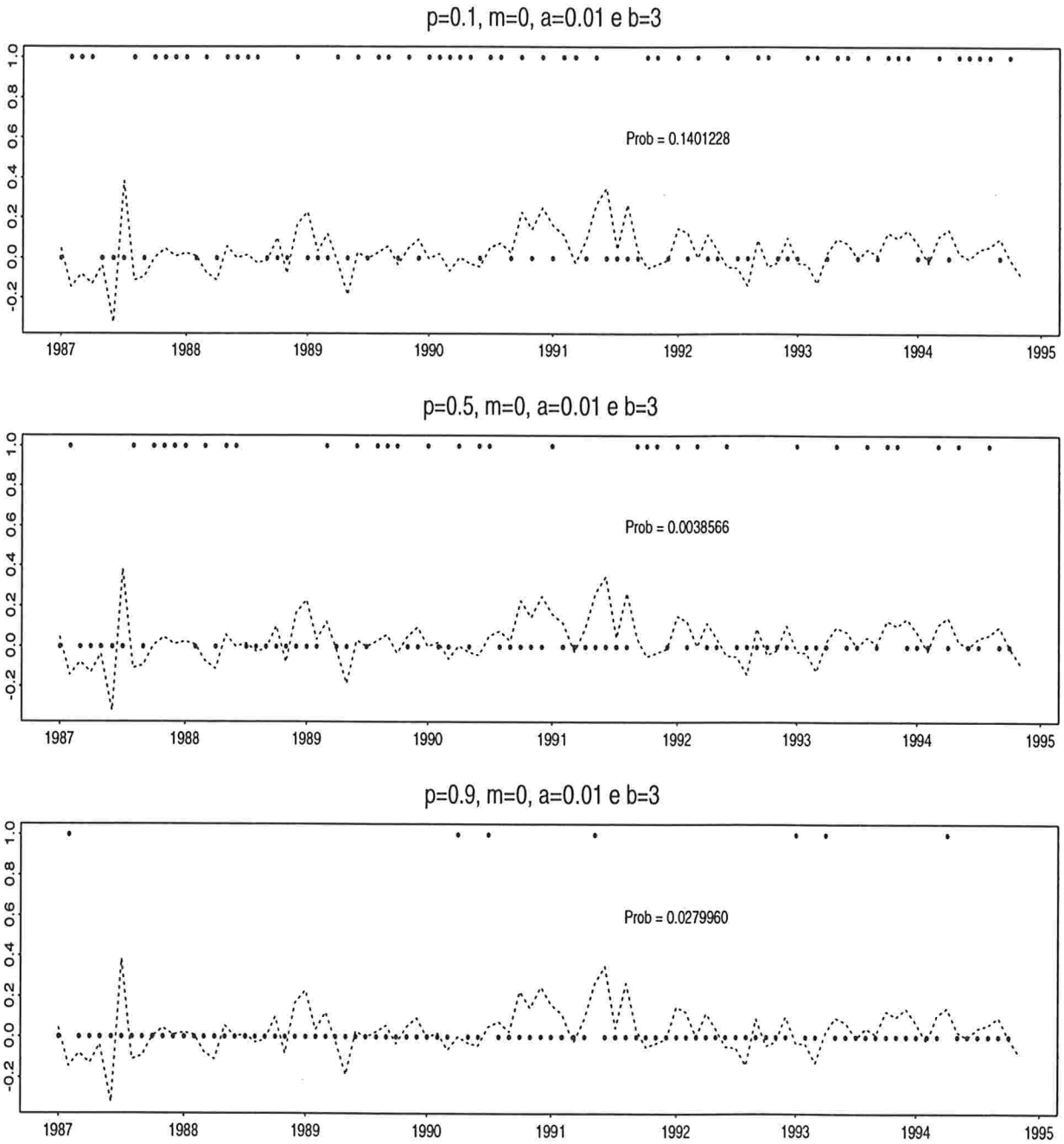
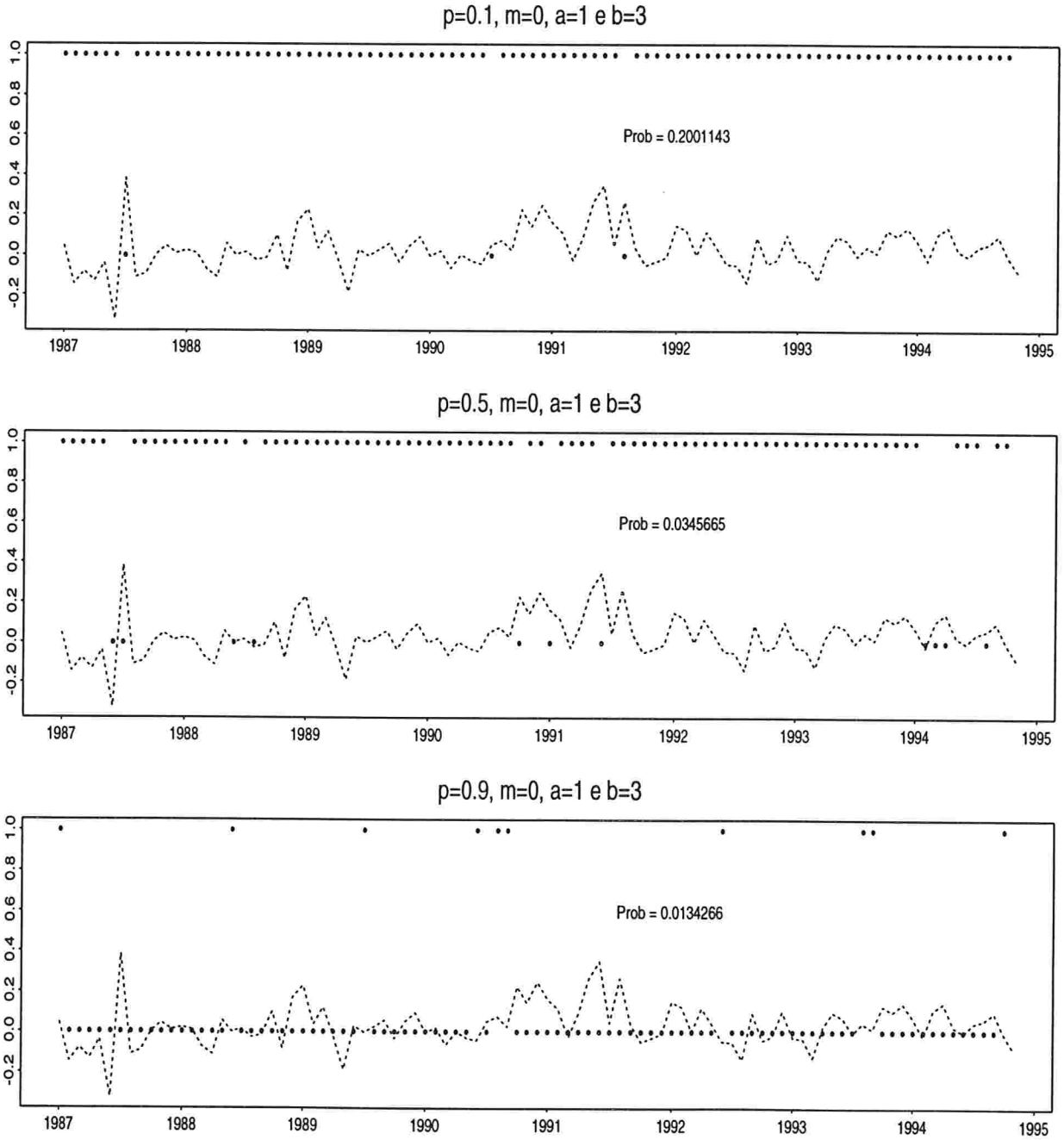


Figura 5.39: Partição cuja ocorrência é a mais provável- $m = 0$, $a = 1$ e $b = 3$



Assumindo que $(\sigma^{ij})^2$ tem seu comportamento descrito pela distribuição $\mathcal{GI}(a/2, d/2)$ para todo i e j as seguintes partições são obtidas.

Figura 5.40: Partição cuja ocorrência é a mais provável- $m = 0$, $a = 0.01$ e $d = 3$

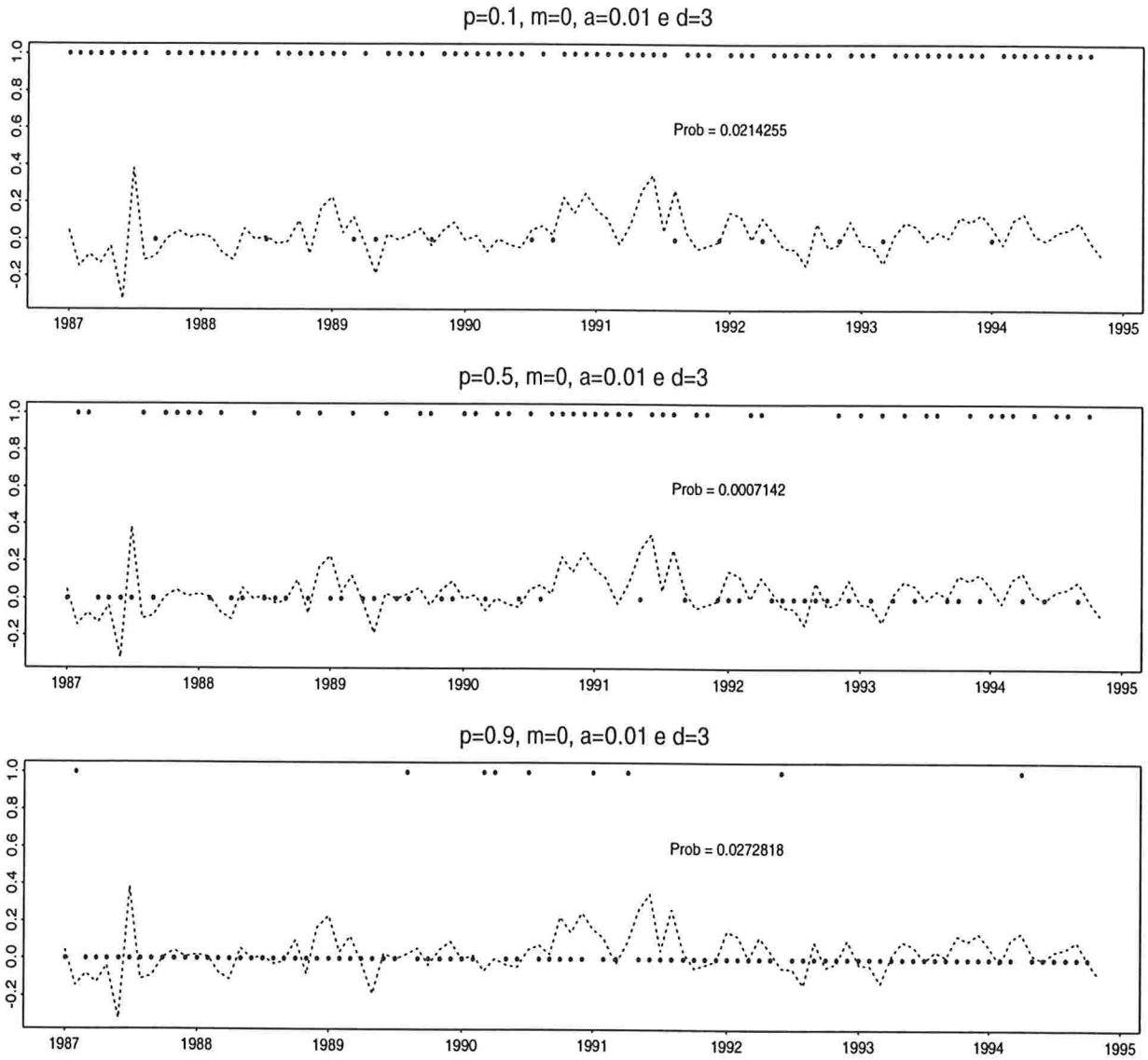
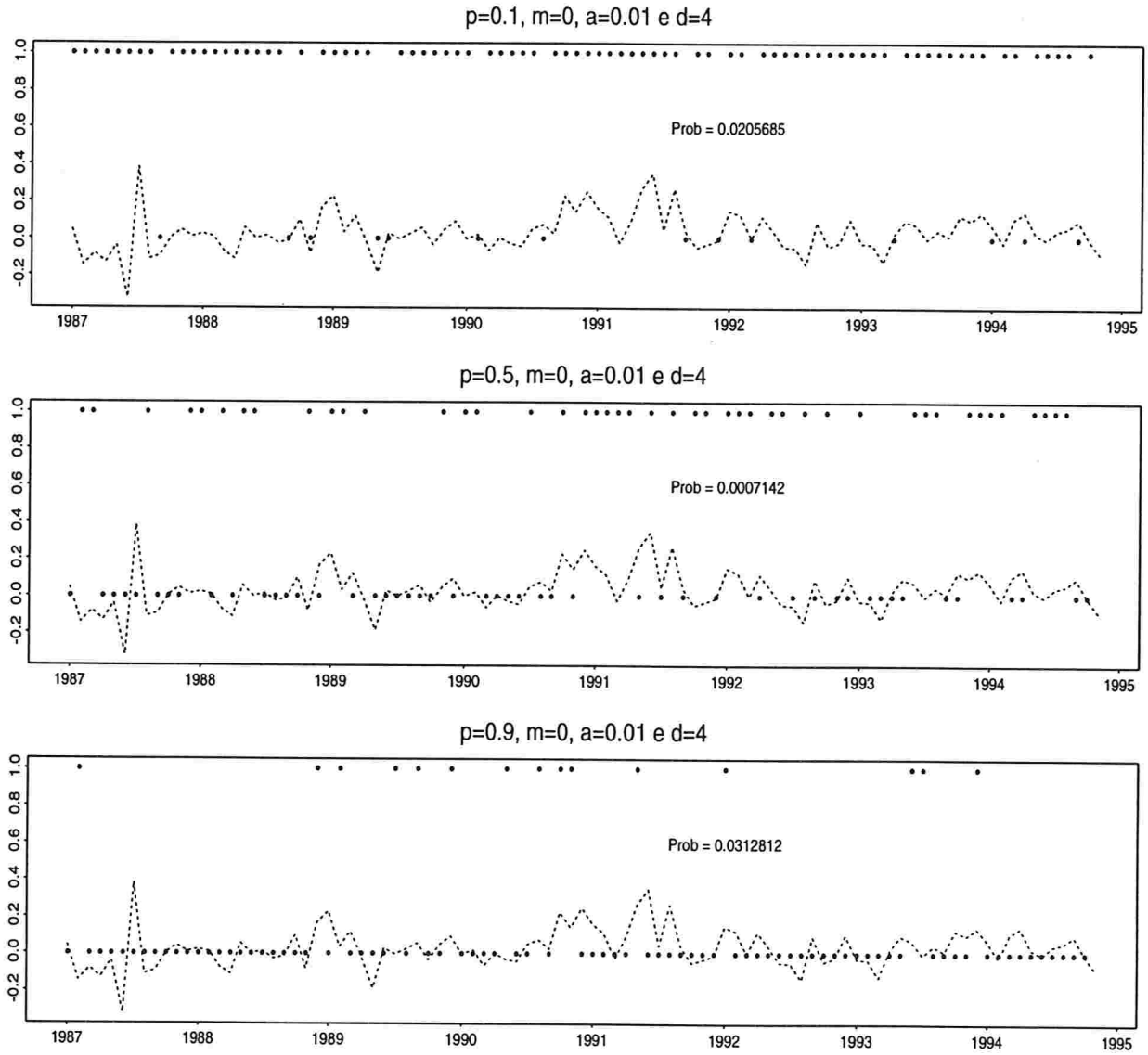


Figura 5.41: Partição cuja ocorrência é a mais provável- $m = 0$, $a = 0.01$ e $d = 4$

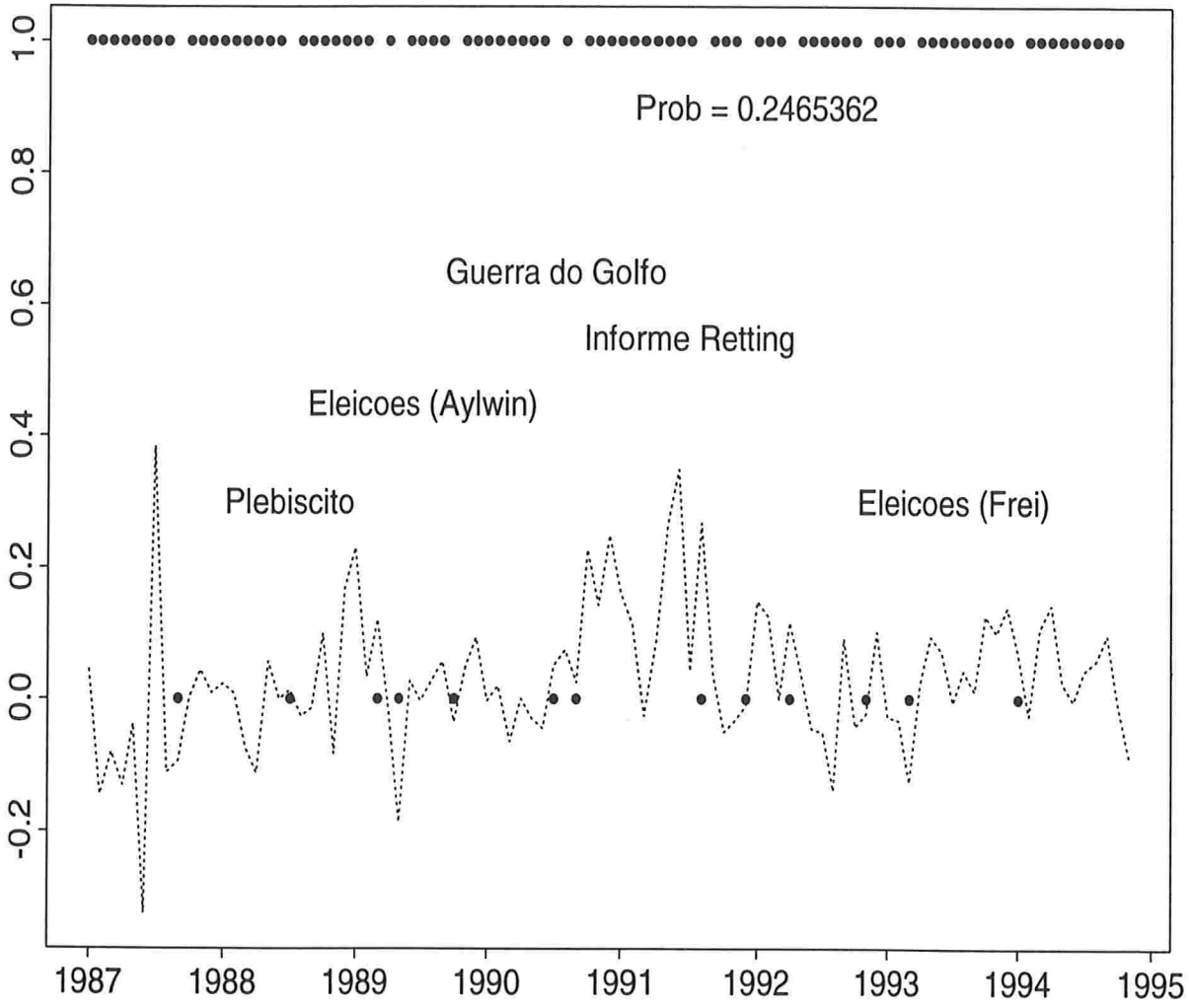


Diante das informações e fatos históricos relatados por Maeda(1996), e que brevemente apresentamos no final da seção 5.3, achamos que o melhor julgamento (dentre os poucos apresentados neste trabalho), que poderia ter sido feito por uma pessoa, sobre o comportamento da série Rentabilidade - ENDESA é aquele que traduz a opinião *a priori* da pessoa sobre μ^{ij} como sendo a distribuição $\mathcal{N}(0; (\sigma^{ij})^2)$, sobre $(\sigma^{ij})^2$ como sendo a distribuição $\mathcal{GI}(0.01/2;\ 3/2)$ e que aponta o valor de p como sendo 0.1.

Ao descrever sua opinião *a priori* sobre o comportamento da série ENDESA dessa forma,

a seguinte partição ocorrerá *a posteriori* com maior probabilidade

Figura 5.42: Pontos de Mudança e Fatos Históricos (Rentabilidade ENDESA - $m = 0$, $a = 0.01$, $d = 3$ e $p = 0.1$



Perceba que a partição obtida revela mudanças em pontos bastante estratégicos. Em 1987, a bolsa de Nova York vive uma de suas maiores crises, comparada, inclusive, à grande quebra de 1929. Esta crise refletiu mundialmente, provocando quedas nas bolsas de valores de todo mundo, inclusive na bolsa chilena. O ponto ocorrido em 1988 pode ser justificado pelo plebiscito realizado em outubro para descobrir sobre a conformidade ou não do povo

chileno com respeito a continuidade do governo do então presidente da república - General Augusto Pinochet.

A vitória do NÃO parecer ser um dos fatores provocadores dos pontos de mudança ocorridos no início do ano seguinte, ainda mais, se aliado a este fator considerarmos o fechamento da ecomonia norte-americana à importações de alguns produtos chilenos. A partir do plebiscito, o mercado financeiro experimenta uma alta na volatilidade a qual perdura até meados de 1989 (ver figura 5.28). No final de 1989, ocorre o primeiro processo eleitoral depois de anos de ditadura militar. Este marco de grande importância para a história Chilena, parece provocar um esfriamento do mercado (observa-se volatilidade baixa a partir do último trimestre de 1989 até final de 1990) gerando possivelmente o ponto de mudança percebido nesta época.

O período de ajuste do governo do presidente eleito Patricio Alwyn, a Guerra do Golfo, precedida pela invasão do Kuwait, e a pressão exercida pelo exército contra Alwyn por causa das ventilações sobre a publicação do Informe Rettig são os mais prováveis responsáveis pelo aquecimento do mercado observado entre o final de 1990 e meados de 1991 e pelos dois pontos de mudanças obsevados em 1990. Neste período, observa-se a maior rentabilidade da série ENDESA entre janeiro de 1987 e dezembro de 1994.

No início de 1991, é divulgado o Informe Rettig o qual divulga o número de mortos e desaparecidos durante o período militar. Este informe provocou bastante insegurança sobre o futuro da democracia chilena. Também neste ano, ocorre uma queda mundial no preço do petróleo. Os dois pontos de mudanças ocorridos neste ano podem ter sido conseqüência destes fatores. A partir do final de 1991, a volatilidade no mercado começa a apresentar uma tendência ao decaimento. Os retornos também são mais baixos que os alcançados no período anterior e permanecem assim até o início de 1993.

O ano de 1992 foi marcado por uma política monetária expansiva e pela queda do dólar nos mercados internacionais e 1993 foi marcado pela dininuição no volume das exportações chilenas devido ao processo recessivo encarado pelos países desenvolvidos. Estes fatos podem ter sido os causadores dos pontos de mudanças ocorridos em 1992 e 1993, os quais parecem

mais óbvios nos retornos esperados que na volatilidade do mercado.

Considerações Finais

Atualizar probabilidades quando imprevistos ocorrem é um tema cujo estudo ainda não foi concluído. Apesar dos vários mecanismos de atualização de probabilidades existentes, nenhum deles responde completamente a questão de como atualizar probabilidades em todas as situações que podemos confrontar. A Regra de Jeffrey, por exemplo, apesar de responder a esta questão em situações mais gerais que o Condicionamento Bayesiano, é completamente inerte quando a informação recebida modifica o espaço amostral, ou quando leva Você a modificar suas opiniões sobre eventos do espaço amostral que não pertencem a uma partição deste e mesmo quando, sob esta informação, a J-condição não puder ser julgada válida. Vimos nos capítulos 1 e 2 exemplos que ilustram algumas destas situações.

Por exemplo, a ocorrência de um *Bayesianus Genuinus* no exemplo de estimação do número de espécies no deserto do Saara, visto no capítulo 1, não é respondido por qualquer proposta apresentada aqui. Também não consideramos razoável a proposta feita por Zabell(1992) a qual inicia com um espaço amostral grande o suficente para incluir tudo o que é possível ocorrer. O que é tudo quando levamos em conta a possibilidade de ocorrer algo completamente novo? Este é um problema ainda em aberto.

Também é digno de mais pesquisas a procura de alguma definição de coerência que norteie a escolha de procedimentos para atualizar-se probabilidades. Ou, quem sabe, provar a inexistência de qualquer definição de coerência temporal que torne normativo este processo de atualização.

Pode-se pensar também, dentro da linha mais fundamental, em um procedimento para

tomada de decisões diante do recebimento de informações imprevistas. Como se comporta a função de utilidade nestes casos? Qual é o similar do Princípio da Maximização da Utilidade Esperada se suas probabilidades são atualizadas via Regra de Jeffrey?

A caracterização preditivista de modelos $O(\mathcal{M})$-invariantes também não foi completamente explorada. Se condições sobre as esperanças condicionais, distintas das consideradas no capítulo 3, forem feitas certamente caracterizaremos modelos outros que não o modelo t. É ainda de nosso interesse investigar as condições necessárias para caracterizar outros modelos $O(\mathcal{M})$-invariantes como, por exemplo, para o modelo exponencial duplo multivariado. Ainda nesta linha de modelagem buscamos uma generalização do resultado apresentado na proposição 2.5.2. Como está apresentado, este resultado não permite, por exemplo, identificar o modelo quando a representação é uma mistura de normais com média e variância desconhecidas.

Devido à complexidade dos aspectos computacionais envolvidos no capítulo 4, poucas contribuições foram feitas nesta parte do trabalho e ainda temos muito o que fazer. Acreditamos que o parâmetro p envolvidos nas coesões usadas na aplicação do capítulo 5 não são fixos e devem ter seu comportamento descrito por uma distribuição de probabilidade. Ao fazer isso a estrutura do Modelo Partição Produto é perdida dificultando um pouco mais as inferências. Também é interessante inferir sobre possíveis mudanças na forma funcional do modelo e não somente nos parâmetros que o indexa, como fizemos aqui. É possível que o comportamento nos diferentes blocos possam ser descritos por modelos distintos e detectar esta mudança pode ser bastante importante. Será que no mercado chileno, as hipóteses da teoria de mercados eficientes são de fato razoáveis a partir do final de 90?

Outro aspecto também de interesse é desenvolver um procedimento para fazer previsões sobre valores futuros de observáveis levando em conta a ocorrência de pontos de mudança na seqüência de observações já feitas e, se possível, respeitando a estrutura do MPP.

Apêndice Um

Neste apêndice mostramos algumas tabelas contendo valores gerados da distribuição t dada na expressão (5.5) para diversos valores de a, b_1 e b_2. Lembramos que a deve pertencer ao intervalo $(0; 0,5)$, b_1 pode ser qualquer real e b_2 é qualquer real positivo que satisfaça a relação $b_2 > \frac{b_1^2}{1-2a}$.

$a = 0.0001$ e $b_1 = 0$

$b_2 = 0.1$	$b_2 = 10$	$b_2 = 100$
0.094	−2.628	0.0861
0.238	1.226	−0.3817
−0.509	0.330	−10.1501
0.447	2.658	−1.3221
0.211	0.958	−3.5971

$a = 0.49$ e $b_1 = 0$

$b_2 = 0.1$	$b_2 = 10$	$b_2 = 100$
0.8095	179.1990	−395.0072
1.77525	−270.4653	70.6545
18.4738	−91.8287	−0.0743
0.6362	−157.1550	253.1096
−0.1919	176.5262	−4.2801

$a = 0.2$ e $b_1 = 0$

$b_2 = 0.1$	$b_2 = 10$	$b_2 = 100$
−0.4397	−1.1168	−14.9614
−0.8842	2.0312	2.5795
0.4866	9.7834	−4.4579
0.6193	3.9720	12.4637
−0.2081	0.3587	3.8529

$a = 0.01$ e $b_1 = 1$

$b_2 = 2$	$b_2 = 103$	$b_2 = 10205$
1.5872	25.7604	−56.2798
1.1007	31.0410	87.8369
1.3628	−14.6032	−229.5013
−0.4099	13.7864	125.2311
−1.5404	16.5268	−107.3629

$a = 0.01$ e $b_1 = 10$

$b_2 = 103$	$b_2 = 1003$	$b_2 = 10205$
4.8337	−0.1690	41.0003
12.1857	19.4260	114.3184
11.3153	54.6330	20.8881
7.7617	62.3210	42.2284
11.3259	27.771	14.9318

$a = 0.01$ e $b_1 = 100$

$b_2 = 10205$	$b_2 = 50000$	$b_2 = 1000000$
−97.8070	547.1301	1511.6486
−85.5320	83.3719	1521.9954
−111.2690	−72.8365	2450.7761
−97.6280	235.9428	−1225.3603
−130.2690	−46.2784	309.2387

$a = 0.2$ e $b_1 = 1$

$b_2 = 1.7$	$b_2 = 167$	$b_2 = 16667$
−0.7720	−3.6080	−116.7338
6.0390	−4.4869	−48.3030
3.0560	−31.6403	−330.3909
4.8140	−11.1442	59.3423
2.9330	12.4663	−215.1994

$a = 0.2$ e $b_1 = 10$

$b_2 = 167$	$b_2 = 200$	$b_2 = 16667$
14.0920	15.9000	348.32288
13.44878	9.2160	19.5378
−0.2590	22.2276	−165.4655
24.4900	−6.5720	85.6929
18.1190	7.2160	−540.6685

$a = 0.2$ e $b_1 = 100$

$b_2 = 16667$	$b_2 = 50000$	$b_2 = 100000$
−24.5016	146.0665	732.5344
144.8118	166.8789	−77.3363
45.9510	449.6085	−36.7773
519.7184	770.8943	570.0135
41.6798	6.1307	−69.1908

$a = 0.49$ e $b_1 = 1$

$b_2 = 51$	$b_2 = 5001$	$b_2 = 500001$
138.0143	−1947.088	18364.996
216.7260	−283.9620	10282.691
121.04245	1089.1130	10630.9360
67.30051	24.483	9917.9670
−41.8420	−277.9740	16720.8110

$a = 0.49$ e $b_1 = 10$

$b_2 = 5001$	$b_2 = 10001$	$b_2 = 500001$
6831.66	−385.2320	18256.2320
−3618.3200	4471.0380	65095.1460
410.4710	−946.4880	27376.7580
3699.1020	−2111.2560	5592.379
1724.9930	−1203.316	24359.7435

$a = 0.49$ e $b_1 = 100$

$b_2 = 500001$	$b_2 = 600000$	$b_2 = 1000000$
−6961.2000	23475.7420	29764.5040
4755.8730	42665.7500	8330.9040
23158.795	42959.4000	25284.6380
5165.479	2192.0750	9371.1410
23000.0530	11135.4090	22317.7260

OBS: Os valores de b_2 encontrados na primeira coluna de cada tabela são valores muito próximos no menor valor possível para b_2 com as particulares escolhas de a e b_1 feitas.

Apêndice Dois

A seguir apresentamos o algoritmo utilizado para fazer as previsões exibidas na seção 5.4 do capítulo 5. Este algoritmo foi implementado em FORTRAN e pode ser compilado utilizando os compiladores Fortran 77 e subsequentes.

Para executar o programa é necessário fornecer os arquivos de dados com nome X.dat e os arquivos contendo os parâmetros a, b_1 e b_2 em cada instante de tempo(mesmo que não haja mudanças nestes). Estes arquivos devem ser nomeados, respectivamente, por A.dat, B1.dat e B2.dat.

```
      Program nuevo1
      implicit none

-------------------------
 Definicao das variaveis
-------------------------

      integer i, max,k,m
      parameter (max=2000)

      double precision X(max),a(max), b1(max), b2(max)
      double precision ecj(max),ecb(max),sj(max),sqj(max),qsj(max)
      double precision sb(max),sqb(max),qsb(max)
      double precision db1,rb1
      double precision vcb(max)
      double precision EMiB(max), EsigmaB(max)
      double precision EMij(max),Esigmaj(max)
      double precision d,r,vcj(max)
```

```
        logical aux1,aux2,aux3

----------------------------
Leitura dos arquivos de dados
----------------------------


        write(*,*) '-------------------------------------------------- '
        write(*,*) 'Lembre-se que os dados do arquivo A.dat devem'
        write(*,*) 'pertencer ao intervalo (0;0.5)'
        write(*,*) ' '
        write(*,*) 'Os dados do Arquivo B1.dat podem  ser quaisquer  reais'
        write(*,*) ' '
        write(*,*) 'E os dados do arquivo B2.dat devem satisfazer a relacao'
        write(*,*) ' b2>b1*b1/(1-2*a)'
        write(*,*) '-------------------------------------------------- '
        write(*,*) ' '
        write(*,*) 'O maximo permitido para o vector e  max'
        write(*,*) 'se o arquivo e maior que max modifique a constante'
        write(*,*) ' "max" do programa'
        write(*,*) ' '
        write(*,*) '-------------------------------------------------- '
         write(*,*) 'Tamanho  dos arquivos ou numero de amostras:'
         read(*,*) m
open(unit=1,file='X.dat',status='old',err=1000)
read(1,*,Err= 1100) (X(i),i=1,m)
close(1)
        open(unit=1,file='A.dat',status='old',err=1000)
         read(1,*) (a(i), i=1,m)
         close(1)
```

```
      open(unit=1,file='B1.dat',status='old',err=1000)
      read(1,*) (b1(i), i=1,m)
      close(1)
      open(unit=1,file='B2.dat',status='old',err=1000)
      read(1,*) (b2(i), i=1,m)
      close(1)

------------------
Inicio do programa
------------------

      a(0)=0.0
      b1(0)=0.0
      b2(0)=0.0
      sb(0)=0.0
      qsb(0)=0.0
      sqb(0)=0.0
      do 11 i=1,m
         ecj(i)=0.0
         ecb(i)=0.0
         sqb(i)=0.0
         sqj(i)=0.0
         sj(i)=0.0
         qsj(i)=0.0
         sb(i)=0.0
         qsb(i)=0.0
         EMiB(i)=0.0
         EMij(i)=0.0
         EsigmaB(i)=0.0
         Esigmaj(i)=0.0
```

```
            vcj(i)=0.0
            vcb(i)=0.0
 11         continue
            open(unit=1,file='Pcambio.dt',status='new',err=1001)
         do 15 k=1,m
            db1=0.0
            rb1=0.0

-------------------------------------------------------
Previsoes utilizando apenas o Condicionamento Bayesiano
-------------------------------------------------------

            db1=1+a(1)*(k-3)
            rb1=a(1)/db1
            ecb(k)=rb1*sb(k-1)+b1(1)/db1
            vcb(k)=b2(1)/a(1)-(b1(1)**2/(a(1)*db1))+sqb(k-1)
            vcb(k)=rb1*(vcb(k)-rb1*qsb(k-1)-2*b1(1)*sb(k-1)/db1)
            sb(k)=sb(k-1)+X(k)
            sqb(k)=sqb(k-1)+X(k)**2
            qsb(k)=sb(k)**2
            EMiB(k)=(a(1)*sb(k-1)+b1(1))/db1
            EsigmaB(k)=b2(1)/a(1)+sqb(k-1)
            EsigmaB(k)=EsigmaB(k)-((a(1)*sb(k-1)+b1(1))**2)/(a(1)*db1)
            EsigmaB(k)=(EsigmaB(k)*a(1))/(a(1)*(k-3)+1)


-----------------------------------------------------
Previsoes utilizando a Regra Preditivista de Jeffrey
-----------------------------------------------------
```

```
   aux1=a(k).eq.a(k-1)
   aux2=b1(k).eq.b1(k-1)
   aux3=b2(k).eq.b2(k-1)
   if ((aux1.and.aux2).and.aux3) then
      d=0.0
      r=0.0
      d=1+a(k)*(k-3)
      r=a(k)/d
      ecj(k)=r*sj(k-1)+b1(k)/d
      vcj(k)=b2(k)/a(k)-b1(k)**2/(a(k)*d)+sqj(k-1)-r*qsj(k-1)
      vcj(k)=r*(vcj(k)-2*b1(k)*sj(k-1)/d)
      EMij(k)=(a(k)*sj(k-1)+b1(k))/d
     Esigmaj(k)=b2(k)/a(k)+sqj(k-1)
     Esigmaj(k)=Esigmaj(k)-(a(k)*sj(k-1)+b1(k))**2/(a(k)*d)
      Esigmaj(k)=a(k)*Esigmaj(k)/(a(k)*(k-3)+1)
      sj(k)=sj(k-1)+X(k)
      sqj(k)=sqj(k-1)+X(k)**2
      qsj(k)=sj(k)**2
else
   write(1,*) k
   ecj(k)=b1(k)/(1-2*a(k))
   vcj(k)=(b2(k)*(1-2*a(k))-b1(k)**2)/(1-2*a(k))**2
   EMij(k)=b1(k)/(1-2*a(k))
Esigmaj(k)=(b2(k)*(1-2*a(k))-b1(k)**2)/((1-a(k))*(1-2*a(k)))
   sj(k)=X(k)
   sqj(k)=X(k)**2
   qsj(k)=sj(k)**2
end if
```

```
-----------------------------------------------------------
Arquivo contendo as previsoes via Condicionamento Bayesiano
-----------------------------------------------------------

 15       continue
          close(1)
         open(unit=2,file='Pbayes.dt',status='new',err=1001)
         do 20 i=1,m
            write(2,*)  ecb(i)

-----------------------------------------------------------------
Arquivo contendo as previsoes via Regra Preditivista de Jeffrey
-----------------------------------------------------------------


 20      continue
         close(2)
         open(unit=3,file='Pjeffrey.dt',status='new',err=1001)
         do 25 i=1,m
            write(3,*) ecj(i)

-----------------------------------------------------------
Arquivo contendo as variancias a posteriori para a proxima
observacao - Condicionamento Bayesiano
-----------------------------------------------------------

 25      continue
         close(3)
         open(unit=4,file='Ebayes.dt',status='new',err=1001)
         do 30 i=1,m
```

```
            write(4,*) vcb(i)

-------------------------------------------------------------
Arquivo contendo as variancias a posteriori para a proxima
observacao - Regra Preditivista de Jeffrey
-------------------------------------------------------------

 30       continue
          close(4)
          open(unit=5,file='Ejeffrey.dt',status='new',err=1001)
          do 35 i=1,m
             write(5,*) vcj(i)

-----------------------------------------------------
Arquivo contendo as estimativas a posteriori para mi-
Condicionamento bayesiano
-----------------------------------------------------


 35       continue
          close(5)
          open(unit=1,file='Mibayes.dt',status='new',err=1001)
          do 40 i=1,m
             write(1,*) EMiB(i)

-------------------------------------------------------
Arquivo contendo as estimativas a posteriori para sigma-
Condicionamento Bayesiano
-------------------------------------------------------
```

```
 40      continue
         close(1)
         open(unit=1,file='Sigbayes.dt',status='new',err=1001)
         do 45 i=1,m
            write(1,*) EsigmaB(i)

-------------------------------------------------------
Arquivo contendo as estimativas a posteriori para mi-
Regra preditivista de Jeffrey
-------------------------------------------------------

 45      continue
         close(1)
         open(unit=1,file='Mijef.dt',status='new',err=1001)
         do 50 i=1,m
            write(1,*) EMij(i)

--------------------------------------------------------
Arquivo contendo as estimativas a posteriori para sigma-
Regra Preditivista de Jeffrey
--------------------------------------------------------


 50      continue
         close(1)
         open(unit=1,file='Sigjef.dt',status='new',err=1001)
         do 55 i=1,m
            write(1,*) Esigmaj(i)
```

```
55      continue
        close(1)

        write(*,*) ''
        write(*,*) '------------------------------------------------ '
        write(*,*) 'os resultados podem ser encontrados nos arquivos:'
        write(*,*) 'Pbayes.dt,Pjeffrey.dt,Ebayes.dt,Ejeffrey.dt'
        write(*,*) 'Sigbayes.dt,Sigjef.dt,Mijef.dt,Mibayes.dt'
        write(*,*) 'os pontos onde ocorrem mudancas nos parametros '
        write(*,*) ' estao no arquivo  "Pcambio.dt"'
1000    write(*,*) ''
1001    write(*,*) ''
1100    write(*,*) ''
        end
```

Apêndice Três

O algoritmo que segue foi implementado em PERL e fornece as estimativas *a posteriori* para os parâmetros μ e σ^2 de uma distribuição normal, a estimativa *a posteriori* para o número de blocos B e as dez partições mais prováveis de terem ocorrido. A técnica usada para estimar as distribuições *a posteriori* para B e ρ foi *Gibbs Sampler*.

Para executar o programa é necessário fornecer o conjuto de dados, os valores de p e τ e o número de seqüências $(U_1, \ldots, U_n)$ a serem geradas, LIM. Além do conjunto com n dados, deve estar disponível para leitura o arquivo "cadena0" contendo a seqüência de U_i's inicial.

O programa gera os arquivos COHES, contendo a matriz de coesões *a posteriori*, RELEV, onde pode-se encontrar a matriz de relevância *a posteriori*, CADENAS, formado pelas médias das seqüências geradas, MEDIAS, constituído pelas estimativas *a posteriori* para μ e σ^2 e PEB, o qual armazena a distribuição *a posteriori* para B e as dez cadeias mais freqüentes com suas respectivas freqüências relativas.

```
   #!/usr/bin/perl

   #programa em perl

   @hora = localtime(time);
   print "$hora[2]:$hora[1]:$hora[0]\n";

# abrir o arquivo de  dados

   open DATA,$ARGV[0] or die "Nao se pode abrir o arquivo de dados";

# leia os dados
```

```
    $i = 1;
    while (<DATA>) {
     $aux = substr($_ , 0, -1);
     if(ord($aux)!=0){ $xk[$i++] = $aux;};
    };
    $p = $ARGV[1];
    $t = $ARGV[2];
    $lim = $ARGV[3];
    $n = $i-1;
    print "leia $n numeros... burp! com p: $p , t: $t e lim: $lim\n";
    close DATA;
    $pi = 3.141592653589793;

#comeco dos calculos

    # calculo dos m_ij^*
    @mPost = &calculaM(\@xk,$t);

    # calculo do factor que inclue as gamas
    @gamas = &calculaG($t);

    # calculo das coesoes c_ij^*
    @c = &calculaC(\@xk,\@gamas,$t,$p);

    # calculo de lambda0
    @lambda0 = &calculaL0(\@c);

    # calculo de LambdaN
    @lambdaN = &calculaLN(\@c);
```

```
# calculo das relevancias r_ij^*
@relev = &calculaR(\@c,\@lambda0,\@lambdaN);

# calculo dos a_ij^*
@aPost = &calculaA(\@xk,$t);

#calculo dos d_ij^*
@dPost = &calculaD;

# calculo das medias a posteriori
@medias = &calculaMedias(\@mPost,\@relev,\@aPost,\@dPost);

# calculo das cadeias
@cadenas = &calculaCadenas(\@c,$lim);

open fMedias,">cadenas";
select fMedias;
print "Medias de las cadenas\n";
for($i = 1;$i <= $lim;$i++) {
 print $cadenas[$i][2]/$n,"\n";
};
close fMedias;

select STDOUT;

# calculos com as cadeias
print "Valor de Estabilizacao = ";
read(STDIN,$est,length($lim));
@cuantas = cuentaCuantas(\@cadenas);
```

```
open C,">cohes";
select C;
print "Matriz de cohesiones\n";
for($i = 0;$i < $n;$i++) {
   for($j = $i+1;$j <= $n;$j++) {
      print $c[$i][$j],"\t";
   };
   print "\n";
};
close C;

open C,">relev";
select C;
print "Matriz de relevancias\n";
for($i = 0;$i < $n;$i++) {
   for($j = $i+1;$j <= $n;$j++) {
      print $relev[$i][$j],"\t";
   };
   print "\n";
};
close C;

open C,">medias";
select C;
print "Medias a Posteriori\n";
for($i = 0;$i <= $n;$i++) {
print $medias[$i][1],"\t",$medias[$i][2],"\n";
};
close C;
```

```
   select STDOUT;
   @hora = localtime(time);
   print "$hora[2]:$hora[1]:$hora[0]\n";



##################################################################
#                                                                #
#  Esta e a secao das funcoes  utilizadas pelo programa          #
#                                                                #
##################################################################


# calculaM devolve o valor dos m_ij^*

sub calculaM {
   my ($pv,$t) = @_;
   local @vv = @$pv;

   for($i = 0;$i < $n;$i++) {
      for($j = $i+1;$j <= $n;$j++) {
         $aux = (1 + ($j-$i)*$t);
         $mp[$i][$j] = $t*suma(\@vv,$i+1,$j)/$aux;
         $mp[$i][$j] = $mp[$i][$j]+funcion($i,$j,1)/$aux;
         #print "m$i $j:", $mp[$i][$j],"\n";
      };
   };
   return(@mp);
};
```

```
# calcula a soma dos dados (pvv) entre os valores i e j
sub suma {
   my ($pv,$i,$j) = @_;
   my $suma = 0;
   local @vv = @$pv;

   for($x = $i;$x <= $j;$x++) {
      $suma += $vv[$x];
   };
   return($suma);
};


#calcula varias parametros a priori

sub funcion{
   my ($i,$j,$tipo) = @_;
   my $aux = 0;

   SWITCH: for($tipo) {
      /1/   && do
                  {  # funcion m^ij
                     $aux = 0;
                     last;
                  };
      /2/   && do
                  {  # funcion a^ij
                     $aux = ($j-$i)/2;
```

```
      #$aux = 0.01;
                        last;
                    };
       /3/   && do
                    {  #funcion d^ij
      #$aux = 4;
                        $aux = (($j-$i)**2)/2 +2;
                        last;
                    };

       die "funcao desconhecida";
   }
   return($aux);
};



# calcula o fator que  acompanha as coesoes

sub calculaG{
   my ($t) = @$_;

   for($d = 1;$d <= $n;$d++) {
      $dij = funcion(0,$d,3);
      $x = ($dij+$d)/2 -1;
      $y = ($dij)/2 -1;
      $gamas[$d] = -$x + $y;
      $gamas[$d] += ($x+0.5)*log($x)-($y+0.5)*log($y);
      $gamas[$d] +=
1/(12*($x)) - 1/(360*($x**3))+1/(1260*($x**5))-1/(1680*($x**7));
      $gamas[$d] -=
```

```
1/(12*($y)) + 1/(360*($y**3))-1/(1260*($y**5))+1/(1680*($y**7));
      $gamas[$d] -= $d/2*log($pi);
      $gamas[$d] += $dij*0.5*log(&funcion(0,$d,2));
      $gamas[$d] -= 0.5*log(1+$t*$d);
      # print "gamas$d :", $gamas[$d],"\n";
   };
   return(@gamas);
}



# calcula a matriz de coesoes c_ij^*

sub calculaC {
   my ($pv,$pg,$t,$p) = @_;
   my $aux = 0;
   local @vv = @$pv;
   local @gamas = @$pg;

   for($i = 0;$i < $n;$i++) {
      for($j = $i+1;$j < $n;$j++) {
         # print "gamas$i $j:", $gamas[$j-$i],"\n";
         $cp[$i][$j] = ($j-$i-1)*log(1-$p);
         $cp[$i][$j] += log($p);
         $cp[$i][$j] += $gamas[$j-$i];
         $cp[$i][$j] +=
(-(funcion($i,$j,3)+$j-$i)/2)*log(funcion($i,$j,2)+M(\@vv,$t,$i,$j));
         $cp[$i][$j] = exp($cp[$i][$j]);
      }
      $j = $n;
      #print "gamas$i $j:", $gamas[$j-$i],"\n";
```

```
      $cp[$i][$j] = ($j-$i-1)*log(1-$p);
      $cp[$i][$j] += $gamas[$j-$i];
      $cp[$i][$j] +=
(-(funcion($i,$j,3)+$j-$i)/2)*log(funcion($i,$j,2)+M(\@vv,$t,$i,$j));
      $cp[$i][$j] = exp($cp[$i][$j]);
   };
   return (@cp);
};



# calcula o valor de M_ij

sub M {
   my ($pv,$t,$i,$j) = @_;
   my $dif = 0;
   my $suma = 0;
   my $sumaC = 0;
   my $Mij = 0;
   local @vv = @$pv;

   for($r = $i+1;$r <= $j;$r++) {
      $dif = $vv[$r]-funcion($i,$j,1);
      $suma += $dif;
      $sumaC += ($dif)**2;
   };
   $Mij = $sumaC - $t/(1+$t*($j-$i))*($suma)**2;
   return($Mij);
}
```

```
# calcula o vetor lambda 0

sub calculaL0{
   my ($pc) = @_;
   local @cp = @$pc;
   my $suma;

   $l0[1] = $cp[0][1];
   for($j = 1;$j < $n;$j++) {
      $suma = 0;
      for($s = 1;$s <= $j;$s++) {
         $suma += $l0[$s]*$cp[$s][$j+1];
      };
      $l0[$j+1] = $cp[0][$j+1] + $suma;
   };
   return(@l0);
};


# calcula o vetor lambda N

sub calculaLN{
   my ($pc) = @_;
   local @cp = @$pc;
   my $suma = 0;

   $lN[$n-1] = $cp[$n-1][$n];
   for($r = $n-2; $r >= 1;$r--) {
      $suma = 0;
      for($s = $r+1;$s <= $n-1;$s++) {
```

```
            $suma += $lN[$s]*$cp[$r][$s];
        };
        $lN[$r] = $cp[$r][$n] + $suma;
    };

    return(@lN);
}



# calcula as relevancias r_ij^*

sub calculaR {
    my ($pc,$pl0,$plN) = @_;
    local @cp = @$pc;
    local @l0 = @$pl0;
    local @lN = @$plN;

    for($j = 1;$j < $n; $j++) {
        $rp[0][$j] = $cp[0][$j]*$lN[$j]/$l0[$n];
         #print "r0 $j:", $rp[0][$j],"\n";
    };
    $rp[0][$n] = $cp[0][$n]/$l0[$n];
     #print "r0 $n:", $rp[0][$n],"\n";
    for($i = 1;$i < $n;$i++) {
for($j = $i+1;$j < $n;$j++) {
            $rp[$i][$j] = $l0[$i]*$cp[$i][$j]*$lN[$j]/$l0[$n];
            # print "r$i $j:", $rp[$i][$j],"\n";
};

$rp[$i][$n] = $l0[$i]*$cp[$i][$n]/$l0[$n];
```

```
        #print "r$i $n:", $rp[$i][$n],"\n";
    };
    return(@rp);
}

# calcula os a_ij^*

sub calculaA{
my ($px,$t) = @_;
local @x = @$px;

for($i = 0;$i < $n;$i++) {
for($j = $i+1;$j <= $n;$j++) {
$denom = (1+$t*($j-$i));
$ap[$i][$j] = funcion($i,$j,2);
                        $ap[$i][$j] += (($j-$i)*(funcion($i,$j,1))**2)/$denom;
$ss = suma(\@x,$i+1,$j);
$ap[$i][$j] += sumaC(\@x,$i+1,$j)-(2*funcion($i,$j,1)/$denom)*$ss;
$ap[$i][$j] -= ($t/$denom)*($ss**2);
                        # print "a$i $j:", $ap[$i][$j],"\n";
};
};
return(@ap);
}


# calcula a soma do quadrado dos dadosentre i e j

sub sumaC{
my ($px,$i,$j) = @_;
```

```
local @x = @$px;

$suma = 0;
for($r = $i;$r <= $j;$r++)
{
$suma += $x[$r]**2;
};
return($suma);
}


# calcula os valores de d_ij^*

sub calculaD{

for($i = 0;$i < $n;$i++)
{
for($j = $i+1;$j <= $n;$j++)
{
$dp[$i][$j] = funcion($i,$j,3)+$j-$i;
                        #print "d$i $j:", $dp[$i][$j],"\n";
}
}
return(@dp);
}


# calcula as medias a posteriori

sub calculaMedias{
```

```
my ($pm,$pr,$pa,$pd) = @_;
local @mP = @$pm;
local @rP = @$pr;
local @aP = @$pa;
local @dP = @$pd;

for($k = 1;$k <= $n;$k++)
{
$sumaM = 0;
$sumaS = 0;
for($i = 0; $i < $k;$i++)
{
for($j = $k;$j <= $n;$j++)
{
                                $sumaM += $mP[$i][$j]*$rP[$i][$j];
$sumaS += ($aP[$i][$j]/($dP[$i][$j]-2))*$rP[$i][$j];
};
};
$med[$k][1] = $sumaM;
$med[$k][2] = $sumaS;
#               print "calculaMedias: ", $med[$k][1], "\t", $med[$k][2], "\n";
};
return(@med);
}


# calcula as cadeias

sub calculaCadenas{
my ($pc,$lim) = @_;
```

```
local @cp = @$pc;

open DATA,"cadena0" or die "No se pode abrir o arquivo de dados";

   # le a  semente

   $i = 0;
   $cAux[$i++]=" ";
   $suma0 = 0;
   while (<DATA>) {
    $aux = substr($_ , 0, -1);
    if(ord($aux)!=0){ $cAux[$i++] = $aux; $suma0 += $aux;};
   };
   close DATA;
   $cad[0][1] = join("",@cAux);
   $cad[0][2] = $suma0;

# calcula o resto das cadeias

for($k = 1;$k <= $lim;$k++)
{
$cad[$k][1] = " ";
$cad[$k][2] = 0;
for($r = 1;$r < $n;$r++)
{
$y = index($cad[$k-1][1],´0´,$r+1);
if($y == -1) { $y = $n;};

$x = rindex($cad[$k][1],´0´,$r-1);
if($x == -1) { $x = 0;};
```

```
$raz = $cp[$x][$y]/($cp[$x][$r]*$cp[$r][$y]);
srand;
$u = rand;
                        while ( $u == 0.0 ) {
   $u = rand;
                        }
if( $raz >= (1-$u)/$u) {
$uk = 1;
$cad[$k][2]++;
}
else {
$uk = 0;
};
$cad[$k][1] = join("",$cad[$k][1],$uk);

};
};
return(@cad);
}


# calcula o numero de cadeias iguais

sub cuentaCuantas{
my ($pcad) = @_;
local @cad = @$pcad;

$listaCad[1][1] = $cad[$est][1];
$listaCad[1][2] = $cad[$est][2];
```

```
$listaCad[1][3] = 1;
$indLista = 1;
$bk[$est] = $n-$cad[$est][2];

for($i = $est+1; $i <= $lim;$i++)
{
$esta = 0;
for($j = 1;$j <= $indLista;$j++)
{
if($cad[$i][1] == $listaCad[$j][1])
{
# si la cadena ya esta registrada
$listaCad[$j][3]++;
$esta = 1;
};
};

# se a cadeia nao estava registrada

if(!$esta)
{
$indLista++;
$listaCad[$indLista][1] = $cad[$i][1];
$listaCad[$indLista][2] = $cad[$i][2];
$listaCad[$indLista][3] = 1;
$esta = 0;
};
$bk[$i] = $n-$cad[$i][2];
};
```

```
open fpeb,">peb";
select fpeb;

# for($i = 1;$i <= $indLista;$i++)
# {
# print
$listaCad[$i][1],"\t",$listaCad[$i][2],"\t",$listaCad[$i][3],"\n";
# };

print "b\tnb\tpeb\tvab\n";
for($i = 1;$i <= $n;$i++)
{
$nb[$i] = 0;
for($j = 1;$j <= $indLista;$j++)
{
if($listaCad[$j][2] == $n-$i)
{
$nb[$i] +=$listaCad[$j][3];
};
};

$peb[$i] = $nb[$i]/($lim-$est+1);

$sqb[$i] =
$nb[$i]*((1-$peb[$i])**2)+($lim-$est+1-$nb[$i])*(($peb[$i])**2);
$vab[$i] = $sqb[$i]/($lim-$est+1);

print $i,"\t",$nb[$i],"\t",$peb[$i],"\t",$vab[$i],"\n";
};
```

```
#ordenar as configuracoes

for($i = 1;$i <= $indLista;$i++)
{
for($j = $i+1;$j <= $indLista;$j++)
{
if($listaCad[$i][3] < $listaCad[$j][3])
{
# es menor
@listaAux = @listaCad[$i];
@listaCad[$i] = @listaCad[$j];
@listaCad[$j] = @listaAux;
};
};
};

if($indLista > 10){ $indLista = 10;};
printf "\n\nCadeias mais frequentes\n";
print "cadena\tPEUI\n";
for($i = 1; $i <= $indLista;$i++)
{
print $listaCad[$i][1],"\t",$listaCad[$i][3]/($lim-$est+1),"\n";
};
close fpeb
}
```

Apêndice Quatro

Os gráficos a seguir mostram a convergência e auto-correlação na amostra gerada para estimar as distribuições *a posteriori* para a partição ρ e para o número de blocos B. Perceba que em alguns casos, a correlação de ordem 1 ultrapassa os limites do intervalo a 95% de confiança, indicando correlação serial de ordem 1. Apesar disto, as inferências exibidas no capítulo 5 não estão comprometidas.

Figura 5.43: Convergência e Correlação para a amostra gerada- $m = 0$, $a = 0.5$ e $b = 0.5$

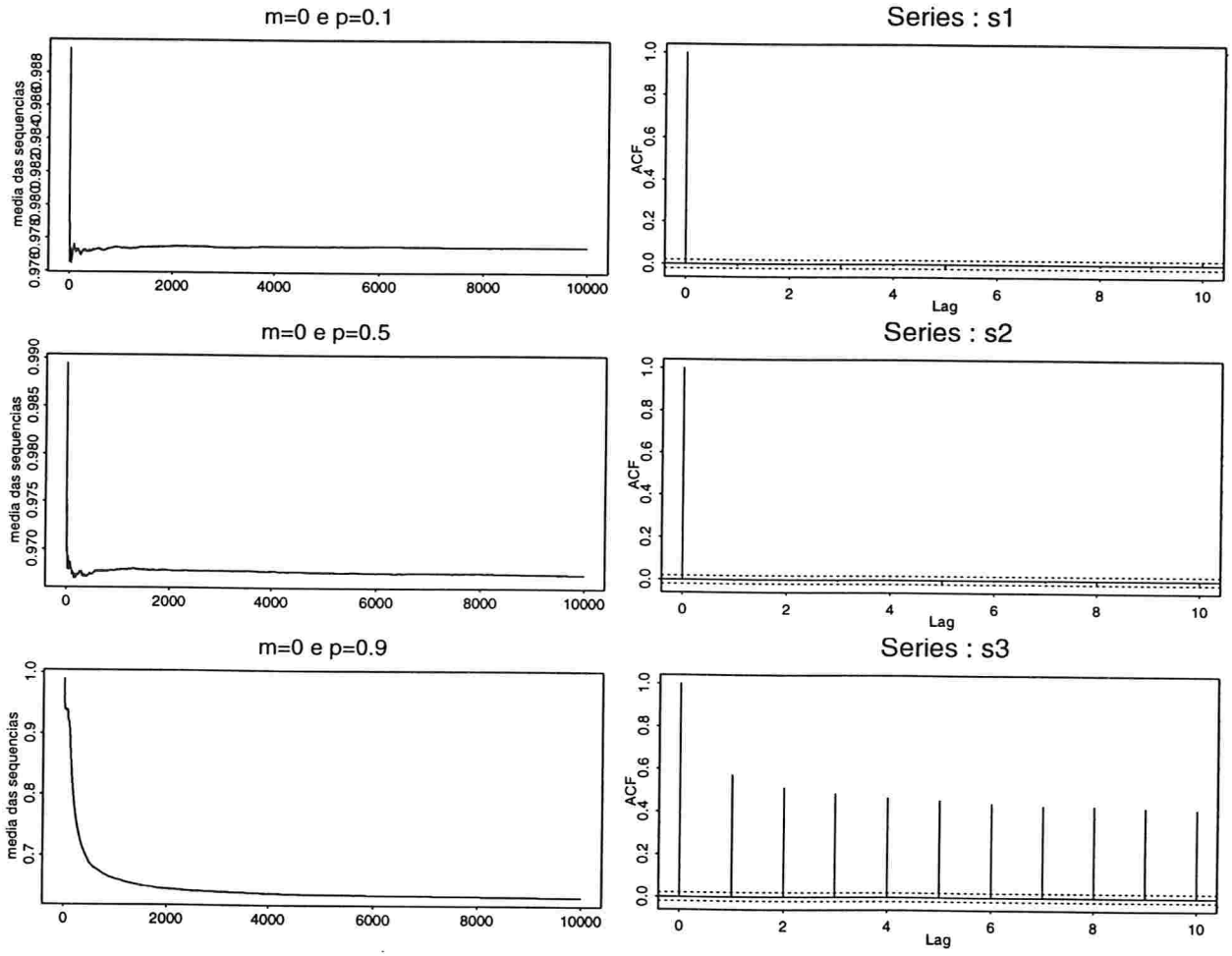


Figura 5.44: Convergência e Correlação para a amostra gerada- $m = 0.05$, $a = 0.5$ e $b = 0.5$

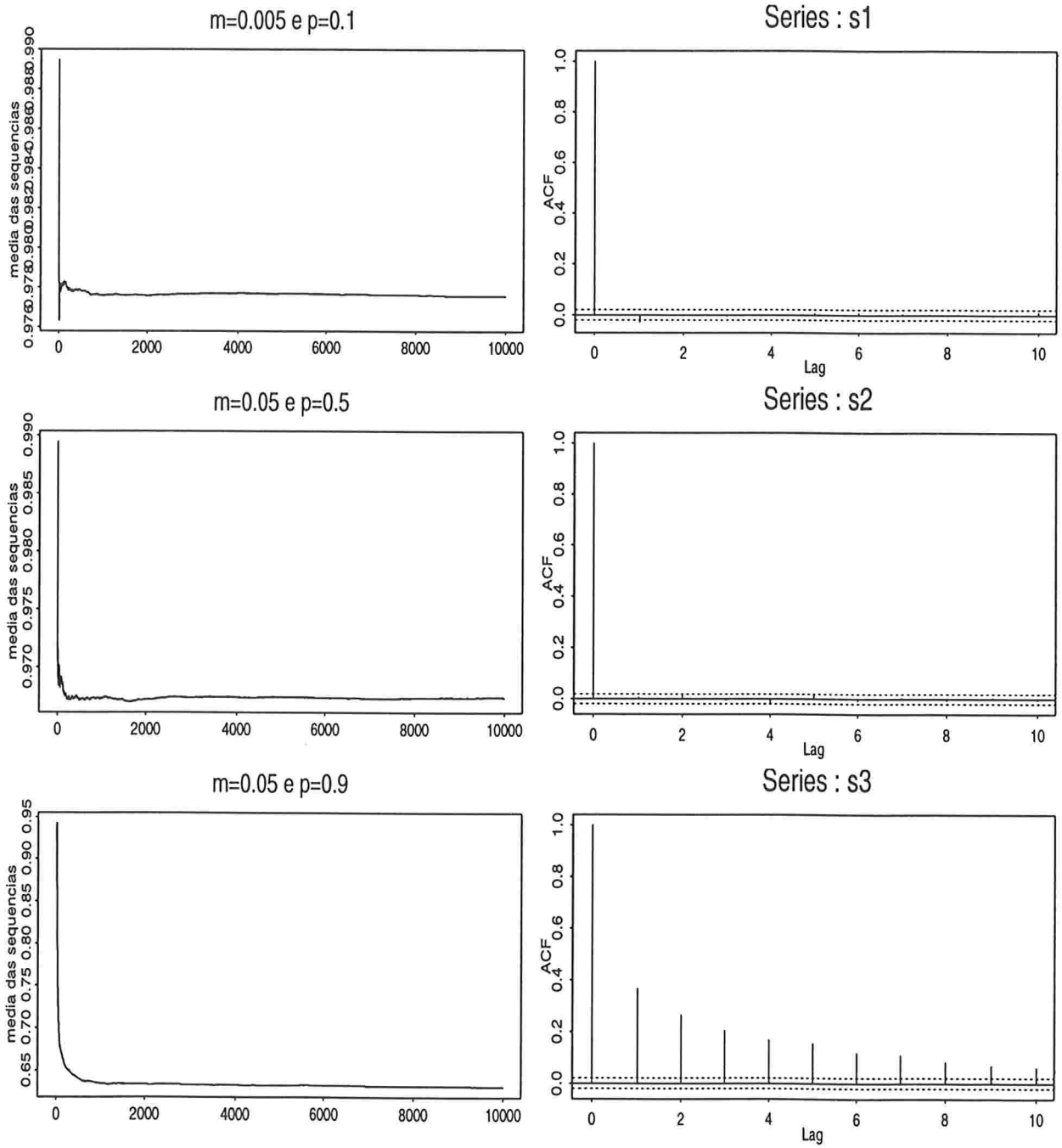


Figura 5.45: Convergência e Correlação para a amostra gerada- $m = 0$, $a = 0.01$ e $b = 3$

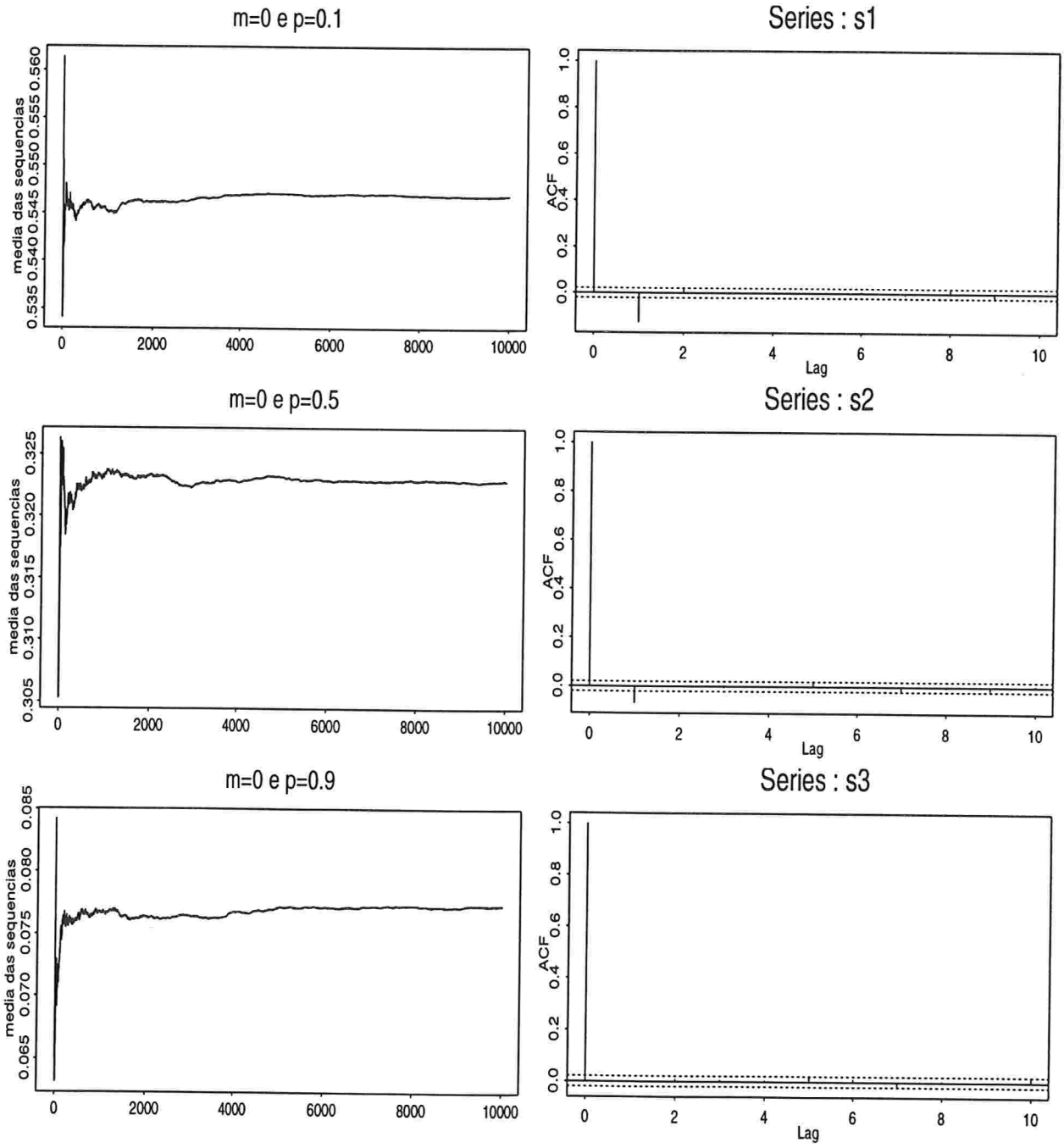
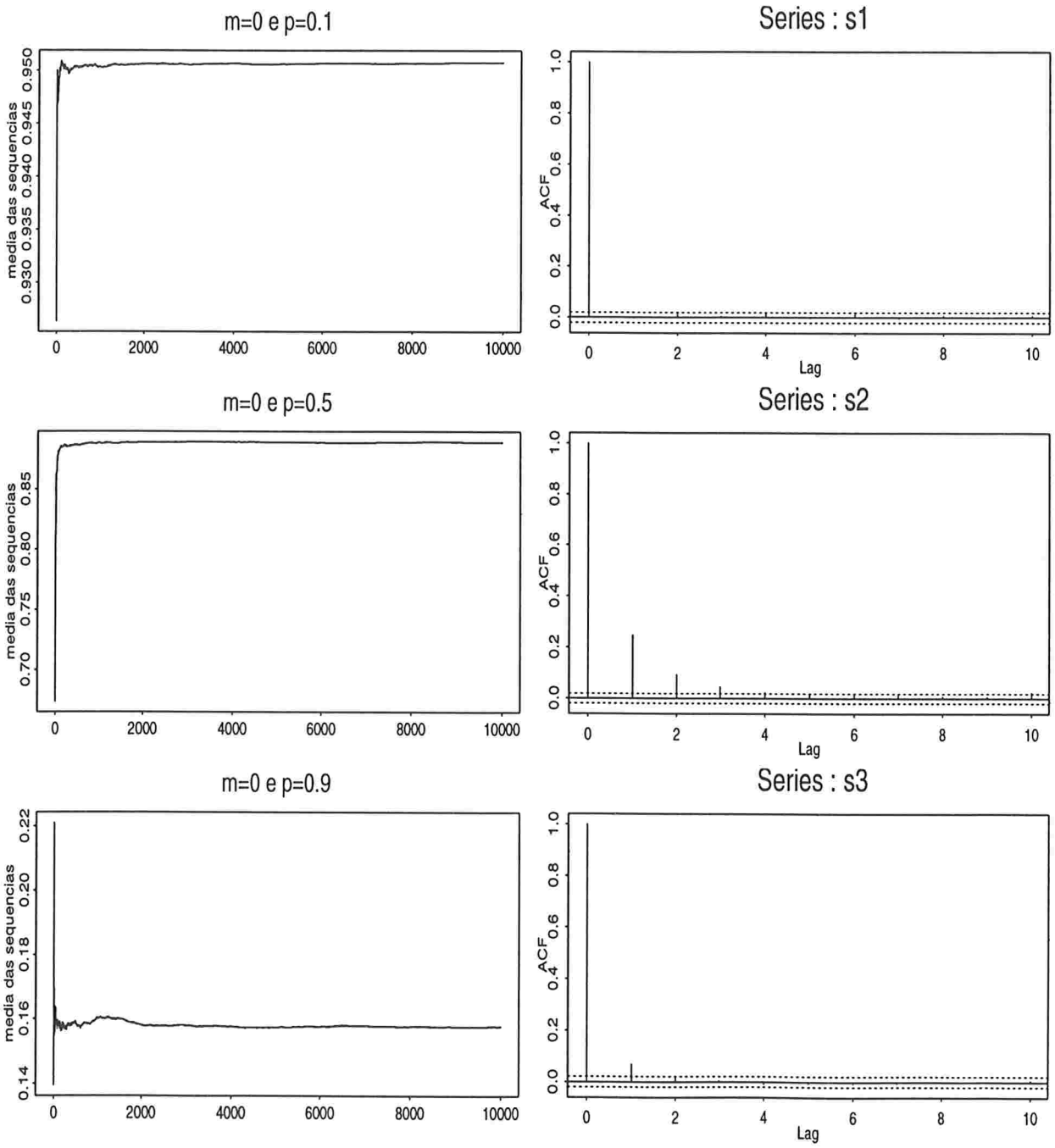


Figura 5.46: Convergência e Correlação para a amostra gerada- $m = 0$, $a = 1$ e $b = 3$



Assumimos que $(\sigma^{ij})^2$ tem seu comportamento descrito pela distribuição $\mathcal{GI}(a/2, d/2)$ para todo i e j obtemos:

Figura 5.47: Convergência e Correlação para a amostra gerada- $m = 0$, $a = 0.01$ e $d = 3$

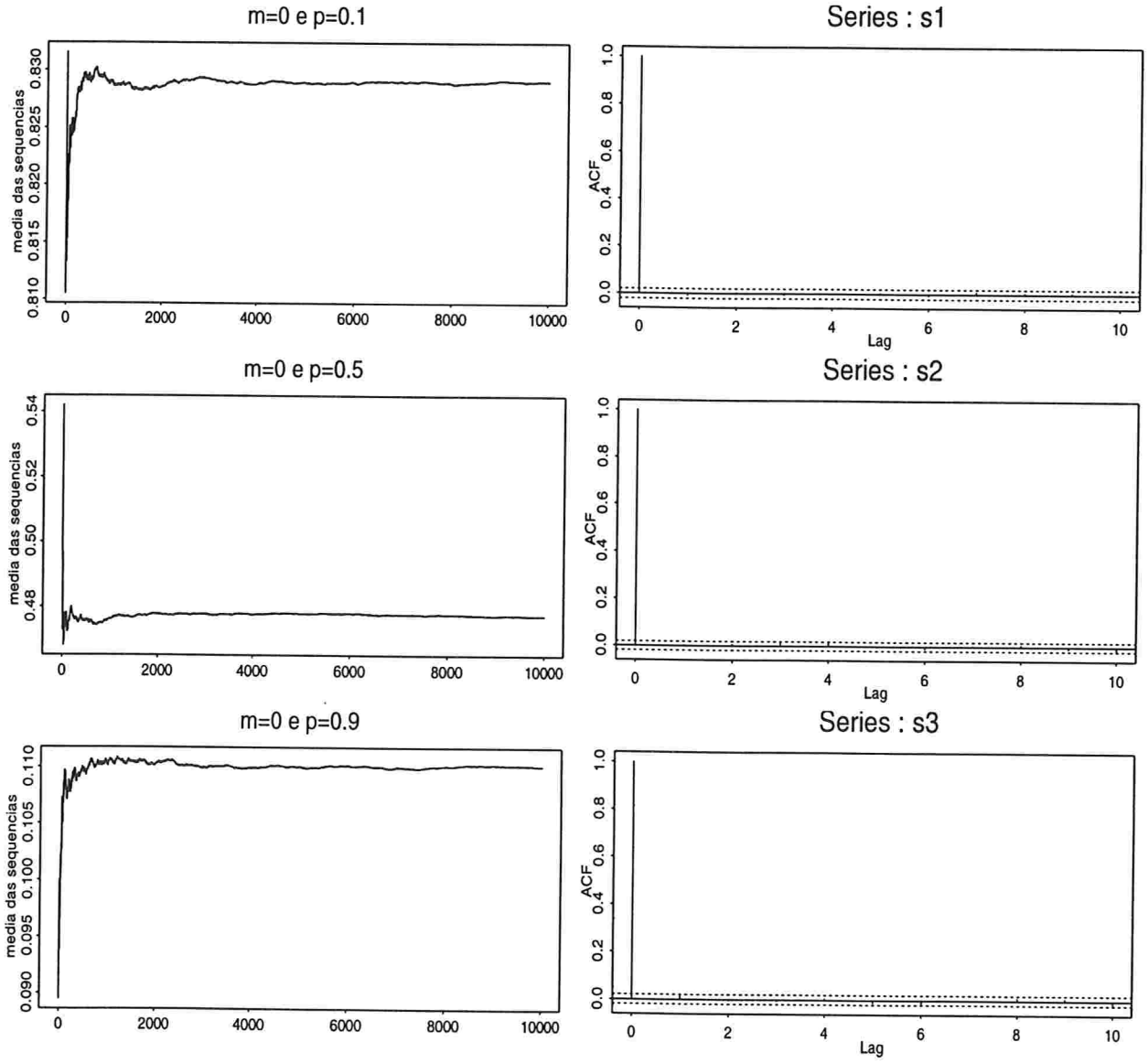
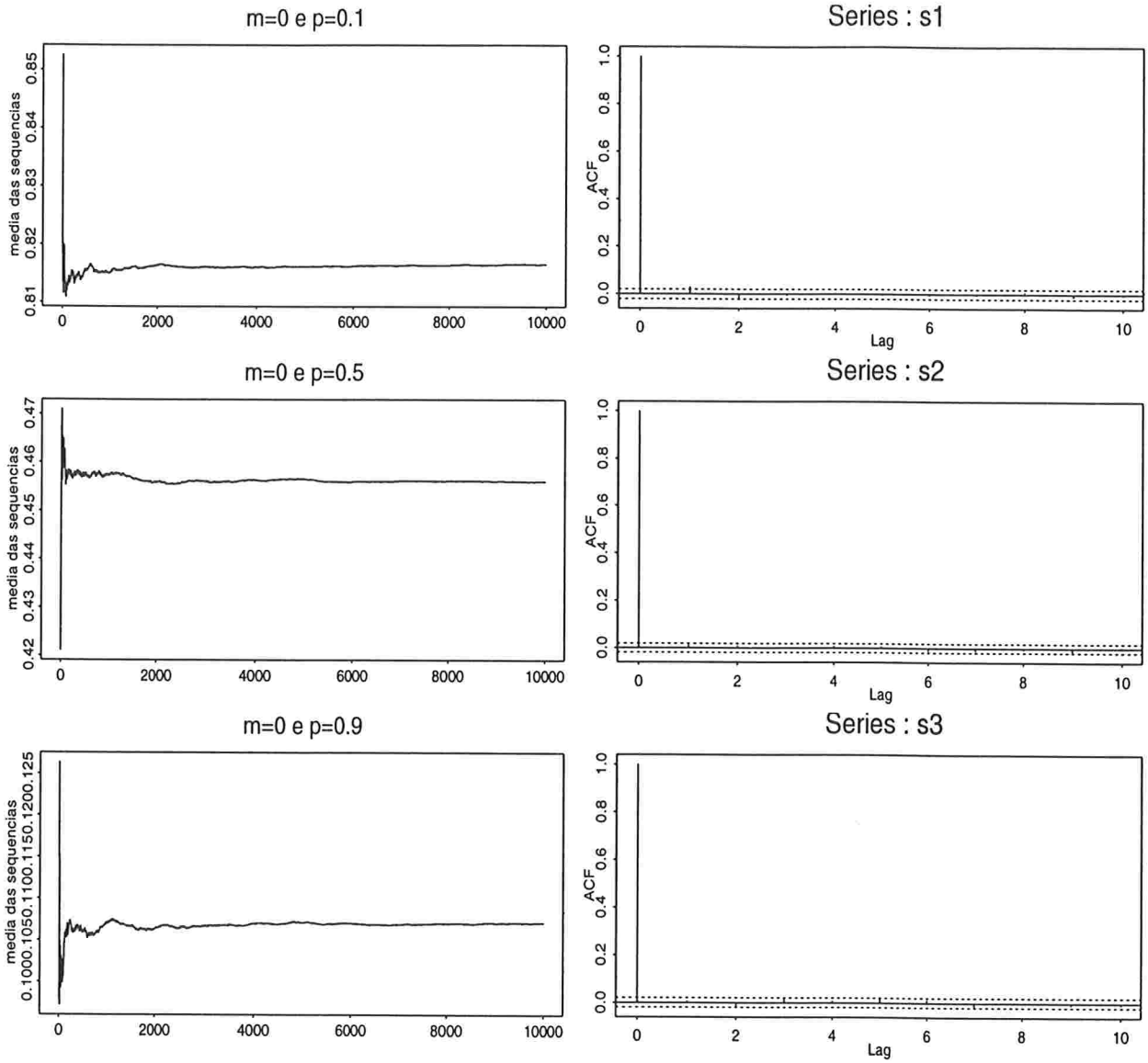


Figura 5.48: Convergência e Correlação para a amostra gerada- $m = 0$, $a = 0.01$ e $b = 4$



Referências Bibliográficas

Anderson, T.W. e Taylor, J.B.(1976) - Strong Consistency of Least Squares Estimates in Normal Linear Regression - *The Annals of Statistics.* V.4 - N.4 - Pg 788-790.

Arellano-Valle, R.B. *et alii* (1994) (com Bolfarine, H.; Iglesias, P.L.)- A Predictivistic Interpretation of the Multivariate t Distribuition. *Test.* V.3 - N.2 - Pg. 221-236.

Arellano-Valle, R.B e Bolfarine, H. (1994)- On Some Characterizations of the t-distribution. *Statistics and Probability Letters.* V.25 - Pg. 79-85.

Armendt, B (1980)- Is There a Dutch Book Argument for Probability Kinematics? *Philosophy of Science.* V.47 - Pg. 583-588.

Barry, D. e Hartigan, J.A.

(1992)- Product Partition Models for Change Point Problems. *The Annals of Statistics.* V.20 - N.1 - Pg. 260-279.

(1993)- A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association.* V.88 - N.421 - Pg. 309-319.

Bayes, T. (1763) - An Ensay towards Solving a Problem in the Doctrine of Chance - Publicado postumamente em *Phil. Trans. Roy. Soc..* V.53 - Pg. 370-418 e V.54 - Pg. 296-325. Reimpresso em (1958). *Biometrika* - V.45 - Pg. 293-315. Reproduzido em Press(1989) - Pg. 185-217.

Blackwell, D. e Girshick, M.A.(1954) - *Theory of Games and Statistical Decision.* John Wiley.

Bernardo, J. M. e Smith, A.F.M.(1994)- *Bayesian Theory* - John Wiley.

Billingsley, P.(1979)- *Probability and Measure.* John Wiley & Sons.

Box, G.E.P e Tiao, G.C.(1973)- *Bayesian Inference in Statistical Analysis.* Addison-Wesley.

Casella,G. e Berger,R.L.(1990)- *Statistical Inference.* Pacific Grove, Califórnia: Waasworth and Brooks/Cole.

Cifarrelli, D. e Regazzini,E.(1982)- Some Considerations about Mathematical Statistics Teaching Methodology Suggested by the Concept of Exchangeability. Em *Exchangeability in Probability and Statistics.* G. Kock e F. Spizzichino (ed.). North Holland.

Chow, Y. e Teicher, H.(1988)-*Probability Theory.* Springer.

Correa, L.(1998) - *Modelación Bayesiana de Puntos de Cambio en la Volatilidad.* Tese de Mestrado. Facultad de Matemáticas - Pontifícia Universidad Católica de Chile.

Crowley, E. M.(1997)- Product Partition Models for Normal Means. *Journal of the American Statistical Association.* V.92 - N.437 - Pg. 192-198.

Csiszár, I.(1975)- I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability.* V.3 - N.1- Pg. 146-158.

Dawid, A.P.(1985)- The Impossibility of Inductive Inference. *Journal of the American Statistical Association.* V.80 - N.390 - pp.340-341.

Dawid, A.P.(1982)- Intersubjective Statistical Models. Em *Exchangeability in Probability and Statistics.* G. Kock e F. Spizzichino (ed.) - North Holland.

De Finetti, B.

(1989)- Probabilidade. *Enciclopédia Einaudi*, V.15. Imprensa Nacional-Casa da Moeda-Lisboa.

(1975)- *Theory of Probability*- V. 2 - New York - John Wiley & Sons Ltd.

(1974)- *Theory of Probability*- V. 1 - New York - John Wiley & Sons Ltd.

(1972)- *Probability, Induction and Statistics*- New York - John Wiley & Sons Ltd.

(1937)- Foresight: Its Logical Laws, Its Subjective Sources. Translated and Reprinted in *Studies in Subjective Probability*, H.E. Kyburg Jr. e H.E. Smokle(ed.). Segunda Edição. New York. Robert E. Krieger Publishing Co. Inc., Huntington - 1980.

(1931)- Sul Significato Soggettivo della Probabilità. *Fundamenta Mathematicae*. V.17 - Pg. 298-329.

DeGroot, M.H.(1970) - *Optimal Statistical Decison*. McGraw-Hill. New York.

Del Pino, G (1996) - Comunicação Privada.

Diaconis, P. *et alii*(1992) (com Eaton,M.L. e Lauritizen, S.L.) - Finite de Finetti Theorems in Linear Models and Multivariate Analysis. *Scandinavian Journal of Statistics* - V. 19 - Pg. 298-315.

Diaconis, P. e Freedman, D.(1990)- Cauchy's Equation and de Finetti's Theorem. *Scandinavian Journal of Statistics*. V.17 - Pg.235 -250.

Diaconis, P. e Zabell, S. L. (1982) - Updating Subjective Probability. *Journal of the American Statistical Association*. V.77 - N. 380 - Pg. 822-830.

Diaconis, P. e Ylvisaker, D.

(1985)- Quantifying Prior Opinion. *Bayesian Statistical 2*- J.M. Bernardo, M.H. DeGroot, D.V. Lindley e A. F. M. Smith (ed.). Elsevier Science Publishers B.V. - North-Holland - Pg. 133-156.

(1979)- Conjugate Priors For Exponential Families. *The Annals of Statistics.* V.7 - N. 2- Pg. 269-281.

Domotor,Z.(1980)- Probability Kinematics and Representation of Belief Change. *Philosophy of Science.* V.47 - Pg. 384-403.

Eaton, M.L.(1989) - Group Invariance Applications in Statistics. Em *Regional Conference Series in Probability and Statistics, Inst. Math. Stat. e Ameriacan Stat. Assoc.* - V. 1.

Fama, E.F.

(1970) - Efficient Capitol Markets: A Review of Theory and Empirical Work. *Journal of Finance.* Pg. 383-423.

(1965) - The Behavior of Stock-Market Prices. *The Journal of Business.* V. 38 - N.1 - Pg. 34-105.

Fang, K. e Zhang, Y.(1980) - *Generalized Multivariate Analysis.* Spring-Verlag.

Fienberg, S.(1970)- An Interative Procedure for Estimation in Contingency Tables. *Annal of Mathematical Statistics.* V. 41 - Pg. 907-917.

Gamerman, D. e Migon, H.S.(1993) - *Inferência Estatística: Uma abordagem Integrada* - Monografia do Instituto de Matemática -UFRJ.

Gamerman, D.(1997) - *Markov Chain Monte Carlo:Stochastic Simulation for Bayesian Inference*

- Chapman & Hall.

Gasco, L.Z.(1997) - *Previsão e Preditivismo em Modelos Lineares com e sem Erros nas Variáveis Preditivas*- Tese de Doutorado - Instituto de Matemática e Estatística. Universidade de São Paulo.

Goldstein, M.

(1985) - Temporal Coherence. *Bayesian Statistical 2* - J. M. Bernardo, M. H. DeGroot, D.V. Lindley e A. F. M. Smith (ed.). Elsevier Science Publishers B.V. - North-Holland - Pg. 231-248.

(1983)- The Prevision of a Prevision. *Journal of the American Statistical Association.* V.78 - N.384 - Pg. 817-819.

Green, P.J.(1995) - Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika.* V. 82 - N. 4 - Pg. 711-732.

Hacking, I(1967)- Slightly More Realistic Personal Probability. *Philosophy of Science.* V.34 - Pg. 311-325.

Hartigan, J.A.(1990) - Partition Models. *Communication in Statistics - Theory and Method.* V.19(8) - Pg. 2745-2756.

Howson, C. Urbach, P.(1993)- *Scientific Reasoning: The Bayesian Approach.* Segunda Edição - Chicago - Open Court.

Howson, C.(1996) - Bayesian Rules of Updating - *Erkenntnis.* V.45 - N. 2 e 3 - Pg.195-208.

Hsu, D.A.(1982)- A Bayesian Robust Detection of Shift in the Risk Struture of Stock Market

Returns. *Journal of the American Statistical Association.* V.77 -N.2 - Pg. 29-39.

Iglesias Z., P.L.(1993)- *Finite Forms of the de Finetti's Theorem: A predictivistic Approach to Statistical Inference in Finite Populations.* Tese de Doutorado. Instituto de Matemática e Estatística - Universidade de São Paulo.

Iglesias Z., P.L., Pereira, C.A. e Tanaka, N.I.(1998)- Characterizations of Multivariate Symmetric Distributions in l_∞ - Norm. A aparecer em *Test.*

Jeffrey, R.

(1992)- *Probability and the Art of Judgment* - Cambridge - Cambridge University Press.

(1965)- *The Logic of Decision* - New York - McGraw-Hill.

Kingman,J.F.C (1972)- On Random Sequences with Spherical Symmetry. *Biometrika.* V. 59 - pg. 183-197.

Lewis, D.(1975)- Probabilities of Conditionals and Conditional Probabilities. *The Philosophical Review.* V. LXXXV - Pg. 297-315.

Loschi, R.H. e Wechsler, S. (1997)- Coherence and Probability. Submetido à *REBRAPE* em outubro de 1997.

Maeda, M. A.(1996) - *Volatilidad Estocástica en el Mercado Accionário Chileno* - Tese de Mestrado. Facultad de Ciencias Económicas y Administrativas. Universidad de Chile.

Mandelbrot, B.(1963) - The Variation of Certain Speculative Prices. *Journal of Business.* V. 36 - Pg. 394-419.

Menzefricke, U.(1981)- A Bayesian Analysis of a Change in the Precision of a Sequence of

Independents Normal Random Variables at an Unknown Time Point. *Applied Statistics.* V. 30 - N.2 - Pg.141-146.

O'Hagan, A. (1994)- *Bayesian Inference.* Em "Kendall's Advanced Theory of Statistics". V.2B - New York - John Wiley.

Pereira, C.A.B.(1995)- Estatística e Informação. *Boletim da ABE*, Ano XI, N.30.

Pereira, C.A.B. *et alii*(1993)- *Teoria Estatística.* 19° Colóquio Brasileiro de Matemática. CNPq - IMPA - RJ.

Press, S.J.(1989) - *Bayesian Statistics* - Wiley.

Rao, C. R.(1973)- *Linear Statistical Inference and its Applications.* John Wiley & Sons . Segunda edição.

Searle, S. R.(1982)- *Matrix Algebra Useful for Statistics.* New York - John Wiley.

Singer, J.M. e Sen,P.K.(1993)- *Large Sample Methods in Statistics: An Introduction with Applications..* Chapman & Hall.

Silva, A.J. da (1997)- *Sobre Condições de Estendibilidade, Simetria Esférica e Inferência Bayesiana em Modelos Elípticos com Erros nas Variáveis.* Tese de Doutorado. Instituto de Matemática e Estatística. Universidade de São Paulo.

Skyrms, B.(1987)- Dynamic Coherence and Probability Kinematics. *Philosophy of Science.* V.54 - Pg. 1-20.

Smith, A.F.M

(1981)- On Random Sequences With Centered Spherical Symmetry. *Journal of the Royal*

Statistical Society B . V.43 - Pg. 203-241.

(1975)- A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables. *Biometrika.* V.62 - N.2. Pg.407-416.

Sthephens, D.A.(1994)- Bayesian Retrospective Multiple-Changepoint Identification. *Applied Statistics.* V. 43 - N.1 - Pg 159-178.

Stigler, S.M.(1982)- Thomas Bayes's Bayesian Inference. *Journal of Royal Statistics A.* V.145 - Parte 2 - Pg. 250-258.

Székely, G. J.(1986)- *Paradoxes in Probability Theory and Mathematical Statistics.* Mathematics and Its Applications - East European Series. D. Reidel Publ.

Teller, P.(1973)- Conditionalization and Observation. *Synthese.* V.26 - Pg. 218-258.

Van Frassen, V.C.(1980)- Rational Belief and Probability Kinematics. *Philosophy of science.* V.47 - Pg. 165-187.

Wechsler, S.(1993)- Exchangeability and Predictivism. *Erkenntnis.* V.38 - Pg. 343-350.

Yao, Y.(1984)- Estimation of a Noisy Discrete-Time Step Function: Bayes and Empirical Bayes Approaches. *The Annals of Statistics.* V.12 - N.4 - Pg. 1434-1447.

Zabell, S. L.(1992)- Predicting the Unpredictable. *Synthese.* V.90 - Pg. 205-232.