

**Estatística gradiente:  
teoria assintótica de alta  
ordem e correção tipo-Bartlett**

Tiago Moreira Vargas

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

Programa: Estatística

Orientador: Profa. Dra. Silvia Lopes de Paula Ferrari

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da

CAPES e CNPq

São Paulo, Março de 2013

**Estatística gradiente:  
teoria assintótica de alta  
ordem e correção tipo-Bartlett**

Esta é a versão original da tese elaborada pelo  
candidato Tiago Moreira Vargas, tal como  
submetida à Comissão Julgadora.

# **Estatística gradiente: teoria assintótica de alta ordem e correção tipo-Bartlett**

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 15/04/2013. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof<sup>a</sup>. Dr<sup>a</sup>. Silvia Lopes de Paula Ferrari (orientadora) - IME-USP
- Prof<sup>a</sup>. Dr<sup>a</sup>. Denise Aparecida Botter - IME-USP
- Prof. Dr. Artur José Lemonte - UFPE
- Prof. Dr. Klaus Leite Pinto Vasconcellos - UFPE
- Prof. Dr. Miguel Angel Uribe Opazo - UNIOESTE

*Ao meu irmão Hélio Vinícius Moreira Vargas (in memoriam)*

# Agradecimentos

Agradeço aos meus pais Hélio Vargas e Adelta Vargas, que são meu alicerce, minha guia.

À professora Silvia Lopes de Paula Ferrari pela orientação, por sua dedicação e incentivo.

A Artur José Lemonte, pela importante colaboração neste trabalho.

Aos amigos do IME: Dylene Barros, Fran Lima, Rodrigo Lambert, Karina Yaginuma, Wagner Barreto de Souza, Alice Lemos, Rosemary Pires, Joelmir Borssoi, Camila Bertini, Débora Lopes, Aníbal Emiliano, Jalmar Carrasco e Eliane Pinheiro, que me acompanharam nessa caminhada.

Aos meus amigos Dimmy Sampaio, Letícia Tavares, Maria Luiza Valdez, Heloísa Rodrigues, Kamilla Palhares, Tainã Goes, Eleen Rodrigues, Silvana Rodrigues, Suzana Guminiak, Paulo Otávio Schotz, Jardel Mesquita e Núbia César, pelos momentos divertidos e agradáveis em São Paulo e Goiânia.

Aos meus familiares da cidade de Piracanjuba-GO, terra natal.

Aos professores da época da UFG: Denise Duarte, Osvaldo Scarpa Magalhães Alves (in Memoriam), Ronaldo Alves Garcia, Alacyr José Gomes e Marcelo Almeida de Souza.

Aos professores Fábio Machado, Heleno Bolfarine e Silvia Elian.

À CAPES e ao CNPq pelo apoio financeiro.

# Resumo

Obtemos uma expansão assintótica da função de distribuição acumulada, sob a hipótese nula, da estatística gradiente para testar hipóteses nulas compostas na presença de parâmetros de perturbação. Esta expansão é derivada utilizando uma rota Bayesiana com base no argumento de encolhimento descrito em Ghosh e Mukerjee (1991). Usando essa expansão, propomos uma estatística gradiente corrigida por um fator de correção tipo-Bartlett que tem distribuição qui-quadrado até um erro de ordem  $o(n^{-1})$  sob a hipótese nula. A partir disso, determinamos fórmulas matriciais e algébricas que auxiliam na obtenção da estatística gradiente corrigida em modelos lineares generalizados com dispersão conhecida e desconhecida. Simulações Monte Carlo são apresentadas. Finalmente, discutimos a obtenção de regiões de credibilidade via inversão da estatística gradiente. Caracterizamos as densidades a priori de correspondência, que asseguram propriedades de cobertura frequentista acuradas para essas regiões.

*Palavras-chave:* Argumento de encolhimento; Correção tipo-Bartlett; Expansão assintótica; *Matching priors*; Rota Bayesiana; Teste gradiente.

# Abstract

We obtain an asymptotic expansion for the null distribution function of the gradient statistic for testing composite null hypotheses in the presence of nuisance parameters. The expansion is derived using a Bayesian route based on the shrinkage argument described in Ghosh and Mukerjee (1991). Using this expansion, we propose a Bartlett-type corrected gradient statistic, which has a chi-square distribution up to an error of order  $o(n^{-1})$  under the null hypothesis. Also, we determined matrix and algebraic formulas that assist in obtaining Bartlett-type corrected statistic in generalized linear models with known and unknown dispersion. Monte Carlo simulations are presented. Finally, we obtain credible regions based by the inversion of gradient statistic. We characterize priori densities, matching priors, that ensure accurate frequentist coverage properties for these regions.

*Keywords:* Asymptotic expansion; Bartlett-type correction; Bayesian route; Gradient test; Matching priors; Shrinkage argument.

# Lista de Figuras

2.1	(a) Distorção de tamanho do teste gradiente (linha contínua) e do teste gradiente corrigido (linha pontilhada), (b) aproximação de primeira ordem (linha contínua) e expansão até ordem $n^{-1}$ (linha pontilhada) para a função distribuição acumulada sob a hipótese nula da estatística gradiente; distribuição Birnbaum–Saunders . . . . .	20
2.2	(a) Distorção de tamanho para o teste gradiente (linha contínua) e para sua versão corrigida (linha pontilhada), (b) Aproximação de primeira ordem (linha contínua) e expansão até ordem $n^{-1}$ (linha pontilhada) da função distribuição acumulada da estatística gradiente sob a hipótese nula; distribuição gama. . . . .	21
4.1	Probabilidades de cobertura frequentista para intervalos de credibilidade obtidos pela inversão do teste gradiente utilizando como referência o quantil (4.20) (linha cheia); probabilidades de cobertura dos intervalos de confiança baseados na inversão da estatística gradiente (tracejada) e de Wald (pontilhada) para $\phi$ utilizando os quantis da distribuição $\chi_1^2$ como referência . . . . .	78



# Lista de Tabelas

3.1	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 4$ . . . . .	49
3.2	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 5$ . . . . .	50
3.3	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 6$ . . . . .	51
3.4	Taxas de rejeição não-nulas (poder) das versões corrigidas dos testes LR ( $\bar{S}_{LR}^*$ ), escore ( $\bar{S}_R^*$ ) e gradiente ( $\bar{S}_T^*$ ) no modelo de regressão gama com $\phi$ desconhecido; $\alpha = 5\%$ , $p = 4$ , $q = 2$ , $n = 30$ . . . . .	52
3.5	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 4$ . . . . .	53
3.6	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 5$ . . . . .	54

3.7	Taxas de rejeição de $\mathcal{H}_0$ no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ , $\bar{S}_R^*$ , $\bar{S}_T^*$ ); $p = 6$ . . . . .	55
3.8	Taxas de rejeição não-nulas (poder) das versões corrigidas dos testes LR ( $\bar{S}_{LR}^*$ ), escore ( $\bar{S}_R^*$ ) e gradiente ( $\bar{S}_T^*$ ) no modelo de regressão normal inverso com $\phi$ desconhecido; $\alpha = 5\%$ , $p = 6$ , $q = 2$ , $n = 30$ . . . . .	57
3.9	Estimativas dos parâmetros do modelo (3.33). . . . .	57
3.10	Valores das estatísticas e $p$ -valores para o teste da hipótese $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$ no modelo (3.33). . . . .	58
3.11	Valores das estatísticas e $p$ -valores para o teste da hipótese $\mathcal{H}_0 : \beta_5 = 0$ no modelo (3.33). . . . .	58
3.12	Estimativas do modelo (3.34). . . . .	59
3.13	Valores das estatísticas e $p$ -valores para o teste da hipótese $\mathcal{H}_0 : \beta_3 = 0$ no modelo (3.34). . . . .	59
3.14	<i>Squid data</i> . . . . .	66
3.15	<i>Breadwrapper data</i> . . . . .	67

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Preliminares . . . . .	1
1.2	O argumento de encolhimento . . . . .	4
1.3	Organização da tese . . . . .	6
<b>2</b>	<b>Correção tipo-Bartlett para a estatística gradiente</b>	<b>8</b>
2.1	Introdução . . . . .	8
2.2	Principais Resultados . . . . .	11
2.3	O caso uniparamétrico . . . . .	15
2.4	Modelos com dois parâmetros ortogonais . . . . .	18
2.5	Discussão . . . . .	22
2.6	Detalhes técnicos . . . . .	23
2.6.1	Demonstração do teorema 2.1 . . . . .	23
<b>3</b>	<b>Aperfeiçoamento do teste gradiente em modelos lineares generalizados</b>	<b>31</b>
3.1	Introdução . . . . .	32
3.2	Testes de hipóteses em MLG . . . . .	33
3.3	Aperfeiçoamento do teste gradiente . . . . .	35
3.4	Aperfeiçoamento do teste gradiente em MLG com $\phi$ conhecido . . . . .	37
3.5	Aperfeiçoamento do teste gradiente em MLG com $\phi$ desconhecido . . . . .	40
3.6	Aperfeiçoamento dos testes LR e escore em MLG com $\phi$ desconhecido . . . . .	43

3.7	Resultados Numéricos: modelo de regressão gama . . . . .	45
3.8	Resultados numéricos: modelo de regressão normal inverso . . . . .	48
3.9	Aplicações . . . . .	57
3.10	Discussão . . . . .	59
3.11	Detalhes técnicos . . . . .	60
3.11.1	Obtenção das quantidades $A_1, A_2$ e $A_3$ . . . . .	60
3.11.2	Obtenção das quantidades $A_{1,\beta\phi}$ e $A_{2,\beta\phi}$ . . . . .	63
3.11.3	Conjuntos de dados utilizados na Seção 3.9 . . . . .	66
<b>4</b>	<b>Prioris de correspondência associadas à estatística gradiente</b>	<b>68</b>
4.1	Introdução . . . . .	68
4.2	Principais resultados . . . . .	70
4.3	O caso uniparamétrico . . . . .	76
4.4	O caso biparamétrico . . . . .	77
4.5	Parâmetro de interesse unidimensional ortogonal aos parâmetros de perturbação . . . . .	79
4.6	Discussão . . . . .	81
4.7	Detalhes técnicos . . . . .	83
4.7.1	Quantidades dadas em (4.4) . . . . .	83
4.7.2	Lema auxiliar . . . . .	83
<b>5</b>	<b>Considerações finais e pesquisas futuras</b>	<b>85</b>
	<b>Referências Bibliográficas</b>	<b>87</b>

# Capítulo 1

## Introdução

### 1.1 Preliminares

Os testes mais utilizados em grandes amostras são baseados nas estatísticas da razão de verossimilhanças (Wilks, 1938), de Wald (Wald, 1943) e escore (Rao, 1948). Esses testes são os mais utilizados em diversas áreas, pela pouca disponibilidade de testes que utilizam estatísticas com distribuição exata. Recentemente, Terrell (2002) propôs um teste baseado na estatística gradiente. Esta estatística não envolve a matriz de informação nem observada, nem esperada, além de ser muito simples de ser calculada. Rao (2005) escreveu “The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic”. Da mesma forma que as estatísticas da razão de verossimilhanças, escore e de Wald, a estatística gradiente tem distribuição assintótica  $\chi^2$  sob a hipótese nula.

A estatística gradiente é tema de pesquisas recentes. Lemonte e Ferrari (2012b) fizeram o estudo do poder local do teste baseado na estatística gradiente. Concluíram que nenhum dos testes, razão de verossimilhanças, escore, de Wald e gradiente, é uniformemente mais poderoso que os outros. Artigos que fizeram o estudo do poder local dos quatro testes em classes de modelos paramétricos podem ser citados, como Lemonte (2011) para modelos não lineares da família exponencial, Lemonte e Ferrari (2011a) em modelos de regressão Birnbaum–Saunders, Lemonte (2012a) para modelos lineares generalizados com dispersão variável Lemonte e Ferrari (2012d) em modelos de dispersão.

A distribuição exata das estatísticas da razão de verossimilhanças, de Wald, escore e gradiente em amostras finitas geralmente é desconhecida. Logo os testes são baseados em aproximações assintóticas. Sob a hipótese nula, utiliza-se uma distribuição qui-quadrado como referência. No entanto, em pequenas amostras, a distribuição qui-quadrado pode não ser uma boa aproximação para a distribuição dessas estatísticas. Neste sentido, fez-se necessário na literatura propor métodos de aperfeiçoamento dos testes baseando-se em estatísticas modificadas, como a obtenção de fatores de correção de Bartlett e tipo-Bartlett.

Podemos citar vários trabalhos sobre a obtenção de fatores de Bartlett e suas aplicações, impulsionados pelo primeiro trabalho nesta linha, em que Bartlett (1937) determinou uma estatística da razão de verossimilhanças corrigida, obtida através da esperança da estatística sob a hipótese nula até ordem  $o(n^{-1})$ , em que  $n$  é o tamanho da amostra. Bartlett (1937, 1947, 1954) determinou fatores de correção na linha de análise multivariada a fim de melhorar a distribuição assintótica sob a hipótese nula da estatística da razão de verossimilhanças em grandes amostras. Lawley (1956) apresentou uma fórmula geral para o primeiro momento da estatística da razão de verossimilhanças envolvendo expressões muito gerais que envolvem momentos do logaritmo da função de verossimilhança e suas derivadas. Uma grande contribuição foi o trabalho de Hayakawa (1977), que determinou uma expansão até ordem  $n^{-1}$  para a função distribuição acumulada da estatística da razão de verossimilhanças.

Em relação ao aperfeiçoamento do teste escore, Taniguchi (1991) determinou uma estatística escore aperfeiçoada para o caso uniparamétrico, e Chandra e Mukerjee (1991) para o caso multiparamétrico, porém, não contemplando casos mais gerais como hipóteses compostas, onde há mais de um parâmetro de perturbação. Cordeiro e Ferrari (1991) desenvolveram uma metodologia geral para a correção de testes baseados em estatísticas que possuem distribuição assintótica qui-quadrado. Essa correção resulta em uma estatística modificada por um fator, que é um polinômio na própria estatística cujos coeficientes são funções de momentos do logaritmo da função de verossimilhança. A estatística modificada possui distribuição qui-quadrado até ordem  $o(n^{-1})$ . Esse fator de correção não é um fator de Bartlett genuíno, pois envolve a própria estatística. Portanto, as correções dessas estatísticas são chamadas de correções tipo-Bartlett. O resultado em Cordeiro e Ferrari (1991) é muito geral, pois

inclui a presença de mais de um parâmetro de perturbação. No entanto, esta metodologia necessita da função distribuição da estatística sob a hipótese nula até um erro de ordem  $o(n^{-1})$ . Assim, os autores propuseram uma estatística escore modificada que tem distribuição qui-quadrado até ordem  $n^{-1}$ , a partir da expansão sob a hipótese nula da função distribuição da estatística escore, obtida por Harris (1985).

Um dos objetivos desta tese é obter um fator de correção tipo-Bartlett para a estatística gradiente, obtido através de uma expansão sob a hipótese nula para a função distribuição acumulada da estatística e utilizando os resultados de Cordeiro e Ferrari (1991). Esta correção resulta em uma estatística modificada, que tem distribuição qui-quadrado com erro de ordem  $o(n^{-1})$ . A expansão sob a hipótese nula para a distribuição da estatística gradiente é obtida nesta tese através do argumento de encolhimento proposto por Ghosh e Mukerjee (1991) e descrito em detalhes em Mukerjee e Reid (2000) (ver Seção 1.2). Essa técnica se baseia em uma rota Bayesiana, cujo objetivo é determinar a esperança de funções mensuráveis de variáveis aleatórias, e reduz substancialmente os cálculos envolvidos em problemas frequentistas como determinação de expansões para probabilidades caudais (Mukerjee e Reid, 2000), expansões assintóticas para estatísticas usuais, fatores de Bartlett (Ghosh e Mukerjee, 1991) e estudo de poder de testes (Ghosh e Mukerjee, 1994), sendo uma alternativa às expansões de Edgeworth.

Aplicações da correção tipo-Bartlett para a estatística gradiente são apresentadas na presente tese, a saber, determinação de fórmulas matriciais simples para o fator de correção tipo-Bartlett da estatística gradiente em modelos lineares generalizados (MLG) com dispersão conhecida e desconhecida. Para isso, foi seguida a mesma linha de raciocínio descrita em Cordeiro (1983, 1987) para a derivação de fórmulas matriciais para fatores de Bartlett da estatística da razão de verossimilhanças em MLG, e de Cordeiro, Ferrari e Paula (1993) para os fatores de correção tipo-Bartlett para a estatística escore nesses modelos.

O argumento de encolhimento é uma técnica crucial para o desenvolvimento de *matching priors*, que são *prioris* nas quais se baseiam regiões de credibilidade que possuem probabilidades de cobertura próximas às de regiões frequentistas (Datta e Mukerjee, 2003). Uma das formas de obter essas

regiões é através da inversão dos testes estatísticos mais comuns, utilizando como referência quantis ajustados que dependem dessas prioris e dos dados. Alguns trabalhos recentes utilizam resultados obtidos através desta técnica. Um exemplo é na abordagem de dados dependentes, como em Chang e Mukerjee (2010), em que os autores estabelecem condições para obtenção de *matching priors* para construção de regiões HPD (*Highest posterior density regions*) e Chang e Mukerjee (2011) que obtiveram regiões de credibilidade baseadas na inversão de um teste da razão de verossimilhanças ajustado. Em análise Bayesiana objetiva, DiCiccio, Kuffner e Young (2012) caracterizaram propriedades Bayesianas da estatística da razão de verossimilhanças sinalizada utilizando essa abordagem. Ventura, Cabras e Racugno (2009) caracterizaram *matching priors* para a obtenção de intervalos de credibilidade com probabilidades de cobertura frequentista acuradas a partir de distribuições a posteriori para um parâmetro de interesse baseados em verossimilhança perfilada, independentemente de modelos específicos. Trabalhos surgiram da aplicação desses resultados em modelos específicos como Min e Sun (2012) e Cabras et al. (2012). Ventura, Sartori e Racugno (2013) propuseram aproximações de alta ordem para a distribuição marginal a posteriori para um parâmetro escalar de interesse baseadas em *matching priors*, e avaliaram as probabilidades de cobertura de intervalos de credibilidade obtidos através dessas distribuições a posteriori em vários modelos específicos.

Nesta tese, utilizamos o argumento de encolhimento para a derivação de *matching priors* que determinam regiões de credibilidade via inversão da estatística gradiente que possuem probabilidades de cobertura frequentista acuradas.

## 1.2 O argumento de encolhimento

Seja  $\mathbf{x} = (x_1, \dots, x_n)^\top$  um vetor aleatório que possui função densidade de probabilidade  $f(\cdot; \boldsymbol{\theta})$  que depende de um vetor de parâmetros  $p$ -dimensional  $\boldsymbol{\theta} \in \Theta$ , onde  $\Theta \subseteq \mathbb{R}^p$  é um subconjunto aberto do espaço Euclidiano. Seja  $Q(\cdot, \boldsymbol{\theta})$  uma função mensurável. Assumimos que  $Q$  é contínua para todo  $\boldsymbol{\theta}$  e que sua esperança existe, ou seja, é finita para todo ponto do espaço paramétrico. Uma rota Bayesiana para obter  $E_{\boldsymbol{\theta}}\{Q(\cdot, \boldsymbol{\theta})\}$  baseada em um argumento de encolhimento envolve os três passos descritos



a seguir.

**Passo 1.** Obtém-se a esperança a posteriori de  $Q(\cdot)$ ,  $E^\pi\{Q(\boldsymbol{\theta}, \mathbf{X})|\mathbf{X} = \mathbf{x}\}$ , sob uma densidade a priori  $\pi(\cdot)$  para  $\boldsymbol{\theta}$ . Essa densidade a priori deve ser própria, ter suporte compacto no espaço paramétrico, ser positiva no interior do seu suporte, ser nula e ter a primeira derivada parcial nula na fronteira do seu suporte.

**Passo 2.** Determina-se  $E_\theta E^\pi\{Q(\boldsymbol{\theta}, \mathbf{X})|\mathbf{X} = \mathbf{x}\} = \Delta(\boldsymbol{\theta})$ , para  $\boldsymbol{\theta}$  no interior do suporte de  $\pi(\cdot)$ .

**Passo 3.** Integra-se  $\Delta(\boldsymbol{\theta})$  com respeito a  $\pi(\cdot)$  e faz-se  $\pi(\cdot)$  convergir em distribuição para a distribuição a priori degenerada no verdadeiro valor de  $\boldsymbol{\theta}$ , onde  $\boldsymbol{\theta}$  está no interior do suporte de  $\pi(\cdot)$ , obtendo, assim,  $E_\theta\{Q(\mathbf{X}, \boldsymbol{\theta})\}$ .

Justificamos agora os passos descritos acima. A densidade a posteriori  $\pi_{post}(\cdot)$  de  $\boldsymbol{\theta}$  sob a priori  $\pi(\cdot)$  é dada por

$$\pi_{post}(\boldsymbol{\theta}|\mathbf{X} = \mathbf{x}) = \frac{f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Portanto, o **Passo 1** fornece

$$E^\pi\{Q(\boldsymbol{\theta}, \mathbf{X})|\mathbf{X} = \mathbf{x}\} = \frac{\int Q(\mathbf{x}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Tomando a esperança da expressão anterior, como descrito no **Passo 2**, temos

$$E_\theta E^\pi\{Q(\boldsymbol{\theta}, \mathbf{X})|\mathbf{X} = \mathbf{x}\} = \Delta(\boldsymbol{\theta}) = \int \left\{ \frac{\int Q(\mathbf{x}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \right\} f(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x}.$$

Em seguida, integra-se  $\Delta(\cdot)$  com respeito a  $\pi$ , como no **Passo 3**,

$$\int \Delta(\boldsymbol{\theta})d\boldsymbol{\theta} = \int \int \left\{ \frac{\int Q(\mathbf{x}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \right\} f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\mathbf{x}d\boldsymbol{\theta}.$$

Como  $Q(\cdot, \cdot)$  é uma função integrável, ou seja, sua esperança existe, podemos trocar a ordem de integração baseando-se no Teorema de Fubini, o que resulta em

$$\begin{aligned}\int \Delta(\boldsymbol{\theta})d\boldsymbol{\theta} &= \int \left\{ \frac{\int Q(\mathbf{x}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \right\} \left\{ \int f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \right\} d\mathbf{x} \\ &= \int \int Q(\mathbf{x}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\mathbf{x} \\ &= \int \left\{ \int Q(\mathbf{y}, \boldsymbol{\theta})f(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x} \right\} \pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int E_{\boldsymbol{\theta}}\{Q(\mathbf{x}, \boldsymbol{\theta})\}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}.\end{aligned}$$

Quando se faz  $\pi(\cdot)$  convergir em distribuição para a priori degenerada no verdadeiro valor de  $\boldsymbol{\theta}$ , a última integral desaparece pois esta é em relação à medida de probabilidade  $\pi(\cdot)$ , que passa a ser uma medida definida em um espaço que contém somente um ponto, a saber, o verdadeiro valor de  $\boldsymbol{\theta}$ .

Na descrição da técnica do argumento de encolhimento fica bem claro o seu objetivo: determinação de esperança de funções mensuráveis de variáveis aleatórias. Mukerjee e Reid (2000) utilizam essa rota Bayesiana para obter uma expansão da função distribuição acumulada do estimador de máxima verossimilhança, simplificando de maneira substancial os cálculos envolvidos no método usual, que são as expansões de Edgeworth. Utilizando essas idéias, é possível obter expansões de funções características das estatísticas usuais e propor fatores de correção para estas estatísticas a fim de melhorar seus desempenhos em amostras pequenas, como em Ghosh e Mukerjee (1991).

Para a obtenção de regiões de credibilidade acuradas via inversão de testes estatísticos esta técnica também é útil. Rao e Mukerjee (1995) determinam regiões de credibilidade com probabilidades de cobertura frequentista acuradas baseando-se na inversão das estatísticas score e de Wald.

### 1.3 Organização da tese

No que se segue, a tese contém três capítulos em que o tema comum é a investigação e a aplicação de metodologias baseadas na estatística gradiente. Na introdução de cada capítulo, é feita uma revisão bibliográfica de cada tema. Cada capítulo é independente dos demais, no sentido de que são introduzidas notações e comentários intrínsecos a cada um deles. Somente nos Capítulos 3 e 4 são citados resultados do Capítulo 2, mas a função de cada um desses capítulos é independente. Nesta

tese abordamos três tópicos relativos à estatística gradiente.

No Capítulo 2 obtemos uma expansão assintótica sob a hipótese nula para a função de distribuição da estatística gradiente para testar hipóteses nulas na possível presença de parâmetros de perturbação. A obtenção dessa expansão é baseada no argumento de encolhimento descrito na Seção 1.2. Uma estatística gradiente corrigida com distribuição qui-quadrado até um erro de ordem  $n^{-1}$  é proposta. Alguns exemplos e simulações são apresentados.

No Capítulo 3 caracterizamos uma expansão para a função distribuição da estatística gradiente em modelos lineares generalizados (MLG) até ordem  $n^{-1}$  sob a hipótese nula. Para isto, foram utilizados os principais resultados do Capítulo 2. O objetivo desse capítulo é derivar fórmulas matriciais de fácil computação para as quantidades que compõem essa expansão. A partir daí, obtém-se uma estatística gradiente corrigida, com distribuição  $\chi^2$  até ordem  $n^{-1}$ . Estudos de simulação com o intuito de verificar o desempenho do teste baseado nessa estatística corrigida e compará-la com outros testes existentes na literatura são apresentados. São apresentadas também aplicações utilizando conjuntos de dados reais.

No Capítulo 4, obtemos densidades a priori, nas quais se baseiam regiões de credibilidade dadas pela inversão da estatística gradiente com boas propriedades frequentistas. Tais prioris são chamadas de prioris de correspondência. Neste capítulo, utilizamos uma parte do desenvolvimento dos resultados do Capítulo 2 para a caracterização de condições para a obtenção dessas regiões. Estas condições correspondem a sistemas de equações diferenciais parciais, envolvendo momentos do logaritmo da função de verossimilhanças e cujas soluções são as prioris de correspondência. A partir daí, temos um quantil modificado que depende dessas prioris e dos dados, com o qual é possível construir essas regiões. Uma propriedade importante dessas regiões de credibilidade é que elas possuem propriedades de cobertura frequentista acuradas. Casos especiais e exemplos são apresentados.

O Capítulo 5 destina-se às conclusões finais e sugestões de pesquisas futuras.

# Capítulo 2

## Correção tipo-Bartlett para a estatística gradiente

### Resumo

Neste capítulo obtemos uma expansão assintótica, sob a hipótese nula, para a função de distribuição da estatística gradiente para testar hipóteses nulas compostas na presença de parâmetros de perturbação. A expansão é derivada utilizando uma rota Bayesiana baseada no argumento de encolhimento descrito em Ghosh e Mukerjee (1991); ver também Seção 1.2. Utilizando essa expansão, propomos uma estatística gradiente corrigida por um fator de correção tipo-Bartlett que possui distribuição qui-quadrado até um erro de ordem  $o(n^{-1})$  sob a hipótese nula. Além disso, utilizamos a expansão para modificar os percentis da distribuição qui-quadrado de referência em grandes amostras. Simulações Monte Carlo e vários exemplos são apresentados e discutidos.

*Palavras-chave:* expansão assintótica, rota Bayesiana, estatística gradiente, argumento de encolhimento.

### 2.1 Introdução

Os testes mais comuns utilizados para amostras de tamanho grande são os testes da razão de verossimilhanças (Wilks, 1938), Wald (Wald, 1943) e escore (Rao, 1948). Estes testes são ampla-

mente utilizados em áreas como economia, biologia e engenharia, entre outros, pois testes exatos não estão sempre disponíveis. Um teste alternativo utiliza a estatística gradiente proposta recentemente por Terrell (2002). Uma vantagem da estatística gradiente sobre as estatísticas de Wald e escore é que não depende da matriz de informação, nem esperada e nem observada. Além disso, a estatística gradiente é muito simples de ser computada. Este fato foi enfatizado por Rao (Rao, 2005), que escreveu: “The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic”.

Sejam  $x_1, \dots, x_n$  uma amostra aleatória de tamanho  $n$  em que cada  $x_i$  possui função densidade de probabilidade  $f(\cdot; \boldsymbol{\theta})$ , que depende de um vetor  $p$ -dimensional de parâmetros desconhecidos  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ . Seja  $\ell(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta})$  e  $\mathbf{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$  o logaritmo da função de verossimilhança e o vetor escore, respectivamente; note que, por conveniência, ambos estão divididos por  $n$ . Queremos testar a hipótese nula  $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$  contra a hipótese alternativa  $\mathcal{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{10}$ , onde  $\boldsymbol{\theta}_{10}$  é um vetor  $q$ -dimensional fixado,  $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_q)^\top$  e  $\boldsymbol{\theta}_2 = (\theta_{q+1}, \dots, \theta_p)^\top$ . A partição de  $\boldsymbol{\theta}$  induz a correspondente partição em  $\mathbf{U}(\boldsymbol{\theta})$ :  $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_1(\boldsymbol{\theta})^\top, \mathbf{U}_2(\boldsymbol{\theta})^\top)^\top$ . Sejam  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top)^\top$  e  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)^\top$  os estimadores de máxima verossimilhança irrestrito e restrito (sob  $\mathcal{H}_0$ ) de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ , respectivamente. A estatística gradiente para testar  $\mathcal{H}_0$  é definida como

$$S = n\mathbf{U}(\tilde{\boldsymbol{\theta}})^\top (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \quad (2.1)$$

e pode ser escrita como  $S = n\mathbf{U}_1(\tilde{\boldsymbol{\theta}})^\top (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})$ , pois  $\mathbf{U}_2(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ . Assim como as estatísticas da razão de verossimilhanças, de Wald e escore, a estatística gradiente tem distribuição assintótica  $\chi_q^2$  sob a hipótese nula, onde  $q$  é o número de restrições impostas por  $\mathcal{H}_0$ .

Embora a estatística gradiente tenha sido derivada por Terrell (2002) a partir das estatísticas escore e de Wald, ela é de natureza diferente. A estatística escore mede o quadrado do comprimento do vetor escore avaliado em  $\mathcal{H}_0$  utilizando a métrica dada pela inversa da matriz de informação de Fisher, enquanto a estatística de Wald determina o quadrado da distância entre os estimadores de máxima verossimilhança irrestrito e restrito de  $\boldsymbol{\theta}$  utilizando a métrica dada pela matriz de informação de Fisher. Além disso, ambas são formas quadráticas. Por outro lado, a estatística gradiente não é uma forma quadrática e mede a distância entre os estimadores de máxima verossimilhança irrestrito e

restrito de  $\theta$  a partir de uma perspectiva diferente. Esta mede a projeção ortogonal do vetor escore em  $\mathcal{H}_0$  sobre o vetor  $\hat{\theta} - \tilde{\theta}$ . Ao contrário da estatística escore e assim como a estatística de Wald, a estatística gradiente não é invariante sob reparametrização do modelo que preserva o parâmetro de interesse.

Recentemente, o teste gradiente tem sido objeto de alguns artigos de pesquisa. Em particular, Lemonte e Ferrari (2012b) obtiveram o poder local do teste gradiente sob alternativas de Pitman (uma sequência de hipóteses alternativas convergindo para a hipótese nula com taxa  $n^{-1/2}$ ); ver também Lemonte e Ferrari (2012c). Os autores compararam o poder local do teste gradiente com o dos testes da razão de verossimilhanças, de Wald e escore. Provaram que nenhum dos testes é uniformemente mais poderoso que os outros, e portanto, o teste gradiente não só é mais simples de ser calculado como também é competitivo em relação aos outros em termos de poder local. Comparações entre o poder local dos testes clássicos e o teste gradiente em modelos de regressão podem ser vistos em Lemonte (2011, 2012).

O principal resultado em Lemonte e Ferrari (2012b) em relação ao poder local do teste gradiente até um erro de ordem  $o(n^{-1/2})$  representa o primeiro passo no estudo de propriedades assintóticas de alta ordem do teste gradiente. No presente capítulo, focamos em derivar uma aproximação de segunda ordem para a distribuição sob a hipótese nula da estatística gradiente. Em outras palavras, o nosso objetivo é a obtenção de uma expansão assintótica para a função distribuição acumulada da estatística gradiente sob a hipótese nula com um erro de ordem  $o(n^{-1})$ .

A rota usual de derivação de expansões para a distribuição de estatísticas de teste que possuem distribuição assintótica qui-quadrado envolve expansões em séries de Edgeworth multivariadas. Embora essa rota tenha sido seguida por muitos autores, é extremamente longa e tediosa (ver, por exemplo, Hayakawa, 1977; Harris, 1985). Aqui, por outro lado, a fim de derivar uma expansão assintótica para a distribuição sob a hipótese nula da estatística gradiente até ordem  $n^{-1}$ , seguimos uma rota Bayesiana baseada em um argumento de encolhimento originalmente sugerida por Ghosh e Mukerjee (1991) e descrita posteriormente em Mukerjee e Reid (2000) (ver Seção 1.2). Embora utilize uma abordagem Bayesiana, essa técnica pode ser utilizada para resolver problemas frequentistas, como a

derivação de correções de Bartlett e probabilidades caudais (Datta e Mukerjee, 2003).

Adicionalmente, obtivemos um fator de correção tipo-Bartlett para a estatística gradiente a partir de resultados dados em Cordeiro e Ferrari (1991). Sob a hipótese nula, a estatística corrigida tem distribuição qui-quadrado até um erro de ordem  $o(n^{-1})$ , enquanto a estatística gradiente não corrigida tem distribuição qui-quadrado até um erro de ordem  $o(n^{-1/2})$ ; isto é, o fator de correção tipo-Bartlett faz o erro de aproximação ser reduzido de  $o(n^{-1/2})$  para  $o(n^{-1})$ . Para um levantamento detalhado sobre correções de Bartlett e tipo-Bartlett, sugerimos a leitura de Cordeiro e Cribari-Neto (1996).

Este capítulo se divide como se segue. Na Seção 2.2, apresentamos nossos principais resultados, a saber, uma expansão assintótica para a função distribuição acumulada da estatística gradiente e sua correção tipo-Bartlett. Nas Seções 2.3 e 2.4, particularizamos nossos resultados gerais para famílias uniparamétricas e famílias com dois parâmetros ortogonais, respectivamente. Simulações de Monte Carlo são apresentadas na Seção 2.4. A Seção 2.5 fecha o capítulo com uma breve discussão. Detalhes técnicos são apresentados na Seção 2.6.

## 2.2 Principais Resultados

Primeiramente, apresentamos algumas notações. Seja  $D_j = \partial/\partial\theta_j$  ( $j = 1, \dots, p$ ) o operador diferencial. Definimos  $U_j = D_j\ell(\boldsymbol{\theta})$ ,  $U_{jr} = D_j D_r \ell(\boldsymbol{\theta})$ ,  $U_{jrs} = D_j D_r D_s \ell(\boldsymbol{\theta})$ , e  $U_{jr su} = D_j D_r D_s D_u \ell(\boldsymbol{\theta})$ . Fazemos algumas suposições, tais como a regularidade das quatro primeiras derivadas de  $\ell(\boldsymbol{\theta})$  com respeito a  $\boldsymbol{\theta}$  e a existência e unicidade do estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ , como outras descritas em Hayakawa (1977). Seja  $\kappa_{jr} = E(U_{jr})$ ,  $\kappa_{jrs} = E(U_{jrs})$ ,  $\kappa_{jr su} = E(U_{jr su})$ ,  $\kappa_{jr}^{(s)} = D_s \kappa_{jr}$ ,  $\kappa_{jr}^{(su)} = D_s D_u \kappa_{jr}$ , e  $\kappa_{jrs}^{(u)} = D_u \kappa_{jrs}$ . Note que todos os  $\kappa$ 's são de ordem  $O(1)$ . Adicionalmente, seja  $\mathbf{K}$  a matriz de informação de Fisher por observação

$$\mathbf{K} = -((\kappa_{jr})) = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix},$$

com  $\mathbf{K}^{-1} = -((\kappa^{jr}))$  denotando sua inversa. Finalmente, definimos as matrizes

$$\mathbf{A} = ((a^{jr})) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{22}^{-1} \end{bmatrix}, \quad \mathbf{M} = ((m^{jr})) = \mathbf{K}^{-1} - \mathbf{A}.$$

No que se segue, usamos a convenção de Einstein para somatórios em que  $\sum'$  denota o somatório sobre todos os componentes de  $\boldsymbol{\theta}$ ; isto é, os índices  $j, r, s, k, l$  e  $u$  variando de 1 a  $p$ . Enunciamos agora o seguinte teorema.

**Teorema 2.1** *Uma expansão assintótica para a distribuição sob a hipótese nula da estatística gradiente para testar  $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$  contra  $\mathcal{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{10}$  é dada por*

$$\Pr(S \leq x) = G_q(x) + \frac{1}{24n} \sum_{i=0}^3 R_i G_{q+2i}(x) + o(n^{-1}), \quad (2.2)$$

onde  $G_z(x)$  é a função distribuição acumulada de uma variável aleatória qui-quadrado com  $z$  graus de liberdade,  $R_1 = 3A_3 - 2A_2 + A_1$ ,  $R_2 = A_2 - 3A_3$ ,  $R_3 = A_3$ ,  $R_0 = -(R_1 + R_2 + R_3)$ ,

$$\begin{aligned} A_1 = & 3 \sum' \kappa_{jrs} \kappa_{klu} [m^{jr} a^{lu} (m^{sk} + 2a^{sk}) + a^{jr} m^{sk} a^{lu} + 2m^{jk} a^{rl} a^{su}] \\ & - 12 \sum' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} (\kappa^{sj} \kappa^{rk} \kappa^{lu} + a^{sj} a^{rk} a^{lu} + \kappa^{sk} \kappa^{lj} \kappa^{ru} + a^{sk} a^{lj} a^{ru}) \\ & - 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [(a^{su} - \kappa^{su}) (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (a^{sk} a^{lu} + \kappa^{sk} \kappa^{lu}) \\ & \quad + 2a^{rs} (\kappa^{jk} \kappa^{lu} - a^{jk} a^{lu}) + 2a^{rk} a^{ls} m^{ju}] \\ & + 6 \sum' \kappa_{jr su} m^{jr} a^{su} - 6 \sum' \kappa_{jrs}^{(u)} [m^{jr} (a^{su} - \kappa^{su}) + 2m^{ju} a^{rs}] \\ & + 12 \sum' \kappa_{rs}^{(ju)} (\kappa^{jr} \kappa^{su} - a^{jr} a^{su}), \end{aligned}$$

$$\begin{aligned} A_2 = & -3 \sum' \kappa_{jrs} \kappa_{klu} [m^{jr} m^{sk} a^{lu} + m^{jr} a^{sk} m^{lu} + 2m^{jk} m^{rl} a^{su} \\ & \quad + \frac{1}{4} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su})] \\ & + 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [m^{su} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (\kappa^{sk} \kappa^{lu} - a^{sk} a^{lu})] \\ & + 6 \sum' \kappa_{jrs}^{(u)} m^{jr} m^{su} - 3 \sum' \kappa_{jr su} m^{jr} m^{su}, \end{aligned}$$

$$A_3 = \frac{1}{4} \sum' \kappa_{jrs} \kappa_{klu} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su}).$$



*Prova.* A prova é apresentada na Seção 2.6.

□

Basicamente, para provar o Teorema 2.1, seguimos uma rota Bayesiana baseada em um argumento de encolhimento. Esse argumento é descrito na Seção 1.2.

Se a hipótese nula é simples, temos  $q = p$ ,  $\mathbf{A} = \mathbf{0}$  e  $\mathbf{M} = \mathbf{K}^{-1}$ . Portanto, uma consequência imediata do Teorema 2.1 é o seguinte corolário.

**Corolário 2.1** *Uma expansão assintótica para a distribuição sob a hipótese nula da estatística gradiente para testar  $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  contra  $\mathcal{H}_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$  é dada por (2.2) com  $q = p$ ,  $R_1 = 3A_3 - 2A_2 + A_1$ ,  $R_2 = A_2 - 3A_3$ ,  $R_3 = A_3$ ,  $R_0 = -(R_1 + R_2 + R_3)$  onde os  $A$ 's são dados por*

$$\begin{aligned} A_1 &= -12 \sum' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} (\kappa^{sj} \kappa^{rk} \kappa^{lu} + \kappa^{sk} \kappa^{lj} \kappa^{ru}) \\ &\quad + 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} (\kappa^{su} \kappa^{jk} \kappa^{lr} + \kappa^{jr} \kappa^{sk} \kappa^{lu}) \\ &\quad + 12 \sum' \kappa_{rs}^{(ju)} \kappa^{jr} \kappa^{su} - 6 \sum' \kappa_{jrs}^{(u)} \kappa^{jr} \kappa^{su}, \\ A_2 &= \frac{3}{4} \sum' \kappa_{jrs} \kappa_{klu} (3\kappa^{jr} \kappa^{sk} \kappa^{lu} + 2\kappa^{jk} \kappa^{rl} \kappa^{su}) \\ &\quad - 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} (\kappa^{su} \kappa^{jk} \kappa^{lr} + \kappa^{jr} \kappa^{sk} \kappa^{lu}) \\ &\quad + 6 \sum' \kappa_{jrs}^{(u)} \kappa^{jr} \kappa^{su} - 3 \sum' \kappa_{jrsu} \kappa^{jr} \kappa^{su}, \\ A_3 &= -\frac{1}{4} \sum' \kappa_{jrs} \kappa_{klu} (3\kappa^{jr} \kappa^{sk} \kappa^{lu} + 2\kappa^{jk} \kappa^{rl} \kappa^{su}). \end{aligned}$$

Podemos agora apresentar uma correção tipo-Bartlett para a estatística gradiente. Uma correção tipo-Bartlett é um fator multiplicativo, que depende da própria estatística e resulta em uma estatística modificada que segue uma distribuição qui-quadrado com erro de aproximação de ordem inferior a  $n^{-1}$ . Cordeiro e Ferrari (1991) obtiveram uma fórmula geral de correção tipo-Bartlett para uma ampla classe de estatísticas que possuem distribuição assintótica qui-quadrado. Um caso especial é quando a função distribuição acumulada da estatística pode ser escrita como em (2.2), independentemente dos coeficientes  $R_1$ ,  $R_2$ , e  $R_3$ . Assim, do Teorema 2.1 e dos resultados de Cordeiro e Ferrari (1991), temos o seguinte corolário.

**Corolário 2.2** *A estatística modificada*

$$S^* = S[1 - (c + bS + aS^2)], \quad (2.3)$$

onde

$$a = \frac{A_3}{12nq(q+2)(q+4)}, \quad b = \frac{A_2 - 2A_3}{12nq(q+2)}, \quad c = \frac{A_1 - A_2 + A_3}{12nq},$$

possui uma distribuição  $\chi_q^2$  até um erro de ordem  $o(n^{-1})$  sob a hipótese nula.

O fator  $[1 - (c + bS + aS^2)]$  em (2.3) pode ser considerado como um fator de correção tipo-Bartlett para a estatística gradiente de tal forma que a distribuição sob a hipótese nula de  $S^*$  é melhor aproximada pela distribuição  $\chi^2$  de referência do que a distribuição da estatística gradiente não corrigida.

Ao invés de modificar a estatística de teste como em (2.3), podemos modificar a distribuição  $\chi^2$  de referência invertendo a fórmula da expansão (2.2) (Hill e Davis, 1968). Para ser mais específico, seja  $\gamma$  o nível desejado do teste e  $x_{1-\gamma}$  o percentil  $1 - \gamma$  da distribuição limite  $\chi^2$  da estatística de teste. Da expansão (2.2), temos o seguinte corolário.

**Corolário 2.3** *Uma expansão assintótica para o percentil  $1 - \gamma$  de  $S$  tem a forma*

$$z_{1-\gamma} = x_{1-\gamma} + \frac{1}{12n} \left[ \frac{A_3 x_{1-\gamma}}{q(q+2)(q+4)} \{x_{1-\gamma}^2 + (q+4)x_{1-\gamma} + (q+2)(q+4)\} \right. \\ \left. + \frac{x_{1-\gamma}(x_{1-\gamma} + q + 2)}{q(q+2)} (A_2 - 3A_3) + \frac{x_{1-\gamma}}{q} (3A_3 - 2A_2 + A_1) \right] + o(n^{-1}), \quad (2.4)$$

onde  $\Pr(\chi_q^2 \geq x_{1-\gamma}) = \gamma$ .

Em geral, (2.3) e (2.4) dependem de parâmetros desconhecidos. Nesse caso, substituímos esses parâmetros desconhecidos por seus estimadores de máxima verossimilhança sob  $\mathcal{H}_0$ . Deve ser notado que a melhoria do teste gradiente sob a hipótese nula  $\mathcal{H}_0$  pode ser realizada de três maneiras: (i) Utilizando a estatística corrigida  $S^*$  em (2.3) e a distribuição  $\chi_q^2$  como referência; (ii) Usando a estatística gradiente  $S$  e sua distribuição acumulada aproximada (2.2); (iii) comparando  $S$  com o percentil superior modificado em (2.4). Estes três procedimentos são equivalentes até ordem  $n^{-1}$ .

Finalmente, os três momentos, até ordem  $n^{-1}$  sob a hipótese nula, da estatística gradiente são apresentados no seguinte corolário.

**Corolário 2.4** *A média, a variância, e o terceiro momento central, até ordem  $n^{-1}$  sob a hipótese nula, da estatística gradiente são*

$$\begin{aligned}\mu'_1(S) &= q + \frac{A_1}{12n}, & \mu_2(S) &= 2q + \frac{A_1 + A_2}{3n}, \\ \mu_3(S) &= 8q + \frac{2(A_1 + 2A_2 + A_3)}{n},\end{aligned}$$

*respectivamente.*

Nas próximas seções, consideramos algumas aplicações dos resultados gerais derivados nesta seção em dois casos especiais: um modelo uniparamétrico e um modelo biparamétrico sob ortogonalidade dos parâmetros.

## 2.3 O caso uniparamétrico

Inicialmente, assumimos que o modelo é indexado por um parâmetro escalar desconhecido,  $\phi$  digamos. O interesse está em testar a hipótese nula  $\mathcal{H}_0 : \phi = \phi_0$  contra  $\mathcal{H}_a : \phi \neq \phi_0$ , onde  $\phi_0$  é um valor fixado. Sejam  $\kappa_{\phi\phi} = E(\partial^2 \ell(\phi)/\partial \phi^2)$ ,  $\kappa_{\phi\phi\phi} = E(\partial^3 \ell(\phi)/\partial \phi^3)$ ,  $\kappa_{\phi\phi\phi\phi} = E(\partial^4 \ell(\phi)/\partial \phi^4)$ ,  $\kappa_{\phi\phi}^{(\phi)} = \partial \kappa_{\phi\phi}/\partial \phi$ ,  $\kappa_{\phi\phi\phi}^{(\phi)} = \partial \kappa_{\phi\phi\phi}/\partial \phi$ , e  $\kappa_{\phi\phi\phi}^{(\phi\phi)} = \partial^2 \kappa_{\phi\phi\phi}/\partial \phi^2$ . A estatística gradiente para testar  $\mathcal{H}_0$  é  $S = nU(\phi_0)(\hat{\phi} - \phi_0)$ , onde  $\hat{\phi}$  é o estimador de máxima verossimilhança de  $\phi$ . Aqui,  $A_1$ ,  $A_2$ , e  $A_3$  dados no Corolário 2.1 reduzem-se a

$$A_1 = \frac{6\kappa_{\phi\phi}(2\kappa_{\phi\phi}^{(\phi\phi)} - \kappa_{\phi\phi\phi\phi}^{(\phi)}) + 12\kappa_{\phi\phi}^{(\phi)}(\kappa_{\phi\phi\phi\phi} - 2\kappa_{\phi\phi\phi}^{(\phi)})}{\kappa_{\phi\phi}^3}, \quad (2.5)$$

$$A_2 = \frac{12\kappa_{\phi\phi}(2\kappa_{\phi\phi\phi}^{(\phi)} - 3\kappa_{\phi\phi\phi\phi}) + 3\kappa_{\phi\phi\phi}^{(\phi)}(5\kappa_{\phi\phi\phi\phi} - 16\kappa_{\phi\phi\phi}^{(\phi)})}{4\kappa_{\phi\phi}^3}, \quad (2.6)$$

$$A_3 = -\frac{5\kappa_{\phi\phi\phi\phi}^2}{4\kappa_{\phi\phi}^3}. \quad (2.7)$$

Apresentamos agora alguns exemplos.

**Exemplo 2.1** *(Distribuição exponencial)*

Sejam  $x_1, \dots, x_n$  uma amostra aleatória da distribuição exponencial com densidade

$$f(x; \phi) = \frac{1}{\phi} e^{-x/\phi}, \quad x > 0, \quad \phi > 0.$$

Aqui,  $\kappa_{\phi\phi} = -\phi^{-2}$ ,  $\kappa_{\phi\phi\phi} = 4\phi^{-3}$  e  $\kappa_{\phi\phi\phi\phi} = -18\phi^{-4}$ . A estatística gradiente é dada por  $S = n(\bar{x} - \phi_0)^2/\phi_0^2$ , onde  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ , igual à estatística escore. É fácil ver que  $A_1 = 0$ ,  $A_2 = 18$ , e  $A_3 = 20$ . Os três primeiros momentos (até ordem  $n^{-1}$ ) de  $S$  são  $\mu_1'(S) = 1$ ,  $\mu_2(S) = 2 + 6/n$ , e  $\mu_3(S) = 8 + 112/n$ . Uma verificação parcial de nossos resultados pode ser realizada pela comparação dos três momentos exatos de  $S$  com os três momentos aproximados dados acima. Uma vez que  $n\bar{x}$  tem distribuição gamma com parâmetros  $n$  e  $1/(n\phi)$ , pode ser provado que os três momentos exatos de  $S$  são  $1$ ,  $2 + 6/n$  e  $8 + 112/n + 120/n^2$  respectivamente. Esses momentos diferem dos momentos aproximados obtidos do Corolário 2.4 por termos de ordem menor que  $n^{-1}$ . A estatística gradiente corrigida pelo fator de correção tipo-Bartlett obtida do Corolário 2.3 é  $S^* = S\{1 - (3 - 11S + 2S^2)/(18n)\}$ .

**Exemplo 2.2** (*Família exponencial uniparamétrica*)

Seja  $x_1, \dots, x_n$  uma amostra aleatória de tamanho  $n$  na qual cada  $x_i$  tem uma distribuição na família exponencial uniparamétrica com densidade

$$f(x; \phi) = \frac{1}{\xi(\phi)} \exp \{-\alpha(\phi)d(x) + v(x)\},$$

onde  $\alpha(\cdot)$ ,  $v(\cdot)$ ,  $d(\cdot)$  e  $\xi(\cdot)$  são funções conhecidas. Também, assumimos que  $\alpha(\cdot)$  e  $\xi(\cdot)$  possuem as três primeiras derivadas contínuas, com  $\xi(\cdot) > 0$ ,  $\alpha'(\phi)$  e  $\beta'(\phi)$  sendo diferentes de zero para todo  $\phi$  no espaço paramétrico, onde  $\beta(\phi) = \xi'(\phi)/(\xi(\phi)\alpha'(\phi))$ . Aqui, ' denota derivada com respeito a  $\phi$ . Por exemplo,  $\beta' = \beta'(\phi) = d\beta(\phi)/d\phi$ . Pode ser provado que  $\kappa_{\phi\phi} = -\alpha'\beta'$ ,  $\kappa_{\phi\phi\phi} = -(2\alpha''\beta' + \alpha'\beta'')$ , e  $\kappa_{\phi\phi\phi\phi} = -3\alpha''\beta'' - 3\alpha'''\beta' - \alpha'\beta''''$ . A estatística gradiente tem a forma  $S = n(\phi_0 - \hat{\phi})\alpha'(\phi_0)(\beta(\phi_0) + \bar{d})$ , onde  $\bar{d} = n^{-1} \sum_{i=1}^n d(x_i)$ . De (2.5), (2.6) e (2.7), temos

$$A_1 = \frac{6}{\alpha'\beta'} \left[ 2 \left( \frac{\beta''}{\beta'} \right)^2 + \frac{\alpha''\beta''}{\alpha'\beta'} - \frac{\beta'''}{\beta'} \right],$$

$$A_2 = \frac{3}{\alpha'\beta'} \left\{ \frac{\beta''}{\beta'} \left( \frac{4\alpha''}{\alpha'} - \frac{\beta''}{4\beta'} \right) + 3 \left[ \left( \frac{\alpha''}{\alpha'} \right)^2 + \left( \frac{\beta''}{\beta'} \right)^2 \right] - \left( \frac{\alpha'''}{\alpha'} - \frac{\beta'''}{\beta'} \right) \right\},$$

$$A_3 = \frac{5}{\alpha'\beta'} \left( \frac{\alpha''}{\alpha'} + \frac{\beta''}{2\beta'} \right)^2.$$

Apresentamos agora alguns casos especiais.

1. Normal ( $\phi > 0, \mu \in \mathbb{R}, x \in \mathbb{R}$ ):

- $\mu$  conhecido:  $\alpha(\phi) = 1/(2\phi)$ ,  $\xi(\phi) = \phi^{1/2}$ ,  $d(x) = (x - \mu)^2$  e  $v(x) = -\log(2\pi)/2$ . Temos  $A_1 = 0$ ,  $A_2 = 36$  e  $A_3 = 40$ . O três primeiros momentos de  $S$  até ordem  $n^{-1}$  são  $\mu'_1(S) = 1$ ,  $\mu_2(S) = 2(1 + 6/n)$ , e  $\mu_3(S) = 8(1 + 29/n)$ . A estatística gradiente corrigida é  $S^* = S\{1 - (1 - 11S/3 + 2S^2/3)/(3n)\}$ .
- $\phi$  conhecido:  $\alpha(\mu) = -\mu/\phi$ ,  $\xi(\mu) = \exp(\mu^2/2\phi)$ ,  $d(x) = x$  e  $v(x) = -x^2/2 - \log(2\pi\phi)/2$ . Aqui,  $A_1 = A_2 = A_3 = 0$ , como esperado.

2. Normal inversa ( $\phi > 0, \mu > 0, x > 0$ ):

- $\mu$  conhecido:  $\alpha(\phi) = \phi$ ,  $\xi(\phi) = 1/\phi^{1/2}$ ,  $d(x) = (x - \mu)^2/(2\mu^2x)$ , e  $v(x) = -\log(2\pi x^3)/2$ . Aqui,  $A_1 = 24$ ,  $A_2 = 30$  e  $A_3 = 10$ , e os três primeiros momentos de  $S$  são  $\mu'_1(S) = 1 + 2/n$ ,  $\mu_2(S) = 2 + 18/n$  e  $\mu_3(S) = 8 + 188/n$ . A estatística gradiente corrigida tem a forma  $S^* = S\{1 - (S + 2)(S + 3)/(18n)\}$ .
- $\phi$  conhecido:  $\alpha(\mu) = \phi/(2\mu^2)$ ,  $\xi(\phi) = \exp(-\phi/\mu)$ ,  $d(x) = x$  e  $v(x) = -\phi/(2x^2) + \log(2\pi x^3)/2$ . Temos  $A_1 = 0$  e  $A_2 = A_3 = 45\mu/\phi$ . Os três primeiros momentos de  $S$  são  $\mu'_1(S) = 1$ ,  $\mu_2(S) = 2 + 15\mu/(n\phi)$  e  $\mu_3(S) = 8 + 270\mu/(n\phi)$ . Também,  $S^* = S\{1 - \mu S(S - 5)/(4n\phi)\}$ .

3. Valor extremo truncado ( $\phi > 0, x > 0$ ):  $\alpha(\phi) = 1/\phi$ ,  $\xi(\phi) = \phi$ ,  $d(x) = \exp(x) - 1$  e  $v(x) = x$ . Temos  $A_1 = 0$ ,  $A_2 = 12$ ,  $A_3 = 20$ ,  $\mu'_1(S) = 1$ ,  $\mu_2(S) = 2 + 4/n$ ,  $\mu_3(S) = 8 + 88/n$  e  $S^* = S\{1 - (12 - 15S + 2S^2)/(18n)\}$ .

4. Pareto ( $\phi > 0, k > 0, k$  conhecido,  $x > k$ ):  $\alpha(\phi) = 1 + \phi$ ,  $\xi(\phi) = (\phi k^\phi)^{-1}$ , e  $v(x) = 0$ . Aqui,  $A_1 = 12$ ,  $A_2 = 15$ ,  $A_3 = 5$ ,  $\mu'_1(S) = 1 + 1/n$ ,  $\mu_2(S) = 2 + 9/n$ ,  $\mu_3(S) = 8 + 94/n$  e  $S^* = S\{1 - (S + 2)(S + 3)/(36n)\}$ .

5. Potência ( $\theta > 0, \phi > 0, \theta$  conhecido,  $x > \theta$ ):  $\alpha(\phi) = 1 - \phi, \xi(\phi) = \phi^{-1}\theta^\phi$  e  $v(x) = 0$ . Os  $A$ 's, os três primeiros momentos e a estatística gradiente corrigida coincidem com aqueles obtidos para a distribuição Pareto.
6. Laplace ( $\theta > 0, k \in \mathbb{R}, k$  conhecido,  $x \in \mathbb{R}$ ):  $\alpha(\theta) = \theta^{-1}, \zeta(\theta) = 2\theta, d(x) = |x - k|$  e  $v(x) = 0$ . Temos  $A_1 = 0, A_2 = 18, A_3 = 20, \mu'_1(S) = 1, \mu_2(S) = 2 + 6/n, \mu_3(S) = 8 + 112/n$  e  $S^* = S\{1 - (3 - 11S + 2S^2)/(18n)\}$ .

## 2.4 Modelos com dois parâmetros ortogonais

As famílias biparamétricas de distribuições com parâmetros ortogonais (Cox e Reid, 1987),  $\phi$  e  $\beta$  digamos, são o tema desta seção. A hipótese nula considerada é  $\mathcal{H}_0 : \phi = \phi_0$ , onde  $\phi_0$  é um valor fixado e  $\beta$  atua como parâmetro de perturbação. A ortogonalidade entre  $\phi$  e  $\beta$  leva a uma considerável simplificação nas fórmulas de  $A_1, A_2$  e  $A_3$ . Aqui,  $\kappa_{\phi\phi\beta} = E(\partial^3 \ell(\boldsymbol{\theta})/\partial\beta\partial\phi^2), \kappa_{\phi\phi\beta}^{(\beta)} = \partial\kappa_{\phi\phi\beta}/\partial\beta$ , etc. Após alguma álgebra, temos

$$A_1 = A_{1\phi} + A_{1\phi\beta}, \quad A_2 = A_{2\phi} + A_{2\phi\beta}, \quad A_3 = -\frac{5\kappa_{\phi\phi\phi}^2}{4\kappa_{\phi\phi}^3}, \quad (2.8)$$

onde  $A_{1\phi}$  e  $A_{2\phi}$  são iguais a  $A_1$  e  $A_2$  dados em (2.5) e (2.6), respectivamente, e

$$A_{1\phi\beta} = \frac{3\{4\kappa_{\phi\phi\beta}\kappa_{\phi\phi}^{(\beta)} + \kappa_{\phi\beta\beta}(4\kappa_{\phi\phi}^{(\phi)} - \kappa_{\phi\phi\phi})\}}{\kappa_{\phi\phi}^2\kappa_{\beta\beta}} + \frac{6(\kappa_{\phi\phi\beta\beta} - 2\kappa_{\phi\phi\beta}^{(\beta)} - 2\kappa_{\phi\beta\beta}^{(\phi)})}{\kappa_{\phi\phi}\kappa_{\beta\beta}}$$

$$+ \frac{3\{2\kappa_{\phi\phi\beta}(2\kappa_{\beta\beta}^{(\beta)} - \kappa_{\beta\beta\beta}) + \kappa_{\phi\beta\beta}(2\kappa_{\beta\beta}^{(\phi)} - 3\kappa_{\phi\beta\beta})\}}{\kappa_{\phi\phi}\kappa_{\beta\beta}^2},$$

$$A_{2\phi\beta} = \frac{3(3\kappa_{\phi\phi\phi}\kappa_{\phi\beta\beta} + \kappa_{\phi\phi\beta}^2)}{\kappa_{\phi\phi}^2\kappa_{\beta\beta}}.$$

As expressões para  $A_{1\phi\beta}$  e  $A_{2\phi\beta}$  em (2.8) podem ser consideradas como uma contribuição adicional introduzida na expansão da função distribuição acumulada da estatística de gradiente devido ao fato de que  $\beta$  é desconhecido e tem que ser estimado a partir dos dados. A seguir, apresentamos alguns exemplos.

### Exemplo 2.3 (Distribuição normal)

Sejam  $x_1, \dots, x_n$  uma amostra aleatória de uma distribuição normal  $N(\phi, \beta)$ . A estatística gradiente para testar  $\mathcal{H}_0 : \phi = \phi_0$  pode ser escrita na forma

$$S = n \frac{T_1 T_2^{-1}}{1 + T_1 T_2^{-1}},$$

onde  $T_1 = n(\bar{x} - \phi_0)^2$ ,  $T_2 = \sum_{i=1}^n (x_i - \bar{x})^2$  e  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ . Sob a hipótese nula,  $T_1/\beta$  e  $T_2/\beta$  são independentes com distribuição  $\chi_1^2$  e  $\chi_{n-1}^2$ , respectivamente. Pode ser mostrado que  $n^{-1}S$  tem uma distribuição beta com parâmetros  $1/2$  e  $(n-1)/2$ . Os três primeiros momentos exatos de  $S$  são  $1$ ,  $2(n-1)/(n+2)$  e  $8(n-1)(n-2)/\{(n+2)(n+4)\}$ , respectivamente. Aqui,  $A_1 = A_3 = 0$  e  $A_2 = -18$ . Os três primeiros momentos aproximados  $S$  são  $\mu'(S) = 1$ ,  $\mu_2(S) = 2 - 6/n$  e  $\mu_3(S) = 8 - 72/n$ . Esses momentos diferem dos momentos exatos somente por termos de ordem menor que  $n^{-1}$ . A estatística gradiente corrigida é  $S^* = S\{1 - (3 - S)/(2n)\}$ .

**Exemplo 2.4** (*Distribuição Birnbaum–Saunders biparamétrica*)

A distribuição Birnbaum–Saunders biparamétrica foi proposta por Birnbaum e Saunders (1969) e possui função distribuição acumulada na forma  $G(x) = \Phi(v)$ , com  $x > 0$ , onde  $v = \phi^{-1}\rho(x/\beta)$ ,  $\rho(z) = z^{1/2} - z^{-1/2}$ ,  $\Phi(\cdot)$  é a função distribuição acumulada de uma variável aleatória normal padrão, e  $\phi > 0$  e  $\beta > 0$  são os parâmetros de forma e escala, respectivamente. Desejamos testar  $\mathcal{H}_0 : \phi = \phi_0$  contra a hipótese alternativa  $\mathcal{H}_a : \phi \neq \phi_0$ , onde  $\phi_0$  é uma constante positiva conhecida. A estatística gradiente para testar  $\mathcal{H}_0$  é

$$S = \frac{n(\hat{\phi} - \phi_0)}{\phi_0^3} \{ \bar{s} + \bar{r} - (2 + \phi_0^2) \},$$

onde  $\bar{s} = (n\tilde{\beta})^{-1} \sum_{i=1}^n x_i$ ,  $\bar{r} = \tilde{\beta}n^{-1} \sum_{i=1}^n x_i^{-1}$  e  $\tilde{\beta}$  é o estimador de máxima verossimilhança de  $\beta$  obtido sob  $\mathcal{H}_0$ . Temos  $\kappa_{\phi\phi} = -2/\phi^2$ ,  $\kappa_{\phi\beta} = 0$  e  $\kappa_{\beta\beta} = -\{1 + \phi(2\pi)^{-1/2}h(\phi)\}/(\phi^2\beta^2)$ , onde  $h(\phi) = \phi(\pi/2)^{1/2} - \pi e^{2/\phi^2} \{1 - \Phi(2/\phi)\}$ . Após alguma álgebra, obtemos  $A_{1\phi} = -3$ ,  $A_{2\phi} = 69/8$ ,  $A_{2\phi\beta} = -45(2 + \phi^2)/[2\{1 + \phi(2\pi)^{-1/2}h(\phi)\}]$ ,  $A_3 = 125/8$  e

$$A_{1\phi\beta} = \frac{9 - 15\phi^2/2}{1 + \phi(2\pi)^{-1/2}h(\phi)} + \frac{3(2 + \phi^2)}{\{1 + \phi(2\pi)^{-1/2}h(\phi)\}^2} \left\{ 2(1 + \phi^2) - \frac{(4 + \phi^2)h(\phi)}{\phi\sqrt{2\pi}} \right\}.$$

Obtidas as quantidades necessárias para obter os  $A$ 's, uma estatística gradiente corrigida por um fator de correção tipo-Bartlett pode ser obtida a partir do Corolário 2.2. É interessante notar que os  $A$ 's, não dependem do parâmetro de escala desconhecido  $\beta$ .

Apresentamos uma pequena simulação de Monte Carlo em relação ao teste da hipótese nula  $\mathcal{H}_0 : \phi = 1$ . As simulações foram realizadas definindo  $\beta = 1$  e tamanhos de amostra variando de 5 a 22 observações. Todos os resultados são baseados em 10.000 réplicas. As distorções de tamanho do teste gradiente e sua versão corrigida (ou seja, o tamanho estimado menos o nominal) para o nível nominal 5% são mostrados na Figura 2.1(a). Fica claro a partir desta figura que o teste gradiente corrigido exibe menores distorções de tamanho do que o teste gradiente original.

Em seguida, definimos  $n = 10$  e consideramos a aproximação de primeira ordem (distribuição  $\chi_1^2$ ) para a distribuição da estatística gradiente e a expansão obtida nesse capítulo. A Figura 2.1(b) apresenta as curvas. A partir dessa figura, fica evidente a diferença entre as curvas, e portanto, a distribuição  $\chi_1^2$  pode não ser uma boa aproximação para a distribuição da estatística gradiente sob a hipótese nula  $\mathcal{H}_0 : \phi = 1$  para o modelo Birnbaum–Saunders biparamétrico se a amostra é pequena.

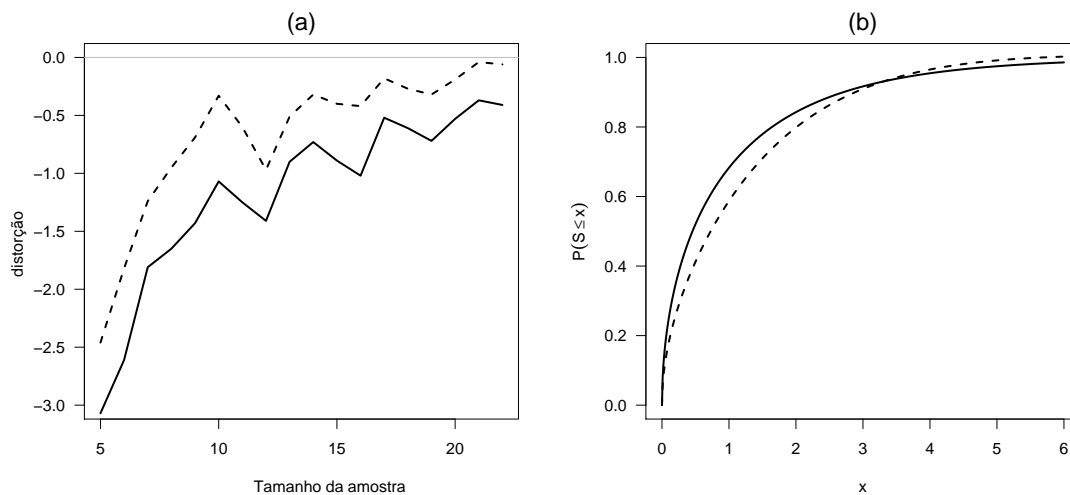


Figura 2.1: (a) Distorção de tamanho do teste gradiente (linha contínua) e do teste gradiente corrigido (linha pontilhada), (b) aproximação de primeira ordem (linha contínua) e expansão até ordem  $n^{-1}$  (linha pontilhada) para a função distribuição acumulada sob a hipótese nula da estatística gradiente; distribuição Birnbaum–Saunders

### Exemplo 2.5 (Distribuição gama)



Sejam  $x_1, \dots, x_n$  uma amostra aleatória de uma distribuição gama com média  $\beta$  e coeficiente de variação  $\phi^{1/2}$ . Aqui, consideramos o problema de testar a hipótese nula  $\mathcal{H}_0 : (\beta, \phi) = (\beta_0, \phi_0)$ , onde  $\beta_0$  e  $\phi_0$  são valores positivos fixados. Note que a hipótese nula é simples, e os  $A$ 's são obtidos do Corolário 2.1 com  $q = 2$ . Após alguma álgebra, obtemos

$$A_1 = 6(d_1 - d_2 + d_3 - 2d_4), \quad A_2 = \frac{18}{\phi} + \frac{3}{4}(d_1 + 2d_2 + 4d_3 - 11d_4), \quad A_3 = \frac{20}{\phi} - \frac{1}{4}(9d_1 - 6d_2 + 5d_4),$$

onde  $d_1 = 1/[\phi(1 - \phi\psi')]$ ,  $d_2 = (1 + \phi^2\psi'')/[\phi(1 - \phi\psi')^2]$ ,  $d_3 = (2 - \phi^3\psi''')/[\phi(1 - \phi\psi')^2]$ , e  $d_4 = (1 + \phi^2\psi'')^2/[\phi(1 - \phi\psi')^3]$ , com  $\psi = \psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ ,  $\Gamma'(\phi) = d\Gamma(\phi)/d\phi$ ,  $\psi' = d\psi/d\phi$ ,  $\psi'' = d^2\psi/d\phi^2$  e  $\psi''' = d^3\psi/d\phi^3$  e  $\Gamma(\cdot)$  representa a função gama.

Apresentamos agora um estudo de simulação baseado em 10.000 réplicas. A hipótese nula é  $\mathcal{H}_0 : \beta = \phi = 1$ . Note que  $\mathcal{H}_0$  significa que os dados vêm de uma distribuição exponencial com média unitária. A Figura 2.2(a) apresenta as distorções de tamanho do teste gradiente e sua versão corrigida para o nível nominal 10% em diferentes tamanhos de amostra. Pode-se notar que o teste gradiente apresenta grandes distorções de tamanho e sua versão corrigida mostra uma distorção menor.

Finalmente, consideramos  $n = 10$  e traçamos as aproximações de primeira e segunda ordem para a distribuição da estatística gradiente. Uma inspeção visual da Figura 2.2(b) revela que a aproximação  $\chi^2_2$  de primeira ordem pode não ser bem acurada em pequenas amostras.

## 2.5 Discussão

Lemonte e Ferrari (2012b) mostraram que o teste de gradiente pode ser uma alternativa interessante aos testes clássicos em grandes amostras, ou seja, os testes da razão de verossimilhanças, de Wald e escore, uma vez que nenhum é uniformemente superior aos outros em termos de poder local de segunda ordem. Além disso, como já observado, a estatística gradiente não envolve a obtenção, estimação, ou inversão da matriz de informação, ao contrário das estatísticas escore e de Wald. Sua forma é simples e atraente. A distribuição da estatística gradiente sob a hipótese nula geralmente é desconhecida, e portanto, o teste se baseia em uma aproximação assintótica. A distribuição qui-quadrado é usada como uma aproximação em grandes amostras para a verdadeira distribuição da estatística sob  $\mathcal{H}_0$ .

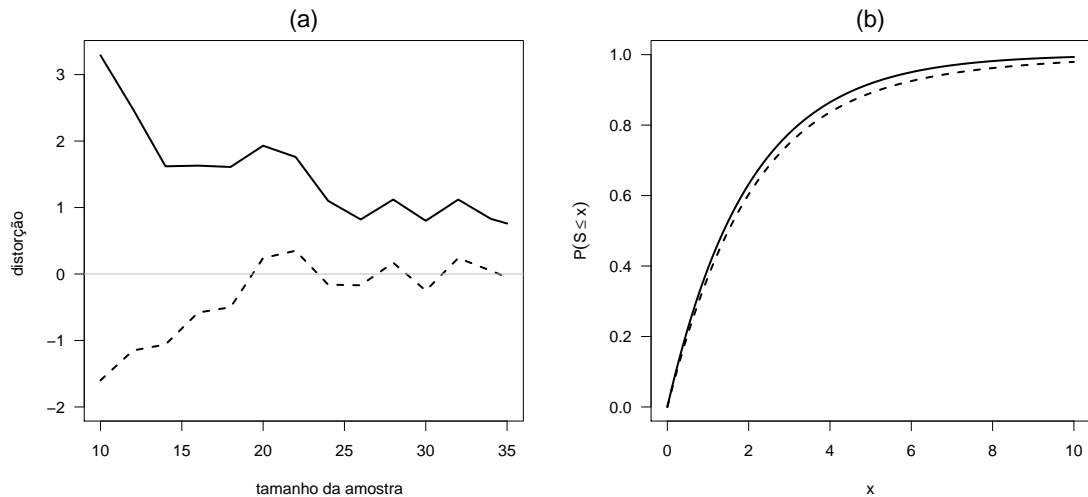


Figura 2.2: (a) Distorção de tamanho para o teste gradiente (linha contínua) e para sua versão corrigida (linha pontilhada), (b) Aproximação de primeira ordem (linha contínua) e expansão até ordem  $n^{-1}$  (linha pontilhada) da função distribuição acumulada da estatística gradiente sob a hipótese nula; distribuição gama.

No entanto, em pequenas amostras, a distribuição qui-quadrado pode ser uma aproximação ruim para essa distribuição, isto é, a aproximação assintótica pode levar a uma inferência não acurada. A fim de diminuir esta imprecisão, uma alternativa estratégica é a utilização de uma teoria assintótica de alta ordem. A expansão assintótica até ordem  $n^{-1}$  para a função distribuição acumulada da estatística gradiente sob a hipótese nula foi derivada nesse capítulo. Uma rota Bayesiana baseada no argumento de encolhimento (Ghosh e Mukerjee, 1991; Mukerjee e Reid, 2000); ver também Seção 1.2; mostrou-se extremamente útil nesse contexto. A expansão é muito geral no sentido que a hipótese nula pode ser composta, na presença de parâmetros de perturbação. Mostramos que os coeficientes que definem a expansão são funções de momentos de derivadas do logaritmo da função de verossimilhança total. Infelizmente, esses coeficientes são difíceis de se interpretar em generalidade. Eles podem ser usados para investigar o quanto a distribuição assintótica se aproxima da verdadeira distribuição (desconhecida) da estatística gradiente.

Cordeiro e Ferrari (1991) mostraram que, de maneira geral, as estatísticas que têm distribuição assintótica qui-quadrado podem ser modificadas por um fator de correção adequado que faz com que a estatística modificada tenha distribuição qui-quadrado até ordem  $n^{-1}$ . Este trabalho que pode ser visto como uma extensão das correções de Bartlett para a estatística da razão de verossimilhanças (Lawley, 1956) para outras estatísticas que têm distribuição assintótica qui-quadrado. O fator de correção vem de coeficientes do termo de ordem  $O(n^{-1})$  na expansão da função distribuição acumulada da estatística de teste de modo que esta se torna melhor aproximada pela distribuição qui-quadrado de referência. Isto é conhecido como correção tipo-Bartlett. As correções de Bartlett e tipo-Bartlett tornaram-se métodos amplamente utilizados para melhorar a aproximação pela distribuição qui-quadrado em grandes amostras para a distribuição sob a hipótese nula das estatísticas da razão de verossimilhanças e score, respectivamente. Nos últimos anos tem sido renovado o interesse por fatores de Bartlett e vários artigos foram publicados, onde foram dadas expressões para o cálculo dessas correções em modelos especiais. Algumas referências são Zucker et al. (2000), Lagos e Morettin (2004), Tu et al. (2005), van Giersbergen (2009), Bai (2009), Lagos et al. (2010) e Noma (2011).

A partir da expansão geral derivada nesse capítulo e utilizando resultados de Cordeiro e Ferrari (1991), obtivemos um fator de correção tipo-Bartlett para a estatística gradiente. A vantagem da estatística corrigida em relação à sua versão não corrigida é que a primeira é melhor aproximada pela distribuição qui-quadrado. Infelizmente, a estatística gradiente corrigida é mais difícil de se obter. Nossos resultados são muito gerais, ou seja, não estão vinculados a classes especiais de modelos, além de permitir que o vetor de parâmetros seja multidimensional e permitir a presença ou não de parâmetros de perturbação. Além disso, assim como os coeficientes da expansão, e consequentemente o fator de correção tipo-Bartlett, podem ser escritos como funções de momentos de derivadas do logaritmo da função de verossimilhança total, estes podem ser obtidos para todas as classes de modelos paramétricos para os quais esses momentos podem ser determinados. Portanto, as aplicações de nossos resultados gerais em vários modelos paramétricos podem ser estudados. No Capítulo 3, aplicamos esses resultados na classe dos modelos lineares generalizados com dispersão conhecida e desconhecida.

## 2.6 Detalhes técnicos

### 2.6.1 Demonstração do teorema 2.1

Exceto quando indicado, os índices  $j, r, s, u, v,$  e  $w$  variam de 1 a  $p$  e os índices  $j', r', s', u', v',$  e  $w'$  variam de 1 a  $q$ . Além disso, um índice repetido tanto com sobrescrito quanto subscrito indica uma soma implícita na variação apropriada. Sejam  $\lambda_{jr} = -\psi_{jr} = -\{D_j D_r \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ ,  $\psi_{jrs} = \{D_j D_r D_s \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ ,  $\psi_{jrsu} = \{D_j D_r D_s D_u \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ , etc. A matriz  $\boldsymbol{\Lambda} = ((\lambda_{jr}))$  é a matriz de informação observada avaliada em  $\hat{\boldsymbol{\theta}}$ . A partição de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$  induz as partições

$$\boldsymbol{\Lambda} = ((\lambda_{jr})) = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{bmatrix}, \quad \boldsymbol{\Lambda}^{-1} = ((\lambda^{jr})) = \begin{bmatrix} \boldsymbol{\Lambda}^{11} & \boldsymbol{\Lambda}^{12} \\ \boldsymbol{\Lambda}^{21} & \boldsymbol{\Lambda}^{22} \end{bmatrix},$$

em que  $\boldsymbol{\Lambda}^{-1}$  é a inversa de  $\boldsymbol{\Lambda}$ . Seja  $\boldsymbol{\Lambda}^{11^{-1}} = ((\lambda_{1w'j'}))$ ,  $\sigma^{jr} = \lambda^{jr} - \lambda^{jw'} \lambda_{1w'j'} \lambda^{j'r}$ ,  $\tau^{jj'} = \lambda^{jw'} \lambda_{1w'j'}$ ,  $\sigma_{suw}^{(1)} = \sigma^{su} \sigma^{vw} [3]$ ,  $\lambda_{j'r's'u'}^{(1)} = \lambda^{j'r'} \lambda^{s'u'} [3]$ , e  $\lambda_{j'r's'u'v'w'}^{(2)} = \lambda^{j'r'} \lambda^{s'u'} \lambda^{v'w'} [15]$ , em que  $[\cdot]$  denota um somatório com o número entre colchetes indicando o número de termos obtidos pela permutação dos índices. Por exemplo,  $\sigma^{su} \sigma^{vw} [3] = \sigma^{su} \sigma^{vw} + \sigma^{sv} \sigma^{uw} + \sigma^{sw} \sigma^{uv}$ . Seja  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_q)^\top = n^{1/2}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)$ ,  $\Psi_{j'}^{(1)} = \psi_{jrs} \sigma^{rs} \tau^{jj'} / 2$ ,  $\Psi_{j'r's'}^{(3)} = \psi_{jrs} \tau^{jj'} \tau^{rr'} \tau^{ss'} / 6$ ,

$$\Psi_{j'r's'u'}^{(4)} = \frac{1}{24} \{ \psi_{jrsu} + \sigma^{vw} (2\psi_{jrs} \psi_{uvw} + 3\psi_{jrv} \psi_{suw}) \} \tau^{jj'} \tau^{rr'} \tau^{ss'} \tau^{uu'}.$$

**Lema 2.1** *Uma expansão assintótica sob a hipótese nula para a estatística gradiente (2.1) é*

$$S = \boldsymbol{\epsilon}^\top \boldsymbol{\Lambda}^{11^{-1}} \boldsymbol{\epsilon} - \frac{3}{\sqrt{n}} \Psi_{j'r's'}^{(3)} \epsilon_{j'} \epsilon_{r'} \epsilon_{s'} - \frac{4}{n} \left( \Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'}^{(3)} \Psi_{u'}^{(1)} \right) \epsilon_{j'} \epsilon_{r'} \epsilon_{s'} \epsilon_{u'} + o_p(n^{-1}). \quad (2.9)$$

*Prova.* Seguimos um procedimento análogo ao de Chang e Mukerjee (2011). Seja  $\boldsymbol{e} = (\boldsymbol{\theta}_1^\top, \widetilde{\boldsymbol{\theta}}_2^\top)^\top = (\epsilon_1, \dots, \epsilon_p)^\top$ , i.e.  $e_{j'} = \theta_{j'}$  para  $j' = 1, \dots, q$  e  $e_j = \widetilde{\theta}_j$  para  $j = q+1, \dots, p$ . A estatística gradiente definida em (2.1) para testar a hipótese nula  $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ , pode ser escrita como

$$S = -n^{1/2} \sum_{v'=1}^q D_{v'} \ell(\boldsymbol{e}) \epsilon_{v'}. \quad (2.10)$$

Primeiramente, note que

$$D_j \ell(\boldsymbol{e}) = 0, \quad j = q+1, \dots, p. \quad (2.11)$$

Por diferenciação de (2.11) com respeito a  $\theta_{j'}$  ( $j' = 1, \dots, q$ ) e avaliando em  $\hat{\boldsymbol{\theta}}$ , temos

$$\lambda_{js} \hat{e}_{sj'} = 0, \quad (2.12)$$

em que  $\hat{e}_{sj'} = [\partial e_s / \partial \theta_{j'}]_{\theta = \hat{\theta}}$ . Agora, diferenciando (2.12) com respeito a  $\theta_{r'}$  and  $\theta_{s'}$  e avaliando em  $\hat{\theta}$ , temos

$$\psi_{jsu} \hat{e}_{ur'} \hat{e}_{sj'} - \lambda_{js} \hat{e}_{sr'j'} = 0, \quad (2.13)$$

$$\psi_{jsuv} \hat{e}_{vs'} \hat{e}_{ur'} \hat{e}_{sj'} + \psi_{jsu} (\hat{e}_{us'r'} \hat{e}_{sj'} + \hat{e}_{us'j'} \hat{e}_{ur'}) + \psi_{jsv} \hat{e}_{sr'j'} \hat{e}_{vs'} - \lambda_{js} \hat{e}_{sr'j's'} = 0, \quad (2.14)$$

em que  $\hat{e}_{sr'j'} = [\partial^2 e_s / \partial \theta_{r'} \partial \theta_{j'}]_{\theta = \hat{\theta}}$  and  $\hat{e}_{sr'j's'} = [\partial^3 e_s / \partial \theta_{r'} \partial \theta_{j'} \partial \theta_{s'}]_{\theta = \hat{\theta}}$ .

Note que  $\sigma^{wj} \lambda_{js} = \delta_{ws} - \tau^{wj'} \delta_{j's}$  e  $\sigma^{wj} \lambda_{jr'} = \delta_{ws} - \tau^{wr'}$ , em que  $\delta$  denota o *delta de Kronecker*<sup>1</sup>. Adicionalmente, note que  $\partial e_{j'} / \partial \theta_{j'} = 1$ , para  $j' = 1, \dots, q$ . Assim, multiplicando ambos os lados de (2.12), (2.13) e (2.14) por  $\sigma^{wj}$ , temos

$$\hat{e}_{wj'} = \tau^{wj'}, \quad \hat{e}_{wr's'} = \sigma^{wj} \psi_{jrs} \tau^{rr'} \tau^{ss'},$$

$$\begin{aligned} \hat{e}_{wr'j's'} = & \sigma^{wj} [\psi_{jsuv} \tau^{vs'} \tau^{ur'} \tau^{sj'} + \psi_{jsu} \psi_{j\tilde{s}\tilde{u}} (\sigma^{\tilde{u}j} \tau^{\tilde{u}s'} \tau^{\tilde{s}r'} \tau^{sj'} + \sigma^{\tilde{s}j} \tau^{\tilde{s}s'} \tau^{\tilde{u}j'} \tau^{ur'}) \\ & + \psi_{jsv} \psi_{j\tilde{s}\tilde{u}} \sigma^{\tilde{s}j} \tau^{vs'} \tau^{\tilde{u}r'} \tau^{sj'}], \end{aligned} \quad (2.15)$$

para  $1 \leq \tilde{j}, \tilde{s}, \tilde{u}, \tilde{v} \leq p$ .

Seja  $\hat{e}$  representando  $e$  avaliado em  $\hat{\theta}$ . Para cada  $v'$  e cada  $1 \leq x, y, z \leq p$ ,

$$\begin{aligned} D_{v'} \ell(e) = & \frac{1}{\sqrt{n}} D_x D_{v'} \ell(\hat{e}) \hat{e}_{xj'} \epsilon_{j'} + \frac{1}{2n} \left[ D_y D_x D_{v'} \ell(\hat{e}) \hat{e}_{yr'} \hat{e}_{xj'} + D_x D_{v'} \ell(\hat{e}) \hat{e}_{xr'j'} \right] \epsilon_{j'} \epsilon_{r'} \\ & + \frac{1}{6n\sqrt{n}} \left[ D_z D_y D_x D_{v'} \ell(\hat{e}) \hat{e}_{zs'} \hat{e}_{yr'} \hat{e}_{xj'} + D_y D_x D_{v'} \ell(\hat{e}) \hat{e}_{yr'} \hat{e}_{xs'j'} \right. \\ & \left. + D_z D_x D_{v'} \ell(\hat{e}) \hat{e}_{zs'} \hat{e}_{xj'r'} + D_y D_x D_{v'} \ell(\hat{e}) \hat{e}_{xj'} \hat{e}_{ys'r'} + D_x D_{v'} \ell(\hat{e}) \hat{e}_{xj'r's'} \right] \epsilon_{s'} \epsilon_{r'} \epsilon_{j'} \\ & + o_p(n^{-3/2}), \end{aligned}$$

que resulta em

$$\begin{aligned} D_{v'} \ell(e) = & -\frac{1}{\sqrt{n}} \lambda_{xv'} \hat{e}_{xj'} \epsilon_{j'} + \frac{1}{2n} \left[ \psi_{yxv'} \hat{e}_{yr'} \hat{e}_{xj'} - \lambda_{xv'} \hat{e}_{xr'j'} \right] \epsilon_{j'} \epsilon_{r'} \\ & + \frac{1}{6n\sqrt{n}} \left[ \psi_{zyxv'} \hat{e}_{zs'} \hat{e}_{yr'} \hat{e}_{xj'} + \psi_{yxv'} \hat{e}_{yr'} \hat{e}_{xs'j'} \right. \\ & \left. + \psi_{zxv'} \hat{e}_{zs'} \hat{e}_{xj'r'} + \psi_{yxv'} \hat{e}_{xj'} \hat{e}_{ys'r'} - \lambda_{xv'} \hat{e}_{xj'r's'} \right] \epsilon_{s'} \epsilon_{r'} \epsilon_{j'} + o_p(n^{-3/2}). \end{aligned} \quad (2.16)$$

Observe que

$$\lambda_{xv'} \hat{e}_{xj'} = \lambda_{xv'} \lambda^{xw'} \lambda_{1w'j'} = \lambda_{v'x} \lambda_{1w'j'} = \lambda_{1v'j'} = \lambda_{1r'j'}.$$

<sup>1</sup>O delta de Kronecker é definido como  $\delta_{wj} = 1$ , se  $w = j$  e  $\delta_{wj} = 0$ , se  $w \neq j$ .

Na última igualdade rearranjamos alguns índices sem perda de generalidade. Agora, note que

$$\begin{aligned}
\psi_{yxv'}\hat{e}_{yr'}\hat{e}_{xj'} - c_{xv'}\hat{e}_{xr'j'} &= \psi_{yxv'}\tau^{yr'}\tau^{xj'} - \lambda_{xv'}(\lambda^{xj} - \lambda^{xw'}\lambda_{1w'j'}\lambda^{j'j})\psi_{j'su}\tau^{ur'}\tau^{sj'} \\
&= \lambda_{1v'j'}\lambda^{j'j}\psi_{j'su}\tau^{ur'}\tau^{sj'} \\
&= \psi_{j'su}\tau^{ur'}\tau^{sj'}\tau^{jv'} \\
&= \psi_{j'rs}\tau^{jj'}\tau^{rr'}\tau^{ss'} \\
&= 6\Psi_{j'r's'}^{(3)},
\end{aligned}$$

e que

$$\begin{aligned}
&\psi_{zyxv'}\hat{e}_{zs'}\hat{e}_{yr'}\hat{e}_{xj'} + \psi_{yxv'}\hat{e}_{yr'}\hat{e}_{xs'j'} + \psi_{zxv'}\hat{e}_{zs'}\hat{e}_{xj'r'} + \psi_{yxv'}\hat{e}_{xj'}\hat{e}_{ys'r'} - \lambda_{xv'}\hat{e}_{xj'r's'} \\
&= \psi_{zyxv'}\tau^{zs'}\tau^{yr'}\tau^{xj'} + \psi_{yxv'}\psi_{j'rs}\sigma^{xj}\tau^{yr'}\tau^{rj'}\tau^{ss'} \\
&\quad + \psi_{zxv'}\psi_{j'rs}\sigma^{xj}\tau^{zs'}\tau^{rj'}\tau^{sr'} + \psi_{yxv'}\psi_{j'rs}\sigma^{yj}\tau^{xj'}\tau^{rs'}\tau^{sr'} \\
&\quad - \psi_{v'rsu}\tau^{rr'}\tau^{ss'}\tau^{uu'} - (\delta_{jv'} - \tau^{jv'})[\psi_{j'suv}\tau^{vs'}\tau^{ur'}\tau^{sj'} \\
&\quad + \psi_{j'su}\psi_{\tilde{j}\tilde{s}\tilde{u}}(\sigma^{uj}\tau^{\tilde{u}s'}\tau^{\tilde{s}r'}\tau^{sj'} + \sigma^{\tilde{s}\tilde{j}}\tau^{\tilde{s}s'}\tau^{\tilde{u}j'}\tau^{ur'}) + \psi_{j'sv}\psi_{\tilde{j}\tilde{s}\tilde{u}}\sigma^{\tilde{s}\tilde{j}}\tau^{vs'}\tau^{\tilde{u}r'}\tau^{sj'}] \\
&= 24(\Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'u'}^{(3)}\Psi_{u'}^{(1)}).
\end{aligned}$$

Novamente fizemos um troca de índices, sem perda de generalidade. Substituindo esses termos em (2.16) e posteriormente (2.16) em (2.10), obtemos (2.9) após alguma álgebra.  $\square$

Sejam  $\pi = \pi(\boldsymbol{\theta})$  uma densidade a priori para  $\boldsymbol{\theta}$ ,  $\pi_j = D_j\pi(\boldsymbol{\theta})$ ,  $\pi_{jr} = D_jD_r\pi(\boldsymbol{\theta})$ ,  $\hat{\pi} = \pi(\hat{\boldsymbol{\theta}})$ ,  $\hat{\pi}_j = \pi_j(\hat{\boldsymbol{\theta}})$ ,  $\hat{\pi}_{jr} = \pi_{jr}(\hat{\boldsymbol{\theta}})$ ,

$$\begin{aligned}
\Psi_{j'r'}^{(2)} &= \left\{ \frac{\hat{\pi}_{jr}}{2\hat{\pi}} + \frac{1}{4}\psi_{j'rsu}\sigma^{su} + \frac{1}{24}(2\psi_{j'rs}\psi_{uvw} + 3\psi_{j'su}\psi_{rvw})\sigma_{suvw}^{(1)} \right\} \tau^{jj'}\tau^{rr'}, \\
\Gamma_{j'}^{(1)} &= \Psi_{j'}^{(1)} + \frac{\hat{\pi}_j}{\hat{\pi}}\tau^{jj'}, \quad \Gamma_{j'r'}^{(2)} = \Psi_{j'r'}^{(2)} + \frac{1}{2\hat{\pi}}(\psi_{j'rs}\hat{\pi}_u + \psi_{j'su}\hat{\pi}_r)\sigma^{su}\tau^{jj'}\tau^{rr'}, \\
\Gamma_{j'r's'u'}^{(4)} &= \Psi_{j'r's'u'}^{(4)} + \frac{\hat{\pi}_u}{6\hat{\pi}}\psi_{j'rs}\tau^{jj'}\tau^{rr'}\tau^{ss'}\tau^{uu'}.
\end{aligned}$$

De Ghosh e Mukerjee (1991), Chang e Mukerjee (2010) derivaram uma expansão até ordem  $n^{-1}$  para a densidade marginal a posteriori de  $\boldsymbol{\epsilon}$ , a qual tem a forma

$$\begin{aligned}
\pi_{post}(\boldsymbol{\epsilon}) &= \phi_q(\boldsymbol{\epsilon}; \boldsymbol{\Lambda}^{11}) \left[ 1 + \frac{1}{\sqrt{n}}(\Gamma_{j'}^{(1)}\epsilon_{j'} + \Gamma_{j'r's'}^{(3)}\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}) \right. \\
&\quad + \frac{1}{n} \left\{ \Gamma_{j'r'}^{(2)}(\epsilon_{j'}\epsilon_{r'} - \lambda^{j'r'}) + \Gamma_{j'r's'u'}^{(4)}(\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'} - \lambda_{j'r's'u'}^{(1)}) \right. \\
&\quad \left. \left. + \frac{1}{2}\Psi_{j'r's'}^{(3)}\Psi_{u'v'w'}^{(3)}(\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\epsilon_{v'}\epsilon_{w'} - \lambda_{j'r's'u'v'w'}^{(2)}) \right\} \right] + o(n^{-1}),
\end{aligned} \tag{2.17}$$

em que  $\phi_q(z; \Sigma)$  denota a densidade de uma distribuição normal  $q$ -variada com média  $\mathbf{0}$  e matriz de covariância  $\Sigma$ .

Agora, seguimos a rota Bayesiana descrita em Mukerjee e Reid (2000); ver Seção 1.2.

**Passo 1.** A função característica posteriori aproximada de  $S$  é

$$M_\pi(t) = E_\pi\{\exp(\xi S)\} = \int \exp(\xi S)\pi_{post}(\epsilon)d\epsilon,$$

em que  $\xi = it$  com  $i = (-1)^{1/2}$ . Do Lema 2.1 e após alguma álgebra, escrevemos

$$\begin{aligned} \exp(\xi S)\pi_{post}(\epsilon) &= (1 - 2\xi)^{-q/2}\phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) \left[1 + \frac{1}{\sqrt{n}}\left\{(1 - 3\xi)\Psi_{j'r's'}^{(3)}\epsilon_{j'}\epsilon_{r'}\epsilon_{s'} + \Gamma_{j'}^{(1)}\epsilon_{j'}\right\}\right. \\ &\quad + \frac{1}{n}\left[\frac{1}{2}\Psi_{j'r's'}^{(3)}\Psi_{u'v'w'}^{(3)}\left\{\frac{1}{9}(1 - 3\xi)^2\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\epsilon_{v'}\epsilon_{w'} - \lambda_{j'r's'u'v'w'}^{(2)}\right\}\right. \\ &\quad - \left.\left[\xi\left\{4\Psi_{j'r's'u'}^{(4)} + \Psi_{j'r's'}^{(3)}(3\Gamma_{u'}^{(1)} - 4\Psi_{u'}^{(1)})\right\} - \Gamma_{j'r's'u'}^{(4)}\right]\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\right. \\ &\quad \left.\left. + \Gamma_{j'r'}^{(2)}(\epsilon_{j'}\epsilon_{r'} - \lambda^{j'r'}) - \Gamma_{j'r's'u'}^{(4)}\lambda_{j'r's'u'}^{(1)}\right]\right] + o_p(n^{-1}). \end{aligned}$$

Usando as seguintes integrais dadas em Datta e Mukerjee (2003, Capítulo. 4):

$$\begin{aligned} \int \phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) d\epsilon &= 1, \quad \int \epsilon_{j'}\phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) d\epsilon = 0, \\ \int \epsilon_{j'}\epsilon_{r'}\phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) d\epsilon &= (1 - 2\xi)^{-1}\lambda^{j'r'}, \quad \int \epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) d\epsilon = 0, \\ \int \epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\phi_q\left(\epsilon; \frac{\Lambda^{11}}{1 - 2\xi}\right) d\epsilon &= (1 - 2\xi)^{-2}\lambda_{j'r's'u'}^{(1)}, \\ \int \epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\epsilon_{v'}\epsilon_{w'}\phi_q\left(h_{(1)}; \frac{1}{1 - 2\xi}C^{11}\right) d\epsilon &= (1 - 2\xi)^{-3}\lambda_{j'r's'u'v'w'}^{(2)}, \end{aligned}$$

fazendo  $\xi = -\frac{1}{2}(1 - 2\xi) + \frac{1}{2}$ ,  $\xi^2 = \frac{1}{4}(1 - 2\xi)^2 - \frac{1}{2}(1 - 2\xi) + \frac{1}{4}$ , e assumindo que  $\theta$  está no interior do suporte de  $\pi$ , obtemos após alguma álgebra

$$M_\pi(t) = (1 - 2\xi)^{-q/2} \left\{1 + \frac{1}{n} \sum_{i=0}^3 H_i (1 - 2\xi)^{-i}\right\} + o_p(n^{-1}), \quad (2.18)$$

em que  $H_0 = -(H_1 + H_2 + H_3)$ ,

$$\begin{aligned} H_1 &= \frac{9}{8}\Psi_{j'r's'}^{(3)}\Psi_{u'v'w'}^{(3)}\lambda_{j'r's'u'v'w'}^{(2)} + \Gamma_{j'r'}^{(2)}\lambda^{j'r'} \\ &\quad + \lambda_{j'r's'u'}^{(1)} \left\{2(\Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'}^{(3)}\Psi_{u'}^{(1)}) + \frac{3}{2}\Psi_{j'r's'}^{(3)}\Gamma_{u'}^{(1)}\right\}, \end{aligned}$$

$$\begin{aligned}
H_2 &= -\frac{3}{4}\Psi_{j'r's'}^{(3)}\Psi_{u'v'w'}^{(3)}\lambda_{j'r's'u'v'w'}^{(2)} \\
&\quad + \lambda_{j'r's'u'}^{(1)} \left\{ \Gamma_{j'r's'u'}^{(4)} - 2(\Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'}^{(3)}\Psi_{u'}^{(1)}) - \frac{3}{2}\Psi_{j'r's'}^{(3)}\Gamma_{u'}^{(1)} \right\}, \\
H_3 &= \frac{1}{8}\Psi_{j'r's'}^{(3)}\Psi_{u'v'w'}^{(3)}\lambda_{j'r's'u'v'w'}^{(2)}.
\end{aligned}$$

**Passo 2.** Seja  $\bar{\pi}(\cdot)$  uma densidade a priori auxiliar para  $\theta$  satisfazendo as condições dadas em Bickel e Ghosh (1990). Agora, obtemos uma função característica posteriori aproximada de  $S$  sob a priori  $\bar{\pi}(\cdot)$ ,  $M_{\bar{\pi}}(t)$  digamos.

De (2.18), temos

$$M_{\bar{\pi}}(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^3 \bar{H}_i (1 - 2\xi)^{-i} \right\} + o_p(n^{-1}),$$

em que  $\bar{H}_i$  denota a contraparte  $H_i$  obtida pela substituição de  $\pi(\cdot)$  por  $\bar{\pi}(\cdot)$ . Após alguma álgebra, temos

$$\Delta(\theta) = E_{\theta}(M_{\bar{\pi}}) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^3 \bar{J}_i (1 - 2\xi)^{-i} \right\} + o(n^{-1}),$$

em que  $\bar{J}_0 = -(\bar{J}_1 + \bar{J}_2 + \bar{J}_3)$ ,

$$\begin{aligned}
\bar{J}_1 &= \frac{1}{32}\kappa_{jrs}\kappa_{uvw}(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}) + \frac{1}{4}\kappa_{jrsm}m^{jr}m^{su} \\
&\quad + \frac{1}{4}\kappa_{jrv}\kappa_{suw}a^{vw}(m^{jr}m^{su}[3]) + \frac{1}{8}\kappa_{jrs}\kappa_{uvw}a^{vw}(m^{jr}m^{su}[3]) + \frac{3}{4}\kappa_{jrs}m^{jr}m^{su}\frac{\bar{\pi}_u}{\bar{\pi}} \\
&\quad + \frac{1}{2}m^{jr} \left\{ \frac{\bar{\pi}_{jr}}{\bar{\pi}} + \frac{1}{2}\kappa_{jrsm}a^{su} + \frac{1}{12}(2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jsu}\kappa_{rvw})(a^{su}a^{vw}[3]) \right. \\
&\quad \left. + \frac{\bar{\pi}_u}{\bar{\pi}}\kappa_{jrs}a^{su} + \frac{\bar{\pi}_r}{\bar{\pi}}\kappa_{jsu}a^{su} \right\},
\end{aligned}$$

$$\begin{aligned}
\bar{J}_2 &= -\frac{1}{48}\kappa_{jrs}\kappa_{uvw}(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}) - \frac{1}{4}\kappa_{jrsm}m^{jr}m^{su} \\
&\quad - \frac{1}{4}\kappa_{jrv}\kappa_{suw}a^{vw}(m^{jr}m^{su}[3]) - \frac{1}{8}\kappa_{jrs}\kappa_{uvw}a^{vw}(m^{jr}m^{su}[3]) - \frac{3}{4}\kappa_{jrs}m^{jr}m^{su}\frac{\bar{\pi}_u}{\bar{\pi}} \\
&\quad + \frac{1}{8} \left( \kappa_{jrsm} + \frac{1}{6}\kappa_{jrs}\frac{\bar{\pi}_u}{\bar{\pi}} \right) m^{jr}m^{su} + \frac{1}{24}(2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jrv}\kappa_{suw})a^{vw}(m^{jr}m^{su}[3]),
\end{aligned}$$

$$\bar{J}_3 = \frac{1}{288}\kappa_{jrs}\kappa_{uvw}(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}).$$

**Passo 3.** Agora, determinamos

$$\int \Delta(\theta)\bar{\pi}(\theta)d\theta = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^3 (1 - 2\xi)^{-i} \int \bar{J}_i\bar{\pi}(\theta)d\theta \right\} + o(n^{-1}),$$



pela integração dos  $\bar{J}$ 's em relação a  $\bar{\pi}$ . O primeiro termo em  $\bar{J}_1$  que depende densidade a priori é  $(3/4)\kappa_{jrs}m^{jr}m^{su}\bar{\pi}_u/\bar{\pi}$ .

Por integração desse termo, obtemos

$$\begin{aligned} \int \frac{3}{4}\kappa_{jrs}m^{jr}m^{su}\frac{\bar{\pi}_u}{\bar{\pi}}\bar{\pi}d\boldsymbol{\theta} &= \frac{3}{4}\int \kappa_{jrs}m^{jr}m^{su}\bar{\pi}_ud\boldsymbol{\theta} \\ &= \frac{3}{4}\left\{(\kappa_{jrs}m^{jr}m^{su}\bar{\pi})|_{\partial\text{supp}\bar{\pi}} - \int D_u(\kappa_{jrs}m^{jr}m^{su})\bar{\pi}d\boldsymbol{\theta}\right\}, \end{aligned}$$

em que  $\partial\text{supp}\bar{\pi}$  denota a fronteira do suporte de  $\bar{\pi}$ . Como  $\bar{\pi}$  é zero em  $\partial\text{supp}\bar{\pi}$ , temos

$$\frac{3}{4}\int \kappa_{jrs}m^{jr}m^{su}\bar{\pi}_ud\boldsymbol{\theta} = -\frac{3}{4}\int D_u(\kappa_{jrs}m^{jr}m^{su})\bar{\pi}d\boldsymbol{\theta}.$$

Analogamente,

$$\int \frac{1}{2}m^{jr}\frac{\bar{\pi}_{jr}}{\bar{\pi}}\bar{\pi}d\boldsymbol{\theta} = \frac{1}{2}\int D_jD_r(m^{jr})\bar{\pi}d\boldsymbol{\theta},$$

$$\int \frac{1}{2}\kappa_{jrs}m^{jr}a^{su}\frac{\bar{\pi}_u}{\bar{\pi}}\bar{\pi}d\boldsymbol{\theta} = -\frac{1}{2}\int D_u(\kappa_{jrs}m^{jr}a^{su})\bar{\pi}d\boldsymbol{\theta},$$

e

$$\int \frac{1}{2}\kappa_{jsu}m^{jr}a^{su}\frac{\bar{\pi}_r}{\bar{\pi}}\bar{\pi}d\boldsymbol{\theta} = -\frac{1}{2}\int D_r(\kappa_{jsu}m^{jr}a^{su})\bar{\pi}d\boldsymbol{\theta}.$$

Após a integração de cada termo dependente da distribuição a priori e fazendo  $\bar{\pi}(\cdot)$  convergir fracamente para priori gerada no verdadeiro valor de  $\boldsymbol{\theta}$ , chegamos a

$$E_{\boldsymbol{\theta}}\{\exp(\xi S)\} = (1 - 2\xi)^{-q/2} \left\{ 1 + n^{-1} \sum_{i=0}^3 \bar{A}_i (1 - 2\xi)^{-i} \right\} + o(n^{-1}),$$

em que os  $\bar{A}$ 's são funções de derivadas de momentos do logaritmo da função de verossimilhança total. Escrevendo  $d = 2\xi/(1 - 2\xi)$  e usando o fato de que  $\sum_{i=0}^3 \bar{A}_i = 0$ , chegamos a

$$M(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{24n} (A_1 d + A_2 d^2 + A_3 d^3) \right\} + o(n^{-1}), \quad (2.19)$$

com  $A_1 = 24(\bar{A}_1 + 2\bar{A}_2 + 3\bar{A}_3)$ ,  $A_2 = 24(\bar{A}_2 + 3\bar{A}_3)$ , e  $A_3 = 24\bar{A}_3$ . Podemos escrever

$$\begin{aligned} A_1 &= 12D_jD_r m^{jr} - 6D_u(\kappa_{jrs}m^{jr}m^{su}) - 12D_u(\kappa_{jrs}m^{jr}a^{su}) - 12D_r(\kappa_{jsu}m^{jr}a^{su}) \\ &\quad + 6\kappa_{jrsu}m^{jr}a^{su} + \kappa_{jrs}\kappa_{uvw}a^{vw}(m^{jr}m^{su}[3]) + (2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jsu}\kappa_{rvw})m^{jr}(a^{su}a^{vw}[3]), \end{aligned}$$

$$\begin{aligned}
A_2 &= 6D_u(\kappa_{jrs}m^{jr}m^{su}) - (\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jrv}\kappa_{suw})a^{vw}(m^{jr}m^{su}[3]) \\
&\quad - 3\kappa_{jrsu}m^{jr}m^{su} - \frac{3}{4}\kappa_{jrs}\kappa_{uvw}(3m^{jr}m^{su}m^{vw} + 2m^{ju}m^{rv}m^{sw}), \\
A_3 &= \frac{1}{4}\kappa_{jrs}\kappa_{uvw}(3m^{jr}m^{su}m^{vw} + 2m^{ju}m^{rv}m^{sw}).
\end{aligned}$$

com

$$D_j D_r(m^{jr}) = -\kappa_{uw}^{(j)}\kappa_{kl}^{(r)}(\kappa^{ju}\kappa^{wk}\kappa^{lr} + \kappa^{jk}\kappa^{lu}\kappa^{wr} + a^{ju}a^{wk}a^{lr} + a^{jk}a^{lu}a^{wr}) + \kappa_{kl}^{(rj)}(\kappa^{jk}\kappa^{lr} - a^{jk}a^{lr}),$$

$$D_u(\kappa_{jrs}m^{jr}m^{su}) = \kappa_{jrs}^{(u)}m^{jr}m^{su} + \kappa_{jrs}\kappa_{kl}^{(u)}(\kappa^{jk}\kappa^{lr}m^{su} - a^{jk}a^{lr}m^{su} + \kappa^{sk}\kappa^{lu}m^{jr} - a^{sk}a^{lu}m^{jr}),$$

$$D_u(\kappa_{jrs}m^{jr}a^{su}) = \kappa_{jrs}^{(u)}m^{jr}a^{su} + \kappa_{jrs}\kappa_{kl}^{(u)}(\kappa^{jk}\kappa^{lr}a^{su} - a^{jk}a^{lr}a^{su} + m^{jr}a^{sk}a^{lu})$$

$$D_r(\kappa_{jrs}m^{jr}a^{su}) = \kappa_{jrs}^{(r)}m^{jr}a^{su} + \kappa_{jrs}\kappa_{kl}^{(r)}(a^{sk}a^{lu}m^{jr} + a^{su}\kappa^{jk}\kappa^{lr} - a^{su}a^{jk}a^{lr}).$$

Invertendo  $M(t)$  em (2.19) e fazendo uma troca de índices adequada, após alguma álgebra, chegamos às expressões para  $A_1$ ,  $A_2$ , e  $A_3$  dadas no Teorema 2.1.

# Capítulo 3

## Aperfeiçoamento do teste gradiente em modelos lineares generalizados

### Resumo

Neste capítulo apresentamos uma expansão assintótica para a função de distribuição da estatística gradiente, sob a hipótese nula e até ordem  $n^{-1}$ , em que  $n$  é o tamanho da amostra, em modelos lineares generalizados com parâmetro de dispersão conhecido e desconhecido. Esta expansão envolve quantidades que dependem de momentos de derivadas do logaritmo da função de verossimilhança. Fórmulas matriciais gerais de fácil implementação computacional para essas quantidades são derivadas. A partir dessa expansão, propomos uma estatística gradiente corrigida por um fator de correção tipo-Bartlett, que possui distribuição  $\chi_q^2$  até ordem  $n^{-1}$ , em que  $q$  é o número de restrições impostas na hipótese nula. Estudos de simulação foram conduzidos com o objetivo de verificar o desempenho do teste que usa esta estatística corrigida e compará-lo com a versão não corrigida do teste gradiente, e com os testes da razão de verossimilhança e score, bem como suas respectivas versões corrigidas, e com o teste de Wald.

*Palavras chave: Correção de Bartlett; Correção tipo-Bartlett; Estatística da razão de verossimilhanças; Estatística de Wald; Estatística score; Estatística gradiente; Modelos lineares generalizados.*

## 3.1 Introdução

Os modelos lineares generalizados (MLG), definidos por Nelder e Wedderburn (1972), compõem uma importante classe de modelos com ampla utilização em análise de dados. Esta classe abrange vários modelos, dentre eles o clássico modelo de regressão linear com erro normal. Além disso, permitem uma flexibilidade para a ligação entre a média da variável resposta e a estrutura linear do modelo. Estudos mais detalhados envolvendo MLG podem ser vistos em McCullagh e Nelder (1989).

Os métodos usuais de testes de hipóteses em MLG baseiam-se nos testes da razão de verossimilhanças, score e de Wald. Alguns trabalhos sobre a utilização desses testes em MLG podem ser vistos em Pregibon (1982), Davison e Tsai (1990), Cook e Weisberg (1983), Atkinson (1982), dentre outros. Cordeiro (1983) derivou fórmulas matriciais para o fator de correção de Bartlett do teste de razão de verossimilhanças nos MLG com dispersão conhecida e, posteriormente, com dispersão desconhecida (Cordeiro, 1987). A partir dos resultados de Cordeiro e Ferrari (1991), Cordeiro, Ferrari e Paula (1993) propuseram fórmulas matriciais para o fator de correção tipo-Bartlett da estatística score em MLG com dispersão conhecida e, posteriormente, Cribari-Neto e Ferrari (1995), consideraram o caso em que a dispersão é desconhecida.

O teste gradiente foi originalmente proposto por Terrell (2002) e é uma alternativa aos testes mais utilizados em grandes amostras, a saber, os testes da razão de verossimilhanças, score e de Wald. Esta estatística possui as mesmas propriedades assintóticas de primeira ordem que as outras três estatísticas. Lemonte e Ferrari (2012b) derivaram uma expansão para a função de distribuição da estatística gradiente para hipóteses compostas sob alternativas de Pitman convergindo para a hipótese nula com taxa  $n^{-1/2}$ , compararam o poder local do teste gradiente com os testes da razão de verossimilhanças, score e de Wald, e concluíram que nenhum dos testes é uniformemente mais poderoso que o outro. Podemos citar na literatura o trabalho de Lemonte (2012a) sobre a utilização do teste gradiente em MLG com dispersão variável. O autor recomenda, a partir de resultados de simulação os testes baseados nas estatísticas score e gradiente, pois apresentam menor distorção em amostras de tamanhos pequeno e moderado.

No Capítulo 2, derivamos uma expansão sob a hipótese nula para a função de distribuição da es-

tatística gradiente até ordem  $n^{-1}$ . A partir disso, obtemos aqui esta expansão em MLG com parâmetro de dispersão desconhecido. Essa expansão envolve quantidades que são funções de momentos do logaritmo da função de verossimilhança. Expressamos essas quantidades em notação matricial. A partir disso, obtemos uma estatística gradiente corrigida por um fator de correção tipo-Bartellett, que possui distribuição  $\chi_q^2$  até um erro de ordem  $o(n^{-1})$ , em que  $q$  é o número de restrições impostas na hipótese nula. Simulações Monte Carlo são apresentadas com o objetivo de verificar o desempenho do teste gradiente corrigido para testar hipóteses compostas em amostras pequenas. Comparações com o teste gradiente não corrigido e com os testes da razão de verossimilhanças e escore, bem como suas respectivas versões corrigidas, e com o teste baseado na estatística de Wald, são apresentadas.

Este capítulo se divide como se segue. Na Seção 3.2 apresentamos os testes da razão de verossimilhanças, escore, gradiente e de Wald em MLG; na Seção 3.3 fazemos um breve comentário sobre aperfeiçoamento do teste gradiente; nas Seções 3.4 e 3.5 derivamos fórmulas auxiliares para as quantidades envolvidas no aperfeiçoamento do teste gradiente em MLG. A Seção 3.6 se destina à apresentação de resultados já conhecidos para o aperfeiçoamento dos testes da razão de verossimilhanças e escore em MLG. As Seções 3.7 e 3.8 mostram simulações Monte Carlo para a avaliação de testes de hipóteses em casos específicos da classe dos MLG, inclusive do teste gradiente corrigido. A Seção 3.9 apresenta aplicações dos resultados do capítulo a experimentos com dados reais. Finalizamos o capítulo com uma breve discussão (Seção 3.10). Os detalhes técnicos estão apresentados na Seção 3.11.

## 3.2 Testes de hipóteses em MLG

Seja  $\mathbf{y} = (y_1, \dots, y_n)^\top$  um vetor de variáveis aleatórias independentes em que cada  $y_l$  tem distribuição na família exponencial, cuja função densidade tem a forma

$$f(y, \theta_l, \phi) = \exp\{\phi[y\theta_l - b(\theta_l) + c(y)] + a(y, \phi)\}, \quad l = 1, \dots, n, \quad (3.1)$$

em que  $a(\cdot, \cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas. Aqui,  $\theta_l$  e  $\phi_l$  são possíveis parâmetros desconhecidos. Para o modelo (3.1), temos que  $E[y_l] = \mu_l = b'(\theta_l)$  e  $\text{Var}[y_l] = V_l/\phi$ , em que  $\phi^{-1}$  é o

parâmetro de dispersão e  $V_l = V(\mu_l) = d\mu_l/d\theta_l$  é a função de variância com  $\theta_l = \int V_l^{-1} d\mu_l = q(\mu_l)$ . Além disso,  $q(\cdot)$  é uma função estritamente monótona da média. As distribuições normal, Poisson, gama, normal inversa e binomial são alguns casos especiais de (3.1). Supomos que o vetor de variáveis auxiliares  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^\top$ , para  $j = 1, \dots, p$ , determinam uma estrutura linear da forma  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ , em que  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$  representa a matriz modelo de dimensão  $n \times p$  e posto completo, isto é,  $\text{posto}(\mathbf{X}) = p$  ( $p < n$ ), e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é um conjunto de parâmetros desconhecidos. Para distribuições na família exponencial na forma natural com parâmetros canônicos  $\phi$  e  $\phi\theta$ , temos  $a(y, \phi) = d_1(\phi) + d_2(y)$ . As distribuições contínuas como normal, normal inversa e gama permitem esta decomposição.

**Definição 3.1** *Um modelo linear generalizado (MLG) é definido por (3.1) e por uma função de ligação  $d(\boldsymbol{\mu}) = \boldsymbol{\eta}$ , estritamente monótona e duas vezes diferenciável, que relaciona o vetor  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ , em que  $\mu_l$  é a média de cada observação, com o vetor  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$ , o preditor linear do modelo.*

Consideramos a partição  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ , em que  $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$  e  $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$  com  $q \leq p$ , o que induz a partição  $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$  da matriz modelo. O interesse é o teste da hipótese composta  $\mathcal{H}_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$  versus  $\mathcal{H}_a: \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$ , em que  $\boldsymbol{\beta}_{10}$  é um vetor de constantes conhecidas e  $\boldsymbol{\beta}_2$  representa um vetor de parâmetros de perturbação. Assumimos por enquanto que o parâmetro de dispersão  $\phi^{-1}$  é conhecido. Seja  $L$  o logaritmo da função de verossimilhança total, dada por

$$L(\boldsymbol{\beta}) = \sum_{l=1}^n [y_l \theta_l - b(\theta_l) + c(y_l)] + \sum_{l=1}^n a(y_l, \phi).$$

Seja  $\mathbf{K} = E[\mathbf{U}\mathbf{U}^\top]$  a matriz de informação de Fisher para  $\boldsymbol{\beta}$  e  $\mathbf{U} = \partial L / \partial \boldsymbol{\beta}$  a função escore total, que pode ser escrita como  $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta}) = \phi \mathbf{X}^\top \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$ , em que  $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$ ,  $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$  com  $w_l = (d\mu_l/d\eta_l)^2 / V_l$ . No Capítulo 2, a função escore foi convenientemente dividida por  $n$  seguindo a notação de Ghosh e Mukerjee (1991). No entanto, adequamos aqui essa notação ao padrão usual utilizado em Cordeiro (1983, 1987), Cordeiro, Ferrari e Paula (1993) e Cribari-Neto e Ferrari (1995), em que a função escore não está dividida por  $n$ . A matriz de informação de Fisher pode ser escrita como  $\mathbf{K} = -((\kappa_{jr})) = \phi \mathbf{X}^\top \mathbf{W} \mathbf{X}$ .

Seja  $\widehat{\beta} = (\widehat{\beta}_1^\top, \widehat{\beta}_2^\top)^\top$  o estimador de máxima verossimilhança irrestrito de  $\beta$  e  $\widetilde{\beta}_2$  o estimador de máxima verossimilhança restrito de  $\beta_2$  obtido sob  $\mathcal{H}_0$ . As funções avaliadas em  $\widetilde{\beta} = (\beta_{10}^\top, \widetilde{\beta}_2^\top)^\top$  e  $\widehat{\beta} = (\widehat{\beta}_1^\top, \widehat{\beta}_2^\top)^\top$  são indicadas pela adição de um ‘til’ sobrescrito e de um ‘chapéu’, respectivamente. De acordo com a partição de  $\beta$  a função escore é particionada como  $U(\beta) = (U_1(\beta)^\top, U_2(\beta)^\top)^\top$ , e a matriz  $K$  nas submatrizes  $K_{11} = \phi X_1^\top W X_1$ ,  $K_{22} = \phi X_2^\top W X_2$  e  $K_{12} = K_{21}^\top = \phi X_1^\top W X_2$ . A inversa da matriz de informação de Fisher particionada é dada por

$$K^{-1} = -((\kappa^{jr})) = \begin{bmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{bmatrix},$$

em que  $K^{11} = \phi^{-1}(R^\top W R)^{-1}$ ,  $K^{12} = (K^{21})^\top = -\phi^{-1}(R^\top W R)^{-1}C^\top$ ,

$$K^{22} = \phi^{-1}[(X_2^\top W X_2)^{-1} + C(R^\top W R)^{-1}C^\top],$$

com  $R = X_1 - X_2C$  e  $C = (X_2^\top W X_2)^{-1}X_2^\top W X_1$ . Seguindo a partição de  $\beta$ , definimos as matrizes  $A$  e  $M$  como

$$A = ((a^{jr})) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K_{22}^{-1} \end{bmatrix}, \quad M = ((m^{jr})) = K^{-1} - A.$$

Assim a estatística gradiente em MLG para testar  $\mathcal{H}_0$  é dada por

$$S_T = \phi^{1/2} \widetilde{s}^\top \widetilde{W}^{1/2} X_1 (\widehat{\beta}_1 - \beta_{10}). \quad (3.2)$$

As estatísticas da razão de verossimilhanças, escore, de Wald são dadas por  $S_{LR}$ ,  $S_R$  e  $S_W$ , respectivamente, sendo

$$S_{LR} = 2[L(\widehat{\beta}_1, \widehat{\beta}_2) - L(\beta_{10}, \widetilde{\beta}_2)], \quad (3.3)$$

$$S_R = \widetilde{s}^\top \widetilde{W}^{1/2} X_1 (R^\top \widetilde{W} R)^{-1} X_1^\top \widetilde{W}^{1/2} \widetilde{s}, \quad (3.4)$$

$$S_W = \phi(\widehat{\beta}_1 - \beta_{10})^\top (\widehat{R}^\top \widehat{W} \widehat{R})(\widehat{\beta}_1 - \beta_{10}), \quad (3.5)$$

em que  $s = \phi^{1/2} V^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$  é o vetor de resíduos de Pearson. As estatísticas  $S_T$ ,  $S_{LR}$ ,  $S_R$  e  $S_W$  possuem distribuição  $\chi_q^2$  sob a hipótese nula assintoticamente.

### 3.3 Aperfeiçoamento do teste gradiente

Sejam  $U_i = \partial L / \partial \beta_i$ ,  $U_{ij} = \partial^2 L / \partial \beta_i \partial \beta_j$ , etc., derivadas do logaritmo da função de verossimilhança.

No que se segue, todos os índices variam de 1 a  $p$ . Defina

$$\begin{aligned} \kappa_{ij} &= E(U_{ij}), \quad \kappa_{ijk} = E(U_{ijk}), \quad \kappa_{ijkl} = E(U_{ijkl}), \\ \kappa_{ij}^{(r)} &= \partial \kappa_{ij} / \partial \beta_r, \quad \kappa_{ij}^{(kr)} = \partial^2 \kappa_{ij} / \partial \beta_k \partial \beta_r, \quad \kappa_{ijk}^{(r)} = \partial \kappa_{ijk} / \partial \beta_r, \end{aligned}$$

e assim por diante, momentos e derivadas de momentos do logaritmo da função de verossimilhança total. O Teorema 2.1 apresenta uma expansão assintótica sob a hipótese nula para a função distribuição acumulada da estatística gradiente até ordem  $o(n^{-1})$ , que pode ser escrita na forma

$$\begin{aligned} \Pr(S_T \leq c) &= G_q(c) + \frac{1}{24} \{ A_3 G_{q+6}(c) (A_2 - 3A_3) G_{q+4}(c) + (A_1 - 2A_2 + 3A_3) G_{q+2}(c) \\ &\quad + (A_2 - A_1 - A_3) G_q(c) \} + o(n^{-1}), \end{aligned} \quad (3.6)$$

em que  $G_r(c) = \Pr(\chi_r^2 \leq c)$ . As quantidades  $A_1, A_2$  e  $A_3$  em (3.6) são funções de momentos do logaritmo da função de verossimilhança, e são dadas no Teorema 2.1. A partir de (3.6), foi apresentada no Corolário 2.2 uma correção tipo-Bartlett para a estatística gradiente. A correção tipo-Bartlett é um fator multiplicativo, ou seja, um polinômio de segundo grau na própria estatística que resulta em uma estatística modificada que possui distribuição  $\chi^2$  até um erro de ordem  $o(n^{-1})$  sob a hipótese nula. Assim, a estatística modificada, obtida usando resultados de Cordeiro e Ferrari (1991) (ver Corolário 2.2), é da forma

$$S_T^* = S_T [1 - (c + bS_T + aS_T^2)], \quad (3.7)$$

com

$$a = \frac{A_3}{12q(q+2)(q+4)}, \quad b = \frac{A_2 - 2A_3}{12q(q+2)}, \quad c = \frac{A_1 - A_2 + A_3}{12q}.$$

A estatística  $S_T^*$  tem distribuição melhor aproximada pela distribuição qui-quadrado de referência do que a distribuição de  $S_T$ . Utilizando a estatística  $S_T^*$  e a distribuição  $\chi_q^2$  como referência, temos um aperfeiçoamento do teste baseado na estatística gradiente. Outra forma de melhorar o teste gradiente, equivalente até ordem  $n^{-1}$ , é comparar o valor observado da estatística  $S_T$  com o percentil superior



modificado dado pela inversão de (3.6) (Hill e Davis, 1968). Denotando por  $\gamma$  o nível de significância desejado do teste e  $x_{1-\gamma}$  o percentil de ordem  $1 - \gamma$  da distribuição  $\chi_q^2$ , o Corolário 2.3 fornece a expansão assintótica para o percentil  $1 - \gamma$  de  $S_T$ . Temos

$$z_{1-\gamma} = x_{1-\gamma} + \frac{1}{12} \left[ \frac{A_3 x_{1-\gamma}}{q(q+2)(q+4)} \{x_{1-\gamma}^2 + (q+4)x_{1-\gamma} + (q+2)(q+4)\} + \frac{x_{1-\gamma}(x_{1-\gamma} + q + 2)}{q(q+2)} (A_2 - 3A_3) + \frac{x_{1-\gamma}}{q} (3A_3 - 2A_2 + A_1) \right] + o(n^{-1}), \quad (3.8)$$

em que  $\Pr(\chi_q^2 \geq x_{1-\gamma}) = \gamma$ .

### 3.4 Aperfeiçoamento do teste gradiente em MLG com $\phi$ conhecido

Apresentamos nesta seção fórmulas matriciais em forma fechada para os coeficientes que definem a estatística gradiente corrigida em MLG assumindo primeiramente que o parâmetro  $\phi$  é conhecido.

Defina

$$(r)_l = \partial \eta_l / \partial \beta_r = x_{lr}, \quad (rs)_l = (\partial \eta_l / \partial \beta_r)(\partial \eta_l / \partial \beta_s) = x_{lr} x_{ls},$$

$$(rsu)_l = (\partial \eta_l / \partial \beta_r)(\partial \eta_l / \partial \beta_s)(\partial \eta_l / \partial \beta_u) = x_{lr} x_{ls} x_{lu},$$

$$(rstu)_l = (\partial \eta_l / \partial \beta_r)(\partial \eta_l / \partial \beta_s)(\partial \eta_l / \partial \beta_t)(\partial \eta_l / \partial \beta_u) = x_{lr} x_{ls} x_{lt} x_{lu}.$$

Em MLG, temos que (Ferrari, Lucâmbio, Cribari-Neto, 2005)

$$\kappa_{ijk} = -\phi \sum_{l=1}^n (f + 2g)_l (ijk)_l, \quad \kappa_{ij}^{(k)} = -\phi \sum_{l=1}^n (f + g)_l (ijk)_l,$$

$$\kappa_{ijk}^{(r)} = -\phi \sum_{l=1}^n t_l (ijk_r)_l, \quad \kappa_{ij}^{(kr)} = -\phi \sum_{l=1}^n d_l (ijk_r)_l, \quad \kappa_{ijk_r} = -\phi \sum_{l=1}^n e_l (ijk_r)_l,$$

em que  $f_l, g_l, t_l, d_l$  e  $e_l$  são escalares dados por

$$f = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2}, \quad g = f - \frac{1}{V^2} \frac{dV}{d\mu} \left( \frac{d\mu}{d\eta} \right)^3,$$

$$t = -9\lambda_1 + 3\lambda_2 + 3\lambda_3 + 4\lambda_4 - 2\lambda_5, \quad d = -5\lambda_1 + 2\lambda_2 + 2\lambda_3 + 2\lambda_4 - \lambda_5,$$

$$e = -12\lambda_1 + 3\lambda_2 + 4\lambda_3 + 6\lambda_4 - 3\lambda_5,$$

em que

$$\lambda_1 = \frac{1}{V^2} \frac{dV}{d\mu} \left( \frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2}, \quad \lambda_2 = \frac{1}{V} \left( \frac{d^2\mu}{d\eta^2} \right)^2, \quad (3.9)$$

$$\lambda_3 = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3}, \quad \lambda_4 = \frac{1}{V^3} \left( \frac{dV}{d\mu} \right)^2 \left( \frac{d\mu}{d\eta} \right)^4, \quad \lambda_5 = \frac{1}{V^2} \frac{d^2V}{d\mu^2} \left( \frac{d\mu}{d\eta} \right)^4. \quad (3.10)$$

Adicionalmente, sejam

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top = \{z_{lc}\}, \quad \mathbf{Z}_2 = \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top = \{z_{2lc}\},$$

$$\mathbf{Z}_d = \text{diag}(z_{11}, \dots, z_{nn}), \quad \mathbf{Z}_{2d} = \text{diag}(z_{211}, \dots, z_{2nn}),$$

$$\mathbf{F} = \text{diag}(f_1, \dots, f_n), \quad \mathbf{G} = \text{diag}(g_1, \dots, g_n),$$

$$\mathbf{T} = \text{diag}(t_1, \dots, t_n), \quad \mathbf{D} = \text{diag}(d_1, \dots, d_n), \quad \mathbf{E} = \text{diag}(e_1, \dots, e_n),$$

em que “ $\otimes$ ” denota o produto de Hadamard (produto elemento a elemento) entre duas matrizes,  $\mathbf{Z}^{(2)} = \mathbf{Z} \otimes \mathbf{Z}$ ,  $\mathbf{Z}^{(3)} = \mathbf{Z}^{(2)} \otimes \mathbf{Z}$ , etc. As matrizes  $\phi^{-1} \mathbf{Z}$  e  $\phi^{-1} \mathbf{Z}_2$  definem as estruturas de covariância assintótica de  $\mathbf{X}\hat{\boldsymbol{\beta}}$  e  $\mathbf{X}\widetilde{\boldsymbol{\beta}}_2$ , respectivamente.

Expressões para os  $A$ 's que definem a expansão da função distribuição acumulada da estatística gradiente em MLG com dispersão conhecida são obtidas substituindo os  $\kappa$ 's em  $A_1$ ,  $A_2$  e  $A_3$  dados no Teorema 2.1. Efetuando primeiramente as somas nos índices que variam no tamanho da amostra e posteriormente efetuando as somas nos índices que definem os parâmetros, podemos definir os  $A$ 's em forma matricial. Assim, teremos termos da forma  $-\sum' (i)_l \kappa^{ij} (j)_l$ ,  $\sum' (i)_l a^{ij} (j)_l$  e  $\sum' (i)_l m^{ij} (j)_l$ , em que  $\kappa^{ij}$ ,  $a^{ij}$  e  $m^{ij}$  são elementos da posição  $(i, j)$  das matrizes  $\mathbf{K}^{-1}$ ,  $\mathbf{A}$  e  $\mathbf{M}$ , respectivamente. Esses somatórios representam elementos de  $\phi^{-1} \mathbf{Z}$ ,  $\phi^{-1} \mathbf{Z}_2$  e  $\phi^{-1} (\mathbf{Z} - \mathbf{Z}_2)$ , respectivamente. A obtenção dessas fórmulas matriciais foi desenvolvida de forma pioneira em Cordeiro (1983), que obteve uma expressão matricial para o fator de correção de Bartlett da estatística da razão de verossimilhanças em MLG e, posteriormente, para o fator de correção tipo-Bartlett para a estatística score em Cordeiro, Ferrari e Paula (1993). Na Seção 3.11.1 detalhamos o cálculo de  $A_1$ ,

$A_2$  e  $A_3$ . As expressões resultantes são

$$\begin{aligned}
A_1 = & 12\phi^{-1}\mathbf{1}^\top(\mathbf{F} + \mathbf{G})[\mathbf{Z}_d\mathbf{Z}\mathbf{Z}_d - \mathbf{Z}_{2d}\mathbf{Z}_2\mathbf{Z}_{2d} + \mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)}](\mathbf{F} + \mathbf{G})\mathbf{1} \\
& - 6\phi^{-1}\mathbf{1}^\top(\mathbf{F} + 2\mathbf{G})[(\mathbf{Z} + \mathbf{Z}_2) \otimes (\mathbf{Z}^{(2)} - \mathbf{Z}_2^{(2)}) + (\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z}\mathbf{Z}_d + \mathbf{Z}_2\mathbf{Z}_{2d}) \\
& + 2\mathbf{Z}_{2d}(\mathbf{Z}\mathbf{Z}_d - \mathbf{Z}_2\mathbf{Z}_{2d}) + 2\mathbf{Z}_2^{(2)} \otimes (\mathbf{Z} - \mathbf{Z}_2)](\mathbf{F} + \mathbf{G})\mathbf{1} \\
& + 3\phi^{-1}\mathbf{1}^\top(\mathbf{F} + 2\mathbf{G})[2(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_2\mathbf{Z}_{2d} + 2\mathbf{Z}_2^{(2)} \otimes (\mathbf{Z} - \mathbf{Z}_2) \\
& + \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{Z}_{2d} + \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d](\mathbf{F} + 2\mathbf{G})\mathbf{1} \\
& - 12\phi^{-1}\mathbf{1}^\top\mathbf{D}(\mathbf{Z}_d^{(2)} - \mathbf{Z}_{2d}^{(2)})\mathbf{1} + 6\phi^{-1}\mathbf{1}^\top\mathbf{T}(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z}_d + 3\mathbf{Z}_{2d})\mathbf{1} \\
& - 6\phi^{-1}\mathbf{1}^\top\mathbf{E}(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_{2d}\mathbf{1}, \tag{3.11}
\end{aligned}$$

$$\begin{aligned}
A_2 = & -3\phi^{-1}\mathbf{1}^\top(\mathbf{F} + 2\mathbf{G})\left[\frac{3}{4}(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d + \frac{1}{2}(\mathbf{Z} - \mathbf{Z}_2)^{(3)} \right. \\
& + \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d + (\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_2(\mathbf{Z} - \mathbf{Z}_2)_d \\
& \left. + 2\mathbf{Z}_2 \otimes (\mathbf{Z} - \mathbf{Z}_2)^{(2)}\right](\mathbf{F} + 2\mathbf{G})\mathbf{1} - 3\phi^{-1}\mathbf{1}^\top(2\mathbf{T} - \mathbf{E})(\mathbf{Z} - \mathbf{Z}_2)_d^{(2)}\mathbf{1} \\
& + 6\phi^{-1}\mathbf{1}^\top(\mathbf{F} + 2\mathbf{G})[(\mathbf{Z} - \mathbf{Z}_2) \otimes (\mathbf{Z}^{(2)} - \mathbf{Z}_2^{(2)}) \\
& + (\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z}\mathbf{Z}_d - \mathbf{Z}_2\mathbf{Z}_{2d})](\mathbf{F} + \mathbf{G})\mathbf{1}, \tag{3.12}
\end{aligned}$$

$$A_3 = \phi^{-1}\mathbf{1}^\top(\mathbf{F} + 2\mathbf{G})\left[\frac{3}{4}(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d + \frac{1}{2}(\mathbf{Z} - \mathbf{Z}_2)^{(3)}\right](\mathbf{F} + 2\mathbf{G})\mathbf{1}, \tag{3.13}$$

em que  $\mathbf{1} = (1, \dots, 1)^\top$  é um vetor  $n$ -dimensional de 1's.

Assim, obtemos os termos que definem (3.6) em MLG com  $\phi$  conhecido. A fim de obter a estatística gradiente corrigida para o teste de hipótese proposto na Seção 3.2, substituímos os  $A$ 's dados em (3.11)–(3.13) em (3.7). Podemos obter os percentis ajustados de  $S_T$  substituindo (3.11)–(3.13) em (3.8). Na eventual presença de parâmetros desconhecidos, substituímos estes por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula.

Se a hipótese nula é simples, então  $\mathbf{Z}_2 = \mathbf{0}$  e as expressões para os  $A$ 's se reduzem de forma significativa. Para modelos com ligação identidade ( $\mu = \eta$ ), temos  $\mathbf{F} = \mathbf{0}$ , porém, os  $A$ 's não sofrem grande redução, exceto no caso de resposta normal ( $V = 1$ ), em que  $A_1 = A_2 = A_3 = 0$ , como

esperado. Em modelos com ligação canônica ( $\theta = \eta$ ), temos  $\mathbf{G} = \mathbf{0}$ , mas os  $A$ 's não sofrem grandes simplificações.

Como observado em Cordeiro, Ferrari e Paula (1993) para o caso do aperfeiçoamento do teste escore em MLG com  $\phi$  conhecido, as quantidades  $A_1$ ,  $A_2$  e  $A_3$  em (3.11)–(3.13) dependem da matriz do modelo do parâmetro de dispersão e parâmetros desconhecidos. Observa-se nesses coeficientes a dependência das funções de ligação e variância bem como suas derivadas. Observe que a implementação computacional dessas expressões é muito simples, pois envolve somente operações matriciais básicas. No entanto, tais quantidades não são de fácil interpretação, uma vez que dependem do sistema de coordenadas adotado.

### 3.5 Aperfeiçoamento do teste gradiente em MLG com $\phi$ desconhecido

Seja  $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$  um vetor de parâmetros  $(p + 1)$ -dimensional, em que  $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ ,  $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$  e  $\phi$  um escalar. O objetivo é testar a hipótese nula da forma  $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ , em que  $(\boldsymbol{\beta}_2^\top, \phi)^\top$  é o vetor de parâmetros de perturbação. A estatística gradiente para testar  $\mathcal{H}_0$  é da forma

$$\bar{S}_T = \tilde{\phi}^{1/2} \tilde{\mathbf{s}}^\top \tilde{\mathbf{W}}^{-1/2} \mathbf{X}_1 (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}), \quad (3.14)$$

em que  $(\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top, \hat{\phi})^\top$  e  $(\boldsymbol{\beta}_{10}^\top, \tilde{\boldsymbol{\beta}}_2^\top, \tilde{\phi})^\top$  são os estimadores de máxima verossimilhança irrestrito e restrito (sob  $\mathcal{H}_0$ ) de  $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$ , respectivamente. A matriz de informação de Fisher de  $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$  é bloco-diagonal devido à ortogonalidade global (Cox e Reid, 1987) dos parâmetros  $\boldsymbol{\beta}$  e  $\phi$ , ou seja,

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{0} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\kappa_{\phi\phi} \end{bmatrix},$$

em que  $\kappa_{\phi\phi} = E(\partial^2 L / \partial \phi^2)$ , e sua inversa assume a forma

$$\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{K}_\beta^{-1} & \mathbf{0} \\ \mathbf{0} & -\kappa_{\phi\phi}^{-1} \end{bmatrix},$$

com

$$\mathbf{K}_\beta^{-1} = \begin{bmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{bmatrix}.$$

As estatísticas da razão de verossimilhanças, escore e de Wald para testar  $\mathcal{H}_0$  são dadas por

$$\bar{S}_{LR} = 2[L(\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top, \phi) - L(\boldsymbol{\beta}_{10}^\top, \tilde{\boldsymbol{\beta}}_2^\top, \tilde{\phi})], \quad (3.15)$$

$$\bar{S}_R = \tilde{\mathbf{s}}^\top \tilde{\mathbf{W}}^{1/2} \mathbf{X}_1 (\mathbf{R}^\top \tilde{\mathbf{W}} \mathbf{R})^{-1} \mathbf{X}_1^\top \tilde{\mathbf{W}}^{1/2} \tilde{\mathbf{s}}, \quad (3.16)$$

$$\bar{S}_W = \hat{\phi}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^\top (\hat{\mathbf{R}}^\top \hat{\mathbf{W}} \hat{\mathbf{R}})(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}), \quad (3.17)$$

respectivamente, em que  $\mathbf{s} = \phi^{1/2} \mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$  é o vetor de resíduos de Pearson.

Defina as matrizes

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_\beta & \mathbf{0} \\ \mathbf{0} & -\kappa_{\phi\phi}^{-1} \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

em que  $\mathbf{M}_\beta = \mathbf{K}_\beta^{-1} - \mathbf{A}_\beta$ . Temos que,  $m^{r\phi}$  e  $a^{r\phi}$  correspondem aos respectivos elementos da posição  $(r, p+1)$  de  $\mathbf{M}$  e  $\mathbf{A}$ . As entradas  $m^{\phi\phi}$  e  $a^{\phi\phi}$  correspondem aos elementos da posição  $(p+1, p+1)$  de  $\mathbf{M}$  e  $\mathbf{A}$  respectivamente. Temos  $m^{r\phi} = m^{\phi r} = m^{\phi\phi} = a^{r\phi} = a^{\phi r} = 0$  ( $r = 1, \dots, p$ ), com  $a^{\phi\phi} = -\kappa_{\phi\phi}^{-1}$ .

O fato da ortogonalidade global de  $\boldsymbol{\beta}$  e  $\phi$  é crucial na determinação dos  $A$ 's que definem a expansão para a função distribuição acumulada da estatística gradiente em MLG sob a hipótese nula com  $\phi$  desconhecido. Assim, cada um dos  $A$ 's dados no Teorema 2.1 são decompostos de forma que as parcelas que estão em função de  $\kappa_{r\phi}$ ,  $m^{r\phi}$ ,  $m^{\phi r}$ ,  $a^{r\phi}$  e  $a^{\phi r}$  ( $r = 1, \dots, p$ ) se anulam devido a essa ortogonalidade. Consequentemente, os  $A$ 's serão compostos de dois termos. O primeiro termo está em função somente de momentos envolvendo o parâmetro  $\boldsymbol{\beta}$ , ou seja, coincide com as quantidades dadas na Seção 3.4 para  $\phi$  conhecido. O segundo termo está em função dos momentos envolvendo os parâmetros  $\boldsymbol{\beta}$  e  $\phi$ . Portanto, denotando por  $A_{T1}$ ,  $A_{T2}$  e  $A_{T3}$  os  $A$ 's para o caso  $\phi$  desconhecido, temos

$$A_{T1} = A_1 + A_{1,\beta\phi}, \quad A_{T2} = A_2 + A_{2,\beta\phi}, \quad A_{T3} = A_3 + A_{3,\beta\phi}, \quad (3.18)$$

em que  $A_1$ ,  $A_2$  e  $A_3$  são dados por (3.11)–(3.13),  $A_{3,\beta\phi} = 0$ ,

$$\begin{aligned}
 A_{1,\beta\phi} = & \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{jr} m^{su} + \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{js} m^{ru} \\
 & + \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{ju} m^{rs} + 6 \sum' \kappa_{jr\phi} \kappa_{\phi\phi\phi} m^{jr} (a^{\phi\phi})^2 \\
 & + 6 \sum' \kappa_{jr\phi} \kappa_{vw\phi} m^{jr} a^{vw} a^{\phi\phi} + 6 \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{jr} a^{su} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{uvw} m^{ju} a^{vw} a^{\phi\phi} + 3 \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{ju} a^{rs} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{u\phi\phi} m^{ju} (a^{\phi\phi})^2 + 3 \sum' \kappa_{jr\phi} \kappa_{uw\phi} m^{ju} a^{rw} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{js\phi} \kappa_{uv\phi} m^{ju} a^{sv} a^{\phi\phi} + 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} a^{rv} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{js\phi} \kappa_{uw\phi} m^{ju} a^{sw} a^{\phi\phi} - 6 \sum' (\kappa_{jr\phi}^{(\phi)} - \kappa_{jr\phi\phi}) m^{jr} a^{\phi\phi} \\
 & - 6 \sum' \kappa_{jr\phi}^{(\phi)} m^{jr} a^{\phi\phi} - 12 \sum' \kappa_{j\phi\phi}^{(u)} m^{ju} a^{\phi\phi} \\
 & - 12 \sum' \kappa_{kl}^{(\phi)} \kappa_{jr\phi} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) a^{\phi\phi} - 12 \sum' \kappa_{\phi\phi}^{(\phi)} \kappa_{jr\phi} m^{jr} (a^{\phi\phi})^2 \\
 & - 12 \sum' \kappa_{j\phi\phi} \kappa_{kl}^{(u)} (\kappa^{jk} \kappa^{lu} - a^{jk} a^{lu}) a^{\phi\phi}, \tag{3.19}
 \end{aligned}$$

$$\begin{aligned}
 A_{2,\beta\phi} = & - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{jr} m^{su} a^{\phi\phi} - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{js} m^{ru} a^{\phi\phi} \\
 & - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{ju} m^{rs} a^{\phi\phi} - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{jr} m^{uv} a^{\phi\phi} \\
 & - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} m^{rv} a^{\phi\phi} - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} m^{ru} a^{\phi\phi}. \tag{3.20}
 \end{aligned}$$

Como mencionado na Seção 3.2, para a família exponencial biparamétrica de posto completo com parâmetros canônicos,  $\phi$  e  $\phi\theta$ , podemos reescrever o termo  $a(y, \phi)$  da equação (3.1) como  $a(y, \phi) = d_1(\phi) + d_2(y)$ . Daqui em diante, assuma que  $a(y, \phi)$  tem esta forma. Seja  $d_{(2)} = \phi^2 d_1''(\phi)$  e  $d_{(3)} = \phi^3 d_1'''(\phi)$ . Temos os seguintes momentos (Cordeiro, 1987):

$$\kappa_{\phi\phi} = n d_1''(\phi), \quad \kappa_{\phi\phi\phi} = \kappa_{\phi\phi}^{(\phi)} = n d_1'''(\phi), \quad \kappa_{ij\phi} = \kappa_{ij}^{(\phi)} = - \sum_{l=1}^n w_l(ij)_l, \tag{3.21}$$

$$\kappa_{i\phi\phi} = \kappa_{i\phi\phi}^{(j)} = \kappa_{ij\phi}^{(\phi)} = \kappa_{jr\phi\phi} = 0. \tag{3.22}$$

Substituindo esses momentos em (3.19) e (3.20), obtemos, após alguma álgebra (ver Seção 3.11.2),

$$A_{1,\beta\phi} = \frac{6q[d_{(3)} + (2 - p + q)d_{(2)}]}{n d_{(2)}^2}, \tag{3.23}$$

$$A_{2,\beta\phi} = \frac{3q(q+2)}{nd_{(2)}}. \quad (3.24)$$

As quantidades  $A_{T1}$ ,  $A_{T2}$  e  $A_{T3}$ , que são utilizadas para definir uma estatística gradiente corrigida em modelos lineares generalizados quando o parâmetro de dispersão é desconhecido, são dadas por (3.18) substituindo  $\phi$  por sua estimativa de máxima verossimilhança sob  $\mathcal{H}_0$ , ou seja,  $\tilde{\phi}$ . Os coeficientes  $A_{1,\beta\phi}$  e  $A_{2,\beta\phi}$  dependem do número de parâmetros do modelo, do número de restrições e independem da matriz modelo  $\mathbf{X}$ .

Para o modelo normal,  $d_{(2)} = -1/2$  e  $d_{(3)} = 1$ , assim,  $A_{1,\beta\phi} = 12q(p-q)/n$  e  $A_{2,\beta\phi} = -6q(q+2)/n$ . Se  $p = q$ , então  $\mathbf{Z}_2 = 0$ ,  $A_{1,\beta\phi} = 6p(d_{(3)} + 2d_{(2)})/(nd_{(2)}^2)$  e  $A_{2,\beta\phi} = 3p(p+2)/(nd_{(2)})$ .

Obtidos os  $A_T$ 's, obtemos a estatística gradiente corrigida para testar  $\mathcal{H}_0$  como

$$\bar{S}_T^* = \bar{S}_T [1 - (\bar{c} + \bar{b}\bar{S}_T + \bar{a}\bar{S}_T^2)], \quad (3.25)$$

em que  $\bar{a} = A_{T3}/[12q(q+2)(q+4)]$ ,  $\bar{b} = (A_{T2} - 2A_{T3})/[12q(q+2)]$ , e  $\bar{c} = (A_{T1} - A_{T2} + A_{T3})/(12q)$ . Uma expansão assintótica para o percentil da distribuição  $\chi_q^2$  de referência, como discutido na Seção 3.3, pode também ser obtido.

## 3.6 Aperfeiçoamento dos testes LR e escore em MLG com $\phi$ desconhecido

O aperfeiçoamento dos testes baseados nas estatísticas da razão de verossimilhanças e escore em MLG com  $\phi$  desconhecido é o tema dessa seção. Considere o teste da hipótese nula  $\mathcal{H}_0$  apresentada na Seção 3.5. A estatística da razão de verossimilhanças para testar essa hipótese nula é dada por (3.15). Cordeiro (1987) derivou fórmulas matriciais para o fator de correção de Bartlett da estatística LR em MLG com dispersão desconhecida, resultando na estatística LR corrigida com distribuição  $\chi_q^2$  até ordem  $n^{-1}$  da forma

$$\bar{S}_{LR}^* = \frac{\bar{S}_{LR}}{1 + a_{LR}}, \quad (3.26)$$

em que  $a_{LR} = A_{LR}/(12q)$ ,  $A_{LR} = A_{LR1} + A_{LR1,\beta\phi}$  e

$$\begin{aligned} A_{LR1} &= -4\phi^{-1}\mathbf{1}^\top \mathbf{G}(\mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)})(\mathbf{F} + \mathbf{G})\mathbf{1} + 3\phi^{-1}\mathbf{1}^\top \mathcal{M}(\mathbf{Z}_d^{(2)} - \mathbf{Z}_{2d}^{(2)})\mathbf{1} \\ &\quad + \phi^{-1}\mathbf{1}^\top \mathbf{F}[2(\mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)}) + 3(\mathbf{Z}_d\mathbf{Z}\mathbf{Z}_d - \mathbf{Z}_{2d}\mathbf{Z}_2\mathbf{Z}_{2d})]\mathbf{F}\mathbf{1}, \end{aligned} \quad (3.27)$$

$$A_{LR1,\beta\phi} = \frac{3q}{nd_{(2)}^2}[d_{(2)}(2 + q - 2p) + 2d_{(3)}].$$

Em (3.27), a matriz  $\mathcal{M}$  é diagonal de ordem  $n$ , ou seja,  $\mathcal{M} = \text{diag}(\mathcal{M}_1, \dots, \mathcal{M}_n)$ , em que  $\mathcal{M} = -4\lambda_1 + \lambda_2 + 2\lambda_4 - \lambda_5$ .

Outras formas de expressar o fator de correção da estatística LR resultam em estatísticas corrigidas equivalentes até ordem  $n^{-1}$ , a saber, a correção obtida diretamente de Cordeiro e Ferrari (1991) em que o fator de correção é  $(1 - a_{LR})$ , ou mesmo o fator de correção  $\exp(-a_{LR})$ . Este último tem a vantagem de determinar uma estatística LR corrigida assumindo somente valores positivos.

Para testar  $\mathcal{H}_0$ , a estatística escore é dada por (3.16). Cordeiro, Ferrari e Paula (1993) obtiveram fórmulas matriciais para os coeficientes da expansão da estatística escore em MLG para  $\phi$  conhecido, e propuseram uma estatística escore corrigida. Posteriormente, Cribari-Neto e Ferrari (1995) obtiveram essas expressões para o caso  $\phi$  desconhecido. Ressaltamos que coeficientes da expansão para a estatística gradiente seguiram um raciocínio similar. Assim, a estatística escore corrigida pelo fator de correção tipo-Bartlett tem a forma

$$\bar{S}_R^* = \bar{S}_R[1 - (\bar{c}_R + \bar{b}_R\bar{S}_R + \bar{a}_R\bar{S}_R^2)], \quad (3.28)$$

em que  $\bar{a}_R = A_{R3}/[12q(q+2)(q+4)]$ ,  $\bar{b}_R = A_{R2} - 2A_{R3}/[12q(q+2)]$ ,  $\bar{c}_R = (A_{R1} - A_{R2} + A_{R3})/(12q)$ , com  $A_{R1} = A_{S1} + A_{S1,\beta\phi}$ ,  $A_{R2} = A_{S2} + A_{S2,\beta\phi}$ ,  $A_{R3} = A_{S3} + A_{S3,\beta\phi}$ , e os coeficientes em forma matricial são

$$\begin{aligned} A_{S1} &= 3\phi^{-1}\mathbf{1}^\top \mathbf{F}\mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{Z}_{2d}\mathbf{F}\mathbf{1} \\ &\quad + 6\phi^{-1}\mathbf{1}^\top \mathbf{F}\mathbf{Z}_{2d}\mathbf{Z}_2(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{F} - \mathbf{G})\mathbf{1} \\ &\quad - 6\phi^{-1}\mathbf{1}^\top \mathbf{F}[\mathbf{Z}_2^{(2)} \otimes (\mathbf{Z} - \mathbf{Z}_2)](2\mathbf{G} - \mathbf{F})\mathbf{1} \\ &\quad - 6\phi^{-1}\mathbf{1}^\top \mathbf{H}(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_{2d}\mathbf{1}, \end{aligned} \quad (3.29)$$



$$\begin{aligned}
A_{S2} = & -3\phi^{-1}\mathbf{1}^\top(\mathbf{F} - \mathbf{G})(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_2(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{F} - \mathbf{G})\mathbf{1} \\
& -6\phi^{-1}\mathbf{1}^\top\mathbf{F}\mathbf{Z}_2d(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{F} - \mathbf{G})\mathbf{1} \\
& -6\phi^{-1}\mathbf{1}^\top(\mathbf{F} - \mathbf{G})[(\mathbf{Z} - \mathbf{Z}_2)^{(2)} \otimes \mathbf{Z}_2](\mathbf{F} - \mathbf{G})\mathbf{1} \\
& +3\phi^{-1}\mathbf{1}^\top\mathbf{B}(\mathbf{Z} - \mathbf{Z}_2)_d^{(2)}\mathbf{1},
\end{aligned} \tag{3.30}$$

$$\begin{aligned}
A_{S3} = & 3\phi^{-1}\mathbf{1}^\top(\mathbf{F} - \mathbf{G})(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{F} - \mathbf{G})\mathbf{1} \\
& +2\phi^{-1}\mathbf{1}^\top(\mathbf{F} - \mathbf{G})(\mathbf{Z} - \mathbf{Z}_2)^{(3)}(\mathbf{F} - \mathbf{G})\mathbf{1}.
\end{aligned}$$

As matrizes  $\mathbf{B} = \text{diag}(b_1, \dots, b_n)$  e  $\mathbf{H} = \text{diag}(h_1, \dots, h_n)$ , em (3.29) e (3.30), em que  $b = \lambda_3 + \lambda_4$  e  $h = \lambda_1 + \lambda_5$ , dependem da função de ligação e derivadas da função de variância. As parcelas envolvendo o parâmetro desconhecido  $\phi$  são idênticas ao caso da estatística gradiente (ver Seção 3.5), ou seja,  $A_{S1,\beta\phi} = A_{1,\beta\phi}$ ,  $A_{S2,\beta\phi} = A_{2,\beta\phi}$  e  $A_{S3,\beta\phi} = 0$ .

Ressaltamos aqui que ainda não temos uma correção tipo-Bartlett para a estatística de Wald na presença de parâmetros de perturbação independente de modelos específicos. Larsen e Jupp (2003) apresentam versões da estatística de Wald invariantes sob mudança de parametrização, baseando-se na geometria de yokes, e versões corrigidas por um fator de correção tipo-Bartlett para essas estatísticas são apresentadas.

Nas seções seguintes apresentamos algumas simulações Monte Carlo para avaliar o desempenho do teste baseado na estatística gradiente corrigida ( $\bar{S}_T^*$ ) e comparar com os testes baseados em  $\bar{S}_W$ ,  $\bar{S}_{LR}$ ,  $\bar{S}_R$  e  $\bar{S}_T$ , e as versões corrigidas  $\bar{S}_{LR}^*$  e  $\bar{S}_R^*$  em amostras de tamanho pequeno. O desempenho dos testes é avaliado de acordo com a taxa de rejeição da hipótese nula de cada teste. Apresentamos também um pequeno estudo do poder de cada teste. Posteriormente, exemplos numéricos com dados reais são apresentados.

### 3.7 Resultados Numéricos: modelo de regressão gama

Considere a classe de MLG com resposta gama, ou seja, cada  $y_l$  tem densidade (3.1) com  $c(y) = \log(y)$  e  $a(y, \phi) = \phi \log(\phi) - \log \Gamma(\phi) - \log(y)$ , em que  $\Gamma(\phi)$  é a função gama,  $b(\theta) = -\log(-\theta)$ ,

$\mu = E(y) = -1/\theta$  e  $V(\mu) = \mu^2$ . Temos

$$\log(\mu_l) = \eta_l = \mathbf{x}_l^\top \boldsymbol{\beta}, \quad l = 1, \dots, n, \quad (3.31)$$

ou seja, a função de ligação é logarítmica. Então,

$$\mu_l = \exp \left( \sum_{j=1}^p \beta_j x_{lj} \right). \quad (3.32)$$

Adicionalmente,  $d\mu/d\eta = \mu$ ,  $w = f = -g = -\mathcal{M} = e = 1$ ,  $b = 6$ ,  $h = 4$  e  $t = d = 0$  são as expressões que definem os momentos do logaritmo da função de verossimilhança. Temos também  $d_{(3)} = \phi(1 + \phi^2\psi'')$  e  $d_{(2)} = \phi(1 - \phi\psi')$ , em que  $\psi' = \psi'(\phi)$  e  $\psi'' = \psi''(\phi)$  são, respectivamente, as funções trigama e tetragama.

Apresentamos aqui resultados de simulações Monte Carlo comparando o desempenho do teste gradiente corrigido ( $\bar{S}_T^*$ ) com o do não corrigido ( $\bar{S}_T$ ) e com a desempenho dos testes da razão de verossimilhanças ( $\bar{S}_{LR}$ ), teste escore ( $\bar{S}_R$ ), teste de Wald ( $\bar{S}_W$ ) e os testes LR ( $\bar{S}_{LR}^*$ ) e escore ( $\bar{S}_R^*$ ) corrigidos, no modelo de regressão gama com ligação logarítmica e parâmetro de dispersão desconhecido, utilizando a linguagem matricial  $\text{Ox}$  (Doornik, 2008), baseando-se em 15,000 réplicas. Consideramos a hipótese nula  $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ , em que  $\mathbf{0}$  é um  $q$ -vetor de zeros. Os outros parâmetros da regressão são iguais a 1 e  $\phi = 1$ . A primeira coluna de covariadas é um vetor de 1's e as demais covariadas foram geradas independentemente com distribuição uniforme  $U(0, 1)$ . O método quasi-Newton (ou BGFS) foi utilizado para a maximização das verossimilhanças. Os modelos em questão possuem  $p + 1$  parâmetros.

As taxas de rejeição da hipótese nula quando esta é verdadeira (erro do tipo I) consideradas para os níveis nominais  $\alpha = 10\%$ ,  $5\%$ ,  $1\%$  no modelo proposto são apresentadas na Tabela 3.1 para  $p = 4$  e  $q = 3, 2, 1$ ; na Tabela 3.2 para  $p = 5$  e  $q = 4, 3, 2$ ; na Tabela 3.3 para  $p = 6$  e  $q = 4, 3, 2$ . Os tamanhos de amostra variam em  $n = 20, 25, 30$ .

O teste de Wald ( $\bar{S}_W$ ) apresenta taxas de rejeição muito acima dos níveis de significância considerados, principalmente quando o número de parâmetros testados no modelo,  $q$ , aumenta. Por exemplo, quando  $n = 20$ ,  $p = 4$ ,  $q = 1$  e  $\alpha = 5\%$  (Tabela 3.1), a taxa de rejeição do teste de Wald foi de

11, 55%, ou seja, mais que o dobro do nível fixado. Nesse mesmo cenário, para  $q = 3$ , esta taxa sobe para 16, 73%, ou seja, deteriora-se ainda mais.

Variando o número de parâmetros do modelo em  $p = 4, 5, 6$ , fixando o tamanho da amostra em  $n = 20$ , o número de parâmetros testados em  $q = 2$  e o nível nominal  $\alpha = 10\%$ , verificamos que as taxas de rejeição do teste LR são 15, 25%, 17, 21% e 18, 80%. Apesar de serem menores do que os do teste de Wald, ainda estão bem acima do nível fixado, principalmente se compararmos com as taxas de rejeição dos testes escore (9, 41%, 10, 56%, 11, 20%) e gradiente (11, 71%, 13, 85%, 15, 24%). Os testes que utilizam as estatísticas corrigidas apresentaram um desempenho superior, visto que, nesta comparação, a rejeição nula para o teste que usa  $\bar{S}_{LR}^*$  variou em 10, 68%, 11, 96%, 11, 80%, o teste que usa  $\bar{S}_R^*$  em 9, 96%, 11, 05% e 10, 58% e do teste  $\bar{S}_T^*$  em 10, 07%, 11, 06% e 10, 27%. Nesse cenário, conclui-se que o teste LR apresenta taxas de rejeição distorcidas, sendo atenuadas pela correção de Bartlett, porém, mesmo o teste corrigido não apresenta desempenho superior aos testes escore e gradiente corrigidos, que apresentaram resultados mais acurados. Verifica-se ainda que, neste contexto, o comportamento do teste gradiente corrigido foi similar ao do teste escore corrigido, ambos com desempenho superior ao teste LR corrigido.

Fixando o número de parâmetros em  $p = 5$  (Tabela 3.2), o tamanho da amostra em  $n = 30$ , o nível nominal em  $\alpha = 5\%$  e variando o número de parâmetros testados no modelo em  $q = 2, q = 3$  e  $q = 4$ , podemos verificar que os testes de Wald e LR tendem a rejeitar a hipótese nula com uma frequência muito acima do nível nominal especificado. Os testes escore e gradiente não corrigidos apresentam uma distorção bem menor do que os testes Wald e LR nesse caso. Por exemplo, para  $q = 2$ , são de 4, 71% e 5, 85%, alterando-se pouco com a variação de  $q$ . As taxas de rejeição dos testes corrigidos se apresentaram bem mais próximas do nível nominal, como no caso  $q = 3$ , em que as taxas de rejeição dos testes que usam  $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$  e  $\bar{S}_T^*$  foram 5, 37%, 5, 04% e 4, 95%, respectivamente, indicando que os testes escore e gradiente corrigidos obtiveram desempenhos consideravelmente melhores.

No caso em que  $p = 6$  (Tabela 3.3), os testes Wald e LR tiveram desempenho semelhante aos casos  $p = 5$  e  $p = 4$ , ou seja, foram extremamente liberais. Ainda na Tabela 3.3, observa-se que o teste gradiente tende a rejeitar a hipótese nula acima do nível nominal para o caso em que  $q = 2$ .

Por exemplo, para  $n = 25$ , as taxas de rejeição são de 13,07% e 6,17% aos níveis  $\alpha = 10\%$  e 5%, respectivamente. Este comportamento é alterado pela correção tipo-Bartlett, pois, nesse caso, as taxas caem para 10,35% e 5,09%, taxas semelhantes ao teste escore corrigido (10,53% e 5,07%) e inferiores ao teste LR corrigido (11,07% e 5,61%). Isto indica que a correção tipo-Bartlett para a estatística gradiente foi efetiva, pois resultou em uma aproximação considerável das taxas de rejeição da hipótese nula com os níveis nominais especificados, obtendo, assim, desempenho competitivo frente aos outros testes corrigidos.

De modo geral, conclui-se que a correção tipo-Bartlett para a estatística gradiente foi efetiva com comportamento semelhante à estatística escore corrigida, tendo distorção menor do que o teste da razão de verossimilhanças corrigido. Os testes da razão de verossimilhanças, escore gradiente não-corrigidos e de Wald apresentaram comportamento mais liberal.

A Tabela 3.4 apresenta as taxas de rejeição não-nulas (poder). Foram fixados  $p = 4$ ,  $q = 2$ , tamanho de amostra  $n = 30$  e nível nominal  $\alpha = 5\%$ . Aqui, estamos levando em consideração a hipótese alternativa  $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$ , para diferentes valores de  $\delta$ . Não consideramos os testes LR, escore e gradiente não-corrigidos e o teste de Wald, pois estes já são mais liberais sob  $\mathcal{H}_0$  do que os testes aperfeiçoados. Como esperado o poder dos testes aumentam à medida que  $|\delta|$  cresce. Fixando  $\delta$ , os poderes dos testes tem comportamento muito parecido, salvo alguns casos como  $\delta = -2$ ,  $\delta = -1$  e  $\delta = 3$ , em que o poder do teste escore corrigido é ligeiramente menor.

### 3.8 Resultados numéricos: modelo de regressão normal inverso

Consideramos agora o caso do modelo de regressão normal inverso. Aqui, cada  $y_i$  tem densidade (3.1) com  $a(y, \phi) = \log[(\phi/2\pi y^3)^{1/2}]$ ,  $c(y) = -1/(2y)$ ,  $b(\theta) = -(-2\theta)^{1/2}$ ,  $\mu = E(y) = (-2\theta)^{1/2}$  e  $V(\mu) = \mu^3$ . Assumimos a estrutura linear (3.31), ou seja, ligação logarítmica. Assim  $d\mu/d\eta = \mu$ ,  $g = -2/\mu$ ,  $t = w = f = \mathcal{M} = 1/\mu$ ,  $h = 9/\mu$ ,  $b = 15/\mu$ ,  $d = 3/\mu$  e  $e = 7/\mu$  definem os momentos do logaritmo da função de verossimilhança. Assim como no modelo normal,  $d_{(2)} = -1/2$  e  $d_{(3)} = 1$ .

A hipótese nula  $\mathcal{H}_0 : \beta_1 = 0$ , em que  $\beta_1$  é um vetor  $q$ -dimensional, é considerada aqui. Os outros

Tabela 3.1: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 4$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 3$								
20	10	24,13	15,78	9,23	10,18	10,73	9,93	9,73
	5	16,73	8,82	4,10	4,37	5,54	4,97	4,77
	1	7,59	2,43	0,55	0,37	1,13	0,99	0,73
25	10	21,17	14,23	9,32	10,26	10,45	9,91	9,91
	5	13,92	7,82	4,16	4,39	5,19	4,96	4,69
	1	5,53	1,81	0,63	0,46	0,99	0,90	0,76
30	10	19,81	13,97	10,24	10,55	10,7	10,64	10,28
	5	12,68	7,62	4,79	5,08	5,48	5,37	5,25
	1	4,75	1,87	0,84	0,67	1,17	1,10	0,97
$q = 2$								
20	10	21,27	15,25	9,41	11,71	10,68	9,96	10,07
	5	14,03	8,71	4,03	5,61	5,61	4,99	5,01
	1	6,15	2,36	0,39	0,73	1,17	0,93	0,91
25	10	19,05	14,3	10,08	11,69	10,59	10,27	10,07
	5	12,11	7,89	4,69	5,59	5,41	5,23	5,07
	1	4,51	1,91	0,63	0,63	0,95	0,90	0,76
30	10	17,52	13,56	10,10	11,46	10,72	10,46	10,31
	5	10,7	7,34	4,45	5,30	5,15	4,96	4,94
	1	3,79	1,83	0,65	0,76	1,05	0,99	0,90
$q = 1$								
20	10	18,04	14,31	10,62	13,02	10,97	10,29	10,39
	5	11,55	8,24	5,15	6,68	5,68	5,19	5,23
	1	4,47	2,28	0,76	1,15	1,17	0,94	0,96
25	10	16,52	13,58	10,30	12,38	10,67	10,49	10,39
	5	10,16	7,45	4,81	6,35	5,47	5,30	5,20
	1	3,67	1,91	0,71	1,19	1,19	1,05	1,11
30	10	15,39	12,39	9,73	11,61	10,15	10,01	9,89
	5	9,22	6,64	4,69	5,97	5,19	5,09	5,04
	1	2,95	1,45	0,74	0,99	0,99	0,97	0,95

Tabela 3.2: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), score ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 5$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 4$								
20	10	31,01	17,2	8,81	9,31	10,61	9,79	9,12
	5	22,41	9,67	3,79	3,37	5,21	4,59	3,95
	1	11,01	2,43	0,42	0,18	0,95	0,77	0,47
25	10	27,38	15,55	9,71	10,34	10,93	10,19	10,16
	5	19,69	9,14	4,66	4,51	5,65	5,01	4,90
	1	9,12	2,54	0,79	0,49	1,23	0,88	0,83
30	10	23,69	14,46	9,78	9,73	10,29	9,78	9,65
	5	15,65	7,81	4,74	4,03	5,03	4,74	4,43
	1	6,51	1,69	0,85	0,52	0,99	0,85	0,79
$q = 3$								
20	10	28,81	17,83	10,53	12,4	11,77	10,76	10,53
	5	21,28	10,76	4,81	5,63	6,17	5,27	5,13
	1	10,6	2,94	0,61	0,54	1,25	0,81	0,78
25	10	24,45	15,87	9,89	11,45	10,97	10,29	10,26
	5	16,63	9,11	4,46	5,37	5,71	4,98	5,03
	1	7,31	2,41	0,81	0,73	1,25	1,01	1,00
30	10	22,02	14,42	9,53	11,19	10,57	9,88	10,09
	5	14,37	8,16	4,57	5,21	5,37	5,04	4,95
	1	5,68	1,98	0,75	0,73	1,10	1,07	0,89
$q = 2$								
20	10	25,07	17,21	10,56	13,85	11,96	11,05	11,06
	5	17,74	10,71	4,87	6,97	6,31	5,75	5,55
	1	8,44	3,26	0,64	0,97	1,31	0,99	0,94
25	10	21,00	15,85	10,40	13,07	11,07	10,53	10,35
	5	14,07	8,92	4,66	6,17	5,61	5,07	5,09
	1	5,68	2,30	0,61	0,94	1,09	0,87	0,93
30	10	19,51	14,25	10,15	12,15	10,66	10,13	10,22
	5	12,29	7,79	4,71	5,85	5,22	4,93	4,85
	1	4,58	1,92	0,75	0,95	1,08	0,89	0,89

Tabela 3.3: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), score ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 6$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 4$								
20	10	35,69	20,23	9,77	11,57	12,15	10,20	10,03
	5	27,14	12,40	4,45	4,99	6,61	4,97	4,82
	1	15,31	3,99	0,64	0,37	1,47	0,76	0,69
25	10	28,99	17,23	8,96	11,04	11,01	10,09	9,82
	5	20,45	9,98	3,95	4,57	5,59	4,81	4,51
	1	9,83	2,77	0,55	0,45	1,15	0,91	0,71
30	10	25,41	15,97	8,97	10,65	10,65	9,93	9,73
	5	17,52	8,93	3,83	4,65	5,29	4,81	4,69
	1	8,05	2,39	0,54	0,59	1,13	0,25	0,82
$q = 3$								
20	10	31,15	19,87	10,13	13,57	11,95	10,33	10,14
	5	22,71	12,17	4,60	6,18	6,17	5,47	4,92
	1	11,53	3,79	0,51	0,61	1,23	0,89	0,70
25	10	27,51	17,65	9,75	12,51	10,95	10,03	9,95
	5	19,09	9,82	4,27	5,75	5,63	4,95	4,77
	1	8,61	2,66	0,59	0,76	1,16	0,83	0,81
30	10	22,89	15,43	9,43	11,87	10,34	9,55	9,65
	5	15,51	8,56	4,15	5,69	5,32	4,62	4,70
	1	6,45	2,15	0,51	0,67	0,90	0,74	0,69
$q = 2$								
20	10	26,77	18,80	11,20	15,24	11,8	10,58	10,27
	5	18,88	11,39	5,11	7,57	6,17	5,45	4,88
	1	8,72	3,43	0,62	1,1	1,32	1,01	0,85
25	10	24,36	16,92	10,56	14,03	11,28	10,29	10,4
	5	16,45	9,89	4,83	7,03	5,67	5,02	4,97
	1	7,07	2,70	0,71	1,04	1,05	1,03	0,84
30	10	20,61	15,05	10,36	12,99	10,71	10,08	10,01
	5	13,5	8,51	4,7	6,37	5,48	4,86	5,04
	1	5,28	2,29	0,61	0,97	1,00	0,81	0,82

Tabela 3.4: Taxas de rejeição não-nulas (poder) das versões corrigidas dos testes LR ( $\bar{S}_{LR}^*$ ), escore ( $\bar{S}_R^*$ ) e gradiente ( $\bar{S}_T^*$ ) no modelo de regressão gama com  $\phi$  desconhecido;  $\alpha = 5\%$ ,  $p = 4$ ,  $q = 2$ ,  $n = 30$ .

$\delta$	-4,00	-3,00	-2,00	-1,00	-0,50	0,50	1,00	2,00	3,00	4,00
$\bar{S}_{LR}^*$	99,99	99,84	96,99	44,38	14,03	14,57	37,01	94,75	99,96	100,00
$\bar{S}_R^*$	99,63	99,31	92,28	40,17	12,14	14,09	37,29	90,44	99,36	99,99
$\bar{S}_T^*$	99,99	99,83	96,73	43,02	13,13	14,05	36,73	94,64	99,97	100,00

parâmetros da regressão são iguais a 1 e  $\phi = 3$ . Com base em 15,000 réplicas, avaliamos a taxa de rejeição de  $\mathcal{H}_0$  para os níveis  $\alpha = 10\%$ ,  $5\%$  e  $1\%$ , dos testes de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ), e as versões corrigidas  $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$  e  $\bar{S}_T^*$ . Aqui, a primeira coluna da matriz de covariadas é um vetor de 1's e as demais covariadas foram geradas independentemente com distribuição normal padrão  $N(0, 1)$ . Na Tabela 3.5 são exibidos os resultados para  $p = 4$  e  $q = 3, 2, 1$ ; Tabela 3.6 com  $p = 5$  e  $q = 4, 3, 2$ ; e Tabela 3.7 para  $p = 6$  e  $q = 4, 3, 2$ . Variamos o tamanho de amostra em  $n = 20, 25, 30$ .

Os testes de Wald e LR mostraram-se muito liberais, com taxas de rejeição da hipótese nula muito distorcidas em todos os casos como, por exemplo, no caso  $p = 5$ ,  $q = 2$ ,  $n = 25$  e  $\alpha = 5\%$  (Tabela 3.6) em que o teste de Wald apresentou uma taxa de rejeição de 23, 51% e o teste LR de 11, 21%. Estas taxas apresentam-se ainda mais distorcidas quando aumenta-se a quantidade de parâmetros testados no modelo. As taxas de rejeição do teste LR são bastante atenuados pela correção de Bartlett, a saber, nesse mesmo cenário, a taxa de rejeição é de 6, 53%, no entanto ainda com desempenho inferior aos testes escore e gradiente não-corrigidos, que apresentaram taxas de 4, 24% e 5, 22%, respectivamente.

Na Tabela 3.5, observa-se que o comportamento dos testes escore e gradiente corrigida foram semelhantes quando  $q = 2$  e  $q = 3$ . Por exemplo, para  $q = 2$  e  $n = 30$ , as taxas de rejeição nula dos testes que utilizam  $\bar{S}_R^*$  e  $\bar{S}_T^*$  foram, respectivamente, de 10, 11% e 10, 15% ao nível de 10%, iguais a 4, 97% ao nível 5%, e de 0, 87% e 0, 88% ao nível de 1%.

Na Tabela 3.6, verificamos que as taxas de rejeição nula do teste gradiente foram atenuadas pela correção tipo-Bartlett, principalmente no caso  $q = 2$ . Por exemplo, para  $n = 20$  e  $\alpha = 10\%$ , o teste



Tabela 3.5: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 4$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 3$								
20	10	41,17	20,03	8,85	7,68	12,95	10,44	10,01
	5	32,72	11,96	4,64	2,81	6,84	2,85	4,99
	1	20,69	3,55	1,19	0,26	1,55	0,02	0,99
25	10	37,37	17,61	8,67	7,83	11,85	10,11	9,50
	5	29,59	10,17	4,44	2,83	5,95	4,10	4,44
	1	18,4	2,82	0,97	0,23	1,35	0,15	0,77
30	10	34,66	15,86	8,76	7,98	10,95	9,86	9,05
	5	26,33	8,69	4,40	2,91	5,43	4,45	4,21
	1	14,88	2,26	0,85	0,17	1,17	0,41	0,65
$q = 2$								
20	10	33,67	18,76	7,81	9,15	12,51	10,93	9,90
	5	25,66	11,19	3,24	3,27	6,53	5,29	4,71
	1	15,01	3,29	0,39	0,21	1,39	0,93	0,73
25	10	27,04	16,03	9,62	10,51	11,32	10,22	9,95
	5	19,35	9,00	4,47	4,55	5,86	4,74	4,93
	1	9,99	2,42	0,9	0,31	1,21	0,79	0,71
30	10	23,68	15,14	8,96	11,03	11,23	10,52	10,46
	5	16,35	8,77	4,09	4,76	5,71	5,09	4,92
	1	7,35	2,20	0,63	0,49	1,30	0,99	0,85
$q = 1$								
20	10	18,39	14,99	11,46	13,19	10,83	10,12	9,94
	5	11,95	8,48	5,67	6,46	5,49	5,09	4,97
	1	4,89	2,49	1,06	1,11	1,26	1,07	0,97
25	10	20,06	14,83	9,50	12,41	11,57	10,73	10,63
	5	13,27	8,46	4,34	5,86	5,68	5,53	5,19
	1	5,31	2,17	0,57	0,77	1,19	1,19	0,96
30	10	15,41	13,19	10,87	11,87	10,32	10,12	9,99
	5	9,48	7,39	5,19	5,83	5,34	4,95	5,14
	1	3,17	1,69	0,75	0,85	0,98	0,81	0,84

Tabela 3.6: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 5$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 4$								
20	10	53,41	23,62	8,14	6,26	13,26	10,74	9,52
	5	44,83	14,67	3,89	2,01	6,74	4,43	4,68
	1	30,77	4,49	0,77	0,21	1,49	0,23	0,96
25	10	49,45	21,07	9,13	6,39	12,55	9,95	8,91
	5	41,37	12,25	4,72	2,15	6,45	2,98	4,05
	1	27,53	3,25	1,25	0,15	1,15	0,07	0,69
30	10	47,20	19,55	9,50	7,47	12,47	9,83	9,09
	5	38,33	11,47	5,18	2,65	6,75	3,67	4,35
	1	25,27	3,43	1,34	0,18	1,37	0,10	0,59
$q = 3$								
20	10	45,93	23,64	7,49	10,20	14,21	11,42	11,44
	5	37,74	15,14	3,65	4,26	7,61	5,66	5,89
	1	23,93	5,01	0,66	0,65	1,73	0,66	1,57
25	10	38,11	19,81	8,67	10,12	12,31	10,83	10,00
	5	29,69	11,9	4,06	3,83	6,51	5,42	4,65
	1	17,35	3,58	0,61	0,33	1,37	0,89	0,85
30	10	31,79	17,04	9,00	9,67	11,33	10,00	9,59
	5	23,61	9,73	4,63	3,96	5,89	4,91	4,59
	1	12,76	2,49	1,06	0,38	1,14	0,71	0,75
$q = 2$								
20	10	36,89	21,54	9,80	11,4	13,23	11,54	9,98
	5	28,93	13,21	5,13	5,36	7,27	4,89	5,36
	1	17,35	4,39	1,25	0,83	1,66	0,14	1,27
25	10	31,27	18,71	9,29	12,15	12,43	11,03	10,69
	5	23,51	11,21	4,24	5,22	6,53	5,33	5,28
	1	13,45	3,36	0,75	0,56	1,54	0,86	0,95
30	10	20,51	14,95	9,59	11,63	10,81	10,11	10,15
	5	13,57	8,67	4,35	5,37	5,49	4,97	4,97
	1	5,35	2,21	0,52	0,66	1,16	0,87	0,88

Tabela 3.7: Taxas de rejeição de  $\mathcal{H}_0$  no modelo de regressão normal inversa com dispersão desconhecida para as estatísticas de Wald ( $\bar{S}_W$ ), LR ( $\bar{S}_{LR}$ ), escore ( $\bar{S}_R$ ), gradiente ( $\bar{S}_T$ ) e versões corrigidas ( $\bar{S}_{LR}^*$ ,  $\bar{S}_R^*$ ,  $\bar{S}_T^*$ );  $p = 6$ .

$n$	$\alpha$	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
$q = 4$								
20	10	51,10	26,65	7,64	8,69	14,21	10,81	10,6
	5	42,37	16,91	3,00	2,79	7,45	5,13	4,96
	1	27,45	5,77	0,39	0,25	1,70	0,94	0,97
25	10	45,01	22,11	10,12	9,67	12,75	10,64	10,07
	5	36,47	13,41	5,49	4,13	6,51	4,68	5,31
	1	23,18	4,02	1,37	0,40	1,55	0,19	1,04
30	10	49,01	22,01	8,86	8,45	12,75	10,63	9,58
	5	39,98	13,09	4,32	3,10	6,46	4,90	4,51
	1	26,23	3,79	0,83	0,23	1,44	0,49	0,80
$q = 3$								
20	10	43,03	24,72	11,5	12,83	14,21	11,52	11,00
	5	34,61	15,87	5,85	5,47	7,92	5,69	5,41
	1	21,27	5,65	1,29	0,51	1,86	0,63	1,03
25	10	35,32	21,32	10,25	12,55	12,91	10,83	10,84
	5	26,8	12,97	4,87	5,40	6,86	5,39	5,27
	1	14,36	3,95	0,94	0,71	1,59	1,09	1,09
30	10	32,05	18,35	9,99	10,67	11,35	10,39	9,78
	5	23,43	10,53	4,45	4,30	5,86	4,87	4,67
	1	12,32	2,95	0,72	0,33	1,16	0,87	0,71
$q = 2$								
20	10	32,61	21,55	12,21	14,85	13,33	11,1	10,59
	5	24,12	13,76	6,02	7,44	7,41	5,47	5,18
	1	13,15	4,70	0,93	0,85	1,71	0,95	0,83
25	10	29,61	19,30	11,42	14,28	12,42	10,51	10,54
	5	21,49	11,99	5,47	7,09	6,78	5,29	5,58
	1	10,61	3,76	0,91	0,91	1,50	1,03	0,94
30	10	28,01	17,51	9,33	11,83	11,29	10,27	10,07
	5	20,23	10,08	4,39	5,20	5,59	5,02	4,83
	1	10,01	2,64	0,90	0,75	1,11	0,98	1,05

gradiente rejeitou a hipótese nula a uma taxa de 11,40%, taxa que foi para 9,98% quando a correção foi aplicada no teste. Quando aumentamos o número de parâmetros testados para  $q = 4$ , as taxas de rejeição do teste gradiente ficaram bem abaixo dos níveis nominais especificados. A correção tipo-Bartlett aproximou estas taxas dos níveis nominais especificados, como no caso  $n = 20$ , em que, para  $\alpha = 10\%$ , as taxas de rejeição empíricas dos testes  $\bar{S}_T$  e  $\bar{S}_T^*$  foram de 6,26% e 9,52%, respectivamente.

Na Tabela 3.7 fica evidente a semelhança das taxas de rejeição nula dos testes escore e gradiente corrigidos, principalmente quando fixamos os níveis  $\alpha = 10\%$  e  $5\%$ . Por exemplo, para  $q = 2$  e  $n = 30$ , as taxas de rejeição empíricas dos testes que usam  $\bar{S}_R^*$  e  $\bar{S}_T^*$  foram, respectivamente, 10,08% e 10,01% para  $\alpha = 10\%$ . Quando aumentamos o número  $q$  de parâmetros testados, as estatísticas escore e gradiente tendem a diminuir a taxa de rejeição nula, fato contrário aos testes de Wald e LR. Para  $q = 4$  e  $n = 20$  por exemplo, as taxa de rejeição de  $\mathcal{H}_0$  ( $\alpha = 10\%$ ) para  $\bar{S}_R$  e  $\bar{S}_T$  foram 9,77% e 11,57%, respectivamente, enquanto a correção desses testes resultou em taxas de 10,20% e 10,03%, respectivamente.

No contexto geral, a correção tipo-Bartlett para a estatística gradiente foi efetiva, com comportamento semelhante à estatística escore corrigida e com menor distorção de tamanho do que o teste baseado na estatística da razão de verossimilhanças corrigida por um fator de correção de Bartlett.

A Tabela 3.8 mostra o poder dos testes baseados nas estatísticas corrigidas. Fixamos  $p = 6$ ,  $q = 2$ ,  $n = 30$  e  $\alpha = 5\%$ . Aqui, analisamos o desempenho dos testes sob  $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$  para diferentes valores de  $\delta$ . Em geral, o poder dos testes aumenta à medida que  $|\delta|$  cresce. Somente quando  $\delta = 0,5$ , há uma queda de poder do teste gradiente corrigido em relação a  $\delta = 0,4$ . Quando  $\delta = -0,5$  o poder do teste gradiente é superior ao poder do teste escore corrigido. De modo geral, os testes apresentam taxas de rejeição não-nulas próximas, fixados os  $\delta$ 's.

Tabela 3.8: Taxas de rejeição não-nulas (poder) das versões corrigidas dos testes LR ( $\bar{S}_{LR}^*$ ), escore ( $\bar{S}_R^*$ ) e gradiente ( $\bar{S}_T^*$ ) no modelo de regressão normal inverso com  $\phi$  desconhecido;  $\alpha = 5\%$ ,  $p = 6$ ,  $q = 2$ ,  $n = 30$ .

$\delta$	-0,50	-0,40	-0,30	-0,20	-0,10	0,10	0,20	0,30	0,40	0,50
$\bar{S}_{LR}^*$	99,08	95,46	80,67	47,63	15,96	19,57	58,06	88,23	97,95	99,67
$\bar{S}_R^*$	94,54	92,17	79,29	49,35	17,65	22,74	62,48	87,91	95,03	96,88
$\bar{S}_T^*$	98,94	95,79	81,64	48,94	15,90	20,55	61,19	90,36	97,91	94,82

### 3.9 Aplicações

No que se segue, apresentamos aplicações a dados reais dos métodos desenvolvidos nas seções anteriores. O primeiro experimento considera o conjunto de dados em Freund (1983), sobre um estudo do tamanho de lulas predadas por tubarões e atuns –“*squid data*”. A variável resposta considerada ( $y$ ) corresponde ao peso em libras de  $n = 22$  espécimes de lulas capturadas. As variáveis regressoras correspondem a características do bico ou boca das lulas, em polegadas, a saber, comprimento do bico ( $x_1$ ), comprimento dos tentáculos ( $x_2$ ), comprimento do entalhe do bico ( $x_3$ ), comprimento do entalhe dos tentáculos ( $x_4$ ) e largura da lula ( $x_5$ ). A Tabela 3.14 da Seção 3.11.3 apresenta os dados. O modelo proposto é dado por

$$\log(\mu_l) = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \beta_4 x_{4l} + \beta_5 x_{5l}, \quad (3.33)$$

$l = 1, \dots, 22$ . Aqui, a variável resposta  $y_l$  tem distribuição gama  $G(\mu_l, \phi)$ , com dispersão  $\phi^{-1}$  desconhecida. As estimativas pontuais dos parâmetros de regressão são dadas na Tabela 3.9;  $\hat{\phi} = 44,0006$ .

Tabela 3.9: Estimativas dos parâmetros do modelo (3.33).

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Estimativa	-2,2899	0,4027	-0,4362	1,2916	1,9420	2,1394
Erro padrão	0,2001	0,5515	0,5944	1,3603	0,7844	1,0407

O teste da hipótese  $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$ , ou seja, dos termos que representam os efeitos dos entalhes do bico e dos tentáculos das lulas, conduziu aos resultados da Tabela 3.10. Notamos que ao nível

Tabela 3.10: Valores das estatísticas e  $p$ -valores para o teste da hipótese  $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$  no modelo (3.33).

	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
Estatística	7,0659	5,8976	4,8382	5,1193	4,6380	4,0842	4,3239
$p$ -valor	0,0292	0,0524	0,0890	0,0773	0,0984	0,1298	0,1151

de 10% de significância, rejeitamos  $\mathcal{H}_0$  quando a inferência é baseada nos testes não-corrigidos ou no teste LR corrigido. No entanto, a decisão se altera quando utilizamos os testes escore e gradiente corrigidos. Para o teste de  $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$ , que corresponde a testar os efeitos dos comprimentos do bico e dos tentáculos das espécimes, nenhum dos testes (inclusive os testes corrigidos) rejeita a hipótese nula. Testando o efeito da largura dos indivíduos,  $\mathcal{H}_0 : \beta_5 = 0$ , notamos que os testes corrigidos indicam menor evidência contra  $\mathcal{H}_0$  (ver Tabela 3.11). Vale ressaltar que o tamanho da amostra ( $n = 22$ ) é pequeno e, por resultados prévios, os testes de Wald e os testes não corrigidos LR, escore e gradiente tendem a apresentar resultados distorcidos, assim, devemos ser mais cautelosos quanto a aceitar as suas conclusões.

Tabela 3.11: Valores das estatísticas e  $p$ -valores para o teste da hipótese  $\mathcal{H}_0 : \beta_5 = 0$  no modelo (3.33).

	$\bar{S}_W$	$\bar{S}_{LR}$	$\bar{S}_R$	$\bar{S}_T$	$\bar{S}_{LR}^*$	$\bar{S}_R^*$	$\bar{S}_T^*$
Estatística	4,2263	3,8505	3,5444	3,5486	2,9754	2,8104	2,7919
$p$ -valor	0,0398	0,0497	0,0597	0,0596	0,0845	0,0937	0,0947

Tomamos agora um segundo experimento envolvendo dados reais. Consideramos um experimento no qual o pesquisador deseja determinar a relação empírica entre a resistência de embalagens ( $y$ ) em g/cm de um estoque de embalagens de pão – “*breadwrapper data*”. Os dados com  $n = 20$  observações estão em Myers (1976) e são dados na Tabela 3.15. As variáveis explicativas são temperatura de empacotamento ( $x_1$ ), temperatura de resfriamento ( $x_2$ ) e porcentagem de polietileno ( $x_3$ ).

O modelo proposto é da forma

$$\log(\mu_l) = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l}, \quad (3.34)$$

com  $l = 1, \dots, 20$ , em que supõe-se cada resposta  $y_l$  com distribuição gamma  $G(\mu_l, \phi)$ , em que  $\phi$  é um parâmetro desconhecido. As estimativas para o modelo são dadas na Tabela 3.12;  $\hat{\phi} = 23,3168$ . Considerando o teste de  $\mathcal{H}_0 : \beta_3 = 0$ , obtivemos as estatísticas e  $p$ -valores dados na Tabela 3.13.

Tabela 3.12: Estimativas do modelo (3.34).

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Estimativa	3,0085	-0,0048	0,0007	0,2352
Erro padrão	0,5974	0,0019	0,0062	0,0934

Tabela 3.13: Valores das estatísticas e  $p$ -valores para o teste da hipótese  $\mathcal{H}_0 : \beta_3 = 0$  no modelo (3.34).

	$\tilde{S}_W$	$\tilde{S}_{LR}$	$\tilde{S}_R$	$\tilde{S}_T$	$\tilde{S}_{LR}^*$	$\tilde{S}_R^*$	$\tilde{S}_T^*$
Estatística	6,3477	4,8229	3,6908	4,2987	3,9449	3,2288	3,7976
$p$ -valor	0,0118	0,0281	0,0547	0,0381	0,0470	0,0724	0,0513

Conclui-se neste caso, que, a um nível de significância 5%, o teste LR e sua versão corrigida rejeitam a hipótese nula. Por outro lado, o teste escore, que já não rejeita a hipótese nula, mantém a mesma conclusão após a correção tipo-Bartlett. O teste gradiente rejeita  $\mathcal{H}_0$  e, após a correção (tipo-Bartlett), altera a sua conclusão, deixando em dúvida a aceitação ou rejeição da hipótese nula já que o  $p$ -valor correspondente é ligeiramente maior que 5%. Observe que o teste LR corrigido em situações anteriores apresentou resultados mais distorcidos do que os testes gradiente e escore corrigidos, principalmente se o tamanho da amostra é pequeno, que é o caso tratado aqui.

### 3.10 Discussão

Cordeiro (1983, 1987) derivou fórmulas matriciais para o fator de correção de Bartlett do teste de razão de verossimilhanças em MLG. Posteriormente, Cordeiro, Ferrari e Paula (1993) e Cribari-Neto

e Ferrari (1995) apresentaram fórmulas matriciais para o fator de correção tipo-Bartlett da estatística escore em MLG baseando-se no trabalho de Cordeiro e Ferrari (1991). Neste capítulo, determinamos fórmulas matriciais para o fator de correção tipo-Bartlett da estatística gradiente em MLG com dispersão conhecida e desconhecida com base nos resultados do Capítulo 2. Estas fórmulas fechadas são de simples computação. Os estudos de simulação indicaram que, em pequenas amostras, a distribuição  $\chi^2$  pode não ser uma boa aproximação para a distribuição assintótica das estatísticas da razão de verossimilhanças, escore, de Wald e gradiente em MLG com dispersão desconhecida. O teste de Wald mostrou-se extremamente liberal, e não deve ser recomendado. Ainda, se as correções não são utilizadas, os testes escore e gradiente são mais recomendados que os testes da razão de verossimilhanças e Wald, pois apresentam menores distorções de tamanho. Os testes gradiente e escore corrigidos apresentam menor distorção de tamanho que o teste da razão de verossimilhanças corrigido, e mostraram comportamento semelhante em amostras pequenas.

## 3.11 Detalhes técnicos

### 3.11.1 Obtenção das quantidades $A_1$ , $A_2$ e $A_3$

Discutimos aqui o cálculo das expressões matriciais para  $A_1$ ,  $A_2$  e  $A_3$ , que definem a expansão para a função de distribuição da estatística gradiente sob a hipótese nula em MLG com dispersão conhecida. Tais quantidades podem ser obtidas substituindo os momentos do logaritmo da função de verossimilhança dados na Seção 3.4 nas fórmulas dos  $A$ 's dadas no Teorema 2.1, efetuando somas sobre o tamanho da amostra e posteriormente avaliando a somas sobre os parâmetros. Ao fazer isso, aparecerão termos da forma  $-\sum'(i)_l \kappa^{ij}(j)_l$ ,  $\sum'(i)_l a^{ij}(j)_l$  e  $\sum'(i)_l m^{ij}(j)_l$ , onde  $-\kappa^{ij}$ ,  $a^{ij}$  e  $m^{ij}$  são elementos da posição  $(i, j)$  das matrizes  $\mathbf{K}^{-1}$ ,  $\mathbf{A}$  e  $\mathbf{M}$  respectivamente. Esses somatórios representam elementos de  $\phi^{-1}\mathbf{Z}$ ,  $\phi^{-1}\mathbf{Z}_2$  e  $\phi^{-1}(\mathbf{Z} - \mathbf{Z}_2)$ , respectivamente. Mostramos aqui, com detalhes, a obtenção de  $A_3$  e com menos detalhes a obtenção de  $A_1$  e  $A_2$ . Do Teorema 2.1, a quantidade  $A_3$  pode ser escrita da forma

$$A_3 = \frac{1}{4} \sum' \kappa_{jrs} \kappa_{uvw} (3m^{jr} m^{su} m^{vw} + 2m^{ju} m^{rv} m^{sw}), \quad (3.35)$$



em que  $\sum'$  representa a soma sobre todos os índices  $1 \leq j, r, s, u, v, w \leq p$ . Fazendo a substituição dos momentos dados na Seção 3.4, tem-se

$$A_3 = \frac{3\phi^2}{4} \sum' \sum (f+2g)_c (jrs)_c \sum (f+2g)_m (uvw)_m m^{jr} m^{su} m^{vw} \\ + \frac{\phi^2}{2} \sum' \sum (f+2g)_c (jrs)_c \sum (f+2g)_m (uvw)_m m^{ju} m^{rv} m^{sw},$$

em que os índices  $c$  e  $m$  variam de 1 a  $n$ . O ponto principal para a obtenção das fórmulas matriciais para os  $A$ 's é a troca dos somatórios  $\sum'$  e  $\sum$ . Fazendo isto, e rearranjando os termos, temos

$$A_3 = \frac{3\phi^2}{4} \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr} (r)_c \right) \left( \sum' (s)_c m^{su} (u)_m \right) \left( \sum' (v)_m m^{vw} (w)_m \right) \\ + \frac{\phi^2}{2} \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{ju} (u)_m \right) \left( \sum' (r)_c m^{rv} (v)_m \right) \left( \sum' (s)_c m^{sw} (w)_m \right),$$

o que nos leva a (3.13).

A partir do Teorema 2.1, o termo  $A_2$  é escrito como

$$A_2 = A_{21} + A_{22} + A_{23},$$

em que

$$A_{21} = -3 \sum' \kappa_{jrs} \kappa_{klu} \left[ m^{jr} m^{sk} a^{lu} + m^{jr} a^{sk} m^{lu} + 2m^{jk} m^{rl} a^{su} \right. \\ \left. + \frac{1}{4} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su}) \right],$$

$$A_{22} = 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} \left[ m^{su} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (\kappa^{sk} \kappa^{lu} - a^{sk} a^{lu}) \right],$$

$$A_{23} = 6 \sum' \kappa_{jrs}^{(u)} m^{jr} m^{su} - 3 \sum' \kappa_{jrsu} m^{jr} m^{su}.$$

Utilizando procedimento análogo feito para a quantidade  $A_3$ , temos

$$A_{21} = -3\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr} (r)_c \right) \left( \sum' (s)_c m^{sk} (k)_m \right) \left( \sum' (l)_m a^{lu} (u)_m \right) \\ - 3\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr} (r)_c \right) \left( \sum' (s)_c a^{sk} (k)_m \right) \left( \sum' (l)_m m^{lu} (u)_m \right) \\ - 6\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jk} (k)_m \right) \left( \sum' (r)_c m^{rl} (l)_m \right) \left( \sum' (s)_c a^{su} (u)_m \right) \\ - \frac{9}{4} \phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr} (r)_c \right) \left( \sum' (s)_c m^{sk} (k)_m \right) \left( \sum' (l)_m m^{lu} (u)_m \right) \\ - \frac{3}{2} \phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jk} (k)_m \right) \left( \sum' (r)_c m^{rl} (l)_m \right) \left( \sum' (s)_c m^{su} (u)_m \right),$$

$$\begin{aligned}
A_{22} = & 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c m^{su}(u)_m \right) \left( \sum' (j)_c \kappa^{jk}(k)_m \right) \left( \sum' (l)_m \kappa^{lr}(r)_c \right) \\
& - 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c m^{su}(u)_m \right) \left( \sum' (j)_c a^{jk}(k)_m \right) \left( \sum' (l)_m a^{lr}(r)_c \right) \\
& + 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (j)_c m^{jr}(r)_c \right) \left( \sum' (s)_c \kappa^{sk}(k)_m \right) \left( \sum' (l)_m \kappa^{lu}(u)_m \right) \\
& - 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (j)_c m^{jr}(r)_c \right) \left( \sum' (s)_c a^{sk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right),
\end{aligned}$$

$$\begin{aligned}
A_{23} = & -6\phi \sum t_m \left( \sum' (j)_m m^{jr}(r)_m \right) \left( \sum' (s)_m m^{su}(u)_m \right) \\
& + 3\phi \sum e_m \left( \sum' (j)_m m^{jr}(r)_m \right) \left( \sum' (s)_m m^{su}(u)_m \right).
\end{aligned}$$

Escrevendo as quantidades acima em forma matricial, e efetuando a soma das mesmas, obtemos (3.12).

O termo  $A_1$  da expansão é dado no Teorema 2.1 por

$$A_1 = A_{11} + A_{12} + A_{13} + A_{14},$$

em que

$$A_{11} = 3 \sum' \kappa_{jrs} \kappa_{klu} \left[ m^{jr} a^{lu} (m^{sk} + 2a^{sk}) + a^{jr} m^{sk} a^{lu} + 2m^{jk} a^{rl} a^{su} \right],$$

$$A_{12} = -12 \sum' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} (\kappa^{sj} \kappa^{rk} \kappa^{lu} + a^{sj} a^{rk} a^{lu} + \kappa^{sk} \kappa^{lj} \kappa^{ru} + a^{sk} a^{lj} a^{ru}),$$

$$\begin{aligned}
A_{13} = & -6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [(a^{su} - \kappa^{su}) (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (a^{sk} a^{lu} + \kappa^{sk} \kappa^{lu}) \\
& + 2a^{rs} (\kappa^{jk} \kappa^{lu} - a^{jk} a^{lu}) + 2a^{rk} a^{ls} m^{ju}],
\end{aligned}$$

$$\begin{aligned}
A_{14} = & 6 \sum' \kappa_{jr su} m^{jr} a^{su} - 6 \sum' \kappa_{jrs}^{(u)} [m^{jr} (a^{su} - \kappa^{su}) + 2m^{ju} a^{rs}] \\
& + 12 \sum' \kappa_{rs}^{(ju)} (\kappa^{jr} \kappa^{su} - a^{jr} a^{su}).
\end{aligned}$$

Assim,

$$\begin{aligned}
A_{11} = & 3\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr}(r)_c \right) \left( \sum' (s)_c m^{sk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right) \\
& + 6\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jr}(r)_c \right) \left( \sum' (s)_c a^{sk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right) \\
& + 3\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (l)_m m^{lr}(r)_c \right) \left( \sum' (s)_c m^{sk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right) \\
& + 6\phi^2 \sum (f+2g)_c (f+2g)_m \left( \sum' (j)_c m^{jk}(k)_m \right) \left( \sum' (r)_c a^{rl}(l)_m \right) \left( \sum' (s)_c a^{su}(u)_m \right),
\end{aligned}$$

$$\begin{aligned}
A_{12} = & -12\phi^2 \sum (f+g)_c (f+g)_m \left( \sum' (s)_c \kappa^{sj}(j)_c \right) \left( \sum' (r)_c \kappa^{rk}(k)_m \right) \left( \sum' (l)_m \kappa^{lu}(u)_m \right) \\
& - 12\phi^2 \sum (f+g)_c (f+g)_m \left( \sum' (s)_c a^{sj}(j)_c \right) \left( \sum' (r)_c a^{rk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right) \\
& - 12\phi^2 \sum (f+g)_c (f+g)_m \left( \sum' (s)_c \kappa^{sk}(k)_m \right) \left( \sum' (j)_c \kappa^{jl}(l)_m \right) \left( \sum' (u)_m \kappa^{ur}(r)_c \right) \\
& - 12\phi^2 \sum (f+g)_c (f+g)_m \left( \sum' (s)_c a^{sk}(k)_m \right) \left( \sum' (j)_c a^{jl}(l)_m \right) \left( \sum' (u)_m a^{ur}(r)_c \right),
\end{aligned}$$

$$\begin{aligned}
A_{13} = & -6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c a^{su}(u)_m \right) \left( \sum' (j)_c \kappa^{jk}(k)_m \right) \left( \sum' (l)_m \kappa^{lr}(r)_c \right) \\
& + 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c a^{su}(u)_m \right) \left( \sum' (j)_c a^{jk}(k)_m \right) \left( \sum' (l)_m a^{lr}(r)_c \right) \\
& + 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c \kappa^{su}(u)_m \right) \left( \sum' (j)_c \kappa^{jk}(k)_m \right) \left( \sum' (l)_m \kappa^{lr}(r)_c \right) \\
& - 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c \kappa^{su}(u)_m \right) \left( \sum' (j)_c a^{jk}(k)_m \right) \left( \sum' (l)_m a^{lr}(r)_c \right) \\
& - 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (r)_c m^{rj}(j)_c \right) \left( \sum' (s)_c a^{sk}(k)_m \right) \left( \sum' (l)_m a^{lu}(u)_m \right) \\
& - 6\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (r)_c m^{rj}(j)_c \right) \left( \sum' (s)_c \kappa^{sk}(k)_m \right) \left( \sum' (l)_m \kappa^{lu}(u)_m \right) \\
& - 12\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c a^{sr}(r)_c \right) \left( \sum' (j)_c \kappa^{jk}(k)_m \right) \left( \sum' (u)_m \kappa^{ul}(l)_m \right) \\
& + 12\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (s)_c a^{sr}(r)_c \right) \left( \sum' (j)_c a^{jk}(k)_m \right) \left( \sum' (u)_m a^{ul}(l)_m \right) \\
& - 12\phi^2 \sum (f+2g)_c (f+g)_m \left( \sum' (r)_c a^{rk}(k)_m \right) \left( \sum' (s)_c a^{sl}(l)_m \right) \left( \sum' (u)_m m^{uj}(j)_c \right),
\end{aligned}$$

$$\begin{aligned}
A_{14} = & -6\phi \sum e_m \left( \sum' (j)_m m^{jr}(r)_m \right) \left( \sum' (s)_m a^{su}(u)_m \right) \\
& + 6\phi \sum t_m \left( \sum' (j)_m m^{jr}(r)_m \right) \left( \sum' (s)_m a^{su}(u)_m \right) \\
& - 6\phi \sum t_m \left( \sum' (j)_m m^{jr}(r)_m \right) \left( \sum' (s)_m \kappa^{su}(u)_m \right) \\
& + 12\phi \sum t_m \left( \sum' (j)_m m^{ju}(u)_m \right) \left( \sum' (r)_m a^{rs}(s)_m \right) \\
& - 12\phi \sum d_m \left( \sum' (j)_m \kappa^{jr}(r)_m \right) \left( \sum' (s)_m \kappa^{su}(s)_u \right) \\
& + 12\phi \sum d_m \left( \sum' (j)_m a^{jr}(r)_m \right) \left( \sum' (s)_m a^{su}(s)_u \right).
\end{aligned}$$

Colocando os termos  $A_{11}$ ,  $A_{12}$ ,  $A_{13}$  e  $A_{14}$  acima na forma matricial, e fazendo a soma dos mesmos, obtemos (3.11).

### 3.11.2 Obtenção das quantidades $A_{1,\beta\phi}$ e $A_{2,\beta\phi}$

Como mencionado anteriormente, os  $A$ 's, que são quantidades que definem a expansão para a função de distribuição da estatística gradiente em MLG com  $\phi$  desconhecido, dependem dos respectivos  $A$ 's obtidos assumindo que o parâmetro de dispersão é conhecido, acrescidos de alguns termos extras. Assim, a partir do Teorema 2.1 do Capítulo 2, pode ser provado que, para a estatística gradiente, esses termos extras podem ser escritos como em (3.19) e (3.20). Substituindo as quantidades dadas em (3.21) e (3.22) em (3.19) e (3.20), podemos perceber que várias quantidades se anulam. Mostramos aqui a derivação das quantidades não-nulas. Primeiramente  $a^{\phi\phi} = -\phi^2/(nd_{(2)})$ . Assim,

$$\begin{aligned} \sum' \kappa_{jr\phi} \kappa_{\phi\phi\phi} m^{jr} (a^{\phi\phi})^2 &= -\frac{d_{(3)}\phi}{nd_{(2)}^2} \sum w_l \sum' (j)_l m^{jr} (r)_l \\ &= -\frac{d_{(3)}}{nd_{(2)}^2} \text{tr}[W(Z - Z_2)]. \\ &= -\frac{qd_{(3)}}{nd_{(2)}^2}, \end{aligned}$$

$$\begin{aligned} \sum' \kappa_{jr\phi} \kappa_{vw\phi} m^{jr} a^{vw} a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_l w_m \left( \sum' (j)_l m^{jr} (r)_l \right) \left( \sum' (v)_m a^{vw} (w)_m \right) \\ &= -\frac{1}{nd_{(2)}} \text{tr}[W(Z - Z_2)] \text{tr}[W Z_2] \\ &= -\frac{q(p - q)}{nd_{(2)}}, \end{aligned}$$

$$\begin{aligned} \sum' \kappa_{kl}^{(\phi)} \kappa_{jr\phi} a^{\phi\phi} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) &= -\frac{\phi^2}{nd_{(2)}} \sum w_l w_m \left( \sum' (j)_m \kappa^{jk} (k)_l \right) \left( \sum' (l)_l \kappa^{lr} (r)_m \right) \\ &\quad + \frac{\phi^2}{nd_{(2)}} \sum w_l w_m \left( \sum' (j)_m a^{jk} (k)_l \right) \left( \sum' (l)_l a^{lr} (r)_m \right) \\ &= -\frac{1}{nd_{(2)}} [\text{tr}(W Z W Z) - \text{tr}(W Z_2 W Z_2)] \\ &= -\frac{q}{nd_{(2)}}, \end{aligned}$$

$$\begin{aligned} \sum' \kappa_{\phi\phi}^{(\phi)} \kappa_{jr\phi} m^{jr} (a^{\phi\phi})^2 &= -\frac{\phi d_{(3)}}{nd_{(2)}^2} \sum w_l \sum' (j)_l m^{jr} r_l \\ &= -\frac{d_{(3)}}{nd_{(2)}^2} \text{tr}[W(Z - Z_2)] \\ &= -\frac{qd_{(3)}}{nd_{(2)}^2}, \end{aligned}$$

$$\begin{aligned}
\sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{jr} m^{uv} a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_l w_m \left( \sum' (j)_l m^{jr}(r)_l \right) \left( \sum' (u)_m m^{uv}(v)_m \right) \\
&= -\frac{1}{nd_{(2)}} \text{tr}[W(Z - Z_2)] \text{tr}[W(Z - Z_2)] \\
&= -\frac{q^2}{nd_{(2)}},
\end{aligned}$$

$$\begin{aligned}
\sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} m^{rv} a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_l w_m \left( \sum' (j)_l m^{ju}(u)_m \right) \left( \sum' (r)_l m^{rv}(v)_m \right) \\
&= -\frac{1}{nd_{(2)}} \text{tr}[W(Z - Z_2)W(Z - Z_2)] \\
&= -\frac{q}{nd_{(2)}}.
\end{aligned}$$

Analogamente,  $\sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{jv} m^{ru} a^{\phi\phi} = -q/nd_{(2)}$ . Substituindo as quantidades acima em (3.19) e (3.20), obtemos  $A_{1,\beta\phi}$  e  $A_{2,\beta\phi}$ . Podemos observar que essas quantidades são idênticas às quantidades  $A_{S1,\beta\phi}$  e  $A_{S2,\beta\phi}$ , obtidas em Cribari-Neto e Ferrari (1995) para a expansão da função de distribuição da estatística escore sob a hipótese nula em MLG com dispersão desconhecida.

### 3.11.3 Conjuntos de dados utilizados na Seção 3.9

Tabela 3.14: *Squid data*.<sup>a</sup>

observação	$y$ (libras)	$x_1$ (pol.)	$x_2$ (pol.)	$x_3$ (pol.)	$x_4$ (pol.)	$x_5$ (pol.)
1	1,95	1,31	1,07	0,44	0,75	0,35
2	2,90	1,55	1,49	0,53	0,90	0,47
3	0,72	0,99	0,84	0,34	0,57	0,32
4	0,81	0,99	0,83	0,34	0,54	0,27
5	1,09	1,05	0,90	0,36	0,64	0,30
6	1,22	1,09	0,93	0,42	0,61	0,31
7	1,02	1,08	0,90	0,40	0,51	0,31
8	1,93	1,27	1,08	0,44	0,77	0,34
9	0,64	0,99	0,85	0,36	0,56	0,29
10	2,08	1,34	1,13	0,45	0,77	0,37
11	1,98	1,30	1,10	0,45	0,76	0,38
12	1,90	1,33	1,10	0,48	0,77	0,38
13	8,56	1,86	1,47	0,60	1,01	0,65
14	4,49	1,58	1,34	0,52	0,95	0,50
15	8,49	1,97	1,59	0,67	1,20	0,59
16	6,17	1,80	1,56	0,66	1,02	0,59
17	7,54	1,75	1,58	0,63	1,09	0,59
18	6,36	1,72	1,43	0,64	1,02	0,63
19	7,63	1,68	1,57	0,72	0,96	0,68
20	7,78	1,75	1,59	0,68	1,08	0,62
21	10,15	2,19	1,86	0,75	1,24	0,72
22	6,88	1,73	1,67	0,64	1,14	0,55

<sup>a</sup>Tabela reproduzida de Myers (1990)

Tabela 3.15: *Breadwrapper data.*<sup>a</sup>

observação	$y$ (g/cm)	$x_1$ (°F)	$x_2$ (°F)	$x_3$ (%)
1	6,60	225,00	46,00	0,50
2	6,90	285,00	46,00	0,50
3	7,90	225,00	64,00	0,50
4	6,10	285,00	64,00	0,50
5	9,20	225,00	46,00	1,70
6	6,80	285,00	46,00	1,70
7	0,40	225,00	64,00	1,70
8	7,30	285,00	64,00	1,70
9	9,80	204,50	55,00	1,10
10	5,00	305,50	55,00	1,10
11	6,90	255,00	39,90	1,10
12	6,30	255,00	70,10	1,10
13	4,00	255,00	55,00	0,09
14	8,60	255,00	55,00	2,11
15	0,10	255,00	55,00	1,10
16	9,90	255,00	55,00	1,10
17	2,20	255,00	55,00	1,10
18	9,70	255,00	55,00	1,10
19	9,70	255,00	55,00	1,10
20	9,60	255,00	55,00	1,10

<sup>a</sup>Tabela reproduzida de Myers (1990)

# Capítulo 4

## Prioris de correspondência associadas à estatística gradiente

### Resumo

O foco deste capítulo é obter densidades a priori que dão validação frequentista com margem de erro  $o(n^{-1})$ , onde  $n$  é o tamanho da amostra, para regiões de credibilidade baseadas na estatística gradiente no caso multiparamétrico. Estas densidades a priori são denominadas prioris de correspondência. A abordagem é geral e considera possível presença de parâmetros de perturbação. Expressões explícitas são derivadas para os quantis a posteriori da estatística gradiente. Alguns exemplos são apresentados e discutidos.

*Palavras-chave:* Estatística gradiente; prioris de correspondência; Regiões de credibilidade; Regiões de confiança.

### 4.1 Introdução

A validação frequentista aproximada de regiões de credibilidade a posteriori diz respeito à determinação de densidades a priori para o vetor de parâmetros do modelo que determinam regiões de credibilidade para o vetor de parâmetros de interesse com propriedades de cobertura frequentista acuradas. Essas



densidades a priori são chamadas *prioris* de correspondência, e têm recebido muita atenção na literatura (Datta e Mukerjee, 2003). Algumas referências são Chang e Mukerjee (2010), que estabelecem condições para obtenção de *prioris* de correspondência para construção de regiões HPD (*higher posterior density regions*) e Chang e Mukerjee (2011), que fizeram o estudo de regiões de credibilidade baseadas na inversão de um teste de razão de verossimilhanças ajustado. Estes dois trabalhos fazem uma abordagem para dados dependentes. Ventura, Cabras e Racugno (2009) obtiveram intervalos de credibilidade com propriedades de cobertura frequentista acuradas a partir de distribuições a posteriori para um parâmetro de interesse obtidos através de uma verossimilhança perfilada. Aplicações deste trabalho a modelos específicos podem ser vistos em Min e Sun (2012) e Cabras et al. (2012). Ventura, Sartori e Racugno (2013) apresentaram aproximações para a distribuição marginal a posteriori para um parâmetro de interesse baseadas em *prioris* de correspondência. Simulações que avaliam a probabilidade de cobertura de regiões obtidas a partir desta distribuição marginal são apresentadas.

As *prioris* de correspondência são obtidas através de sistemas de equações diferenciais parciais obtidos através do argumento de encolhimento sugerido em Ghosh e Mukerjee (1991) e descrito em detalhes por Mukerjee e Reid (2000); ver Seção 1.2. A técnica do argumento de encolhimento consiste em três passos nos quais o objetivo principal é a determinação da esperança de funções mensuráveis de variáveis aleatórias. Essa sequência de passos é utilizada como uma rota Bayesiana para a obtenção de *prioris* de correspondência. Pode-se utilizar esta técnica a fim de reduzir a álgebra que é utilizada nos cálculos assintóticos frequentistas; ver Ghosh e Mukerjee (1991).

Uma das utilizações do argumento de encolhimento é a determinação de regiões de credibilidade via inversão dos testes estatísticos mais utilizados. Estabelecemos nesse capítulo condições para a validação frequentista aproximada de regiões de credibilidade para um vetor de parâmetros de interesse baseados na inversão da estatística gradiente. Esta validação frequentista se dá através da construção de regiões de credibilidade pela inversão da estatística, tomando como referência um quantil ajustado que depende dos dados e de uma *priori* para o vetor de parâmetros. Essa *priori* é a *matching prior*, que é determinada através de condições que são discutidas aqui.

O teste de hipóteses baseado na estatística gradiente proposta por Terrell (2002) é uma alternativa

aos testes mais comuns utilizados para amostras de tamanho grande, a saber, os testes da razão de verossimilhanças (Wilks, 1938), Wald (Wald, 1943) e escore (Rao, 1948). A estatística gradiente não depende da matriz de informação, nem esperada, nem observada, e é, portanto muito mais simples de ser computada do que as estatísticas de Wald e escore. Lemonte e Ferrari (2012b) fizeram o estudo do poder local do teste baseado na estatística gradiente, determinando sua expansão até ordem  $O(n^{-1/2})$  sob alternativas de Pitmann, e concluíram que nenhum dos testes, a saber, razão de verossimilhanças, escore, de Wald e gradiente, é uniformemente mais poderoso que os demais.

O estudo da validação frequentista de intervalos de credibilidade baseados na estatística gradiente complementa estudos similares feitos para o caso da estatística da razão de verossimilhanças (Ghosh e Mukerjee, 1991, 1992) e para as estatísticas escore e Wald (Rao e Mukerjee, 1995).

Este capítulo se divide do seguinte modo. Na Seção 4.2 estão os resultados principais do capítulo, a saber, as condições para determinação de prioris de correspondência associadas à inversão da estatística gradiente. As Seções 4.3, 4.4 e 4.5 se destinam a particularizações dos resultados da Seção 4.2 aos modelos uniparamétricos, biparamétricos e aos modelos multiparamétricos com parâmetro escalar de interesse ortogonal ao vetor de parâmetros de perturbação, respectivamente. Fazemos uma breve discussão na Seção 4.6 e detalhes técnicos estão na Seção 4.7.

## 4.2 Principais resultados

Sejam  $x_1, \dots, x_n$  uma amostra aleatória de tamanho  $n$  em que cada  $x_i$  possui função densidade de probabilidade  $f(\cdot; \boldsymbol{\theta})$ , que depende de um vetor  $p$ -dimensional de parâmetros desconhecidos  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ . Seja  $\ell(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta})$  e  $\mathbf{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$  o logaritmo da função de verossimilhança e o vetor escore, respectivamente; convenientemente ambos estão divididos por  $n$ . Suponha que o interesse esteja no vetor de parâmetros  $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_q)^\top$  e que  $\boldsymbol{\theta}_2 = (\theta_{q+1}, \dots, \theta_p)^\top$  seja um vetor de parâmetros de perturbação. A partição de  $\boldsymbol{\theta}$  induz uma partição em  $\mathbf{U}(\boldsymbol{\theta})$ :  $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_1(\boldsymbol{\theta})^\top, \mathbf{U}_2(\boldsymbol{\theta})^\top)^\top$ . Seja  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top)^\top$  e  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^\top, \tilde{\boldsymbol{\theta}}_2^\top)^\top$  os estimadores de máxima verossimilhança

irrestrito e restrito de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ , respectivamente. A estatística gradiente para  $\boldsymbol{\theta}$  é dada por

$$S = n\mathbf{U}(\tilde{\boldsymbol{\theta}})^\top(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \quad (4.1)$$

e pode ser escrita como  $S = n\mathbf{U}_1(\tilde{\boldsymbol{\theta}})^\top(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$ , pois  $\mathbf{U}_2(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ . Assim como as estatísticas da razão de verossimilhanças, de Wald e escore, a estatística gradiente tem distribuição assintótica  $\chi_q^2$  sob a hipótese nula, em que  $q$  é o número de entradas do vetor de parâmetros de interesse.

Dada uma densidade a priori  $\pi(\cdot)$  para  $\boldsymbol{\theta}$ , estamos interessados em regiões de credibilidade a posteriori para o parâmetro de interesse  $\boldsymbol{\theta}_1$  baseadas na inversão de  $S$ . A região de credibilidade é dada por

$$R_S^{(1-\alpha)}(\pi, \mathbf{x}) = \{\boldsymbol{\theta}_1 : S \leq k_{1-\alpha}(\pi, \mathbf{x})\}, \quad (4.2)$$

em que  $0 < \alpha < 1$ ,  $\mathbf{x} = (x_1, \dots, x_n)^\top$  e  $k_{1-\alpha}(\pi, \mathbf{x})$  é o quantil aproximado de ordem  $1 - \alpha$  da distribuição a posteriori de  $S$ , que depende de  $\pi(\cdot)$  e  $\mathbf{x}$ , e é escolhido através da relação

$$P^\pi[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = 1 - \alpha + o(n^{-1}), \quad (4.3)$$

em que  $P^\pi[\cdot | \mathbf{X} = \mathbf{x}]$  designa a medida de probabilidade a posteriori sob a densidade a priori  $\pi(\cdot)$ . O objetivo aqui é determinar a probabilidade de cobertura frequentista  $P_\theta[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{x})]$  com erro de aproximação  $o(n^{-1})$ , a fim de caracterizar as prioris de correspondência para  $\boldsymbol{\theta}$ , ou seja, as densidades a priori  $\pi(\cdot)$  para  $\boldsymbol{\theta}$  que satisfazem

$$\left| P_\theta[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{x})] - P^\pi\{\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}\} \right| = o(n^{-1}),$$

ou seja, que aproximem as probabilidades de cobertura frequentista e Bayesiana das regiões de credibilidade baseadas na inversão da estatística gradiente.

Introduzimos agora algumas notações. Exceto quando indicado, os índices  $j, r, s, u, v$ , e  $w$  variam de 1 a  $p$  e os índices  $j', r', s', u', v'$ , e  $w'$  variam de 1 a  $q$ . Além disso, utilizamos a convenção de Einstein, em que um índice repetido tanto com sobrescrito quanto subscrito indica uma soma implícita na variação indicada. Seja  $D_j$  o operador diferencial em relação ao parâmetro  $\theta_j$ ,  $\lambda_{jr} = -\psi_{jr} = -\{D_j D_r \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ ,  $\psi_{jrs} = \{D_j D_r D_s \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ ,  $\psi_{jrsu} = \{D_j D_r D_s D_u \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ , etc. A

matriz  $\Lambda = ((\lambda_{jr}))$  é a matriz de informação observada avaliada em  $\widehat{\theta}$ . A partição de  $\theta = (\theta_1^\top, \theta_2^\top)^\top$  induz as partições

$$\Lambda = ((\lambda_{jr})) = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad \Lambda^{-1} = ((\lambda^{jr})) = \begin{bmatrix} \Lambda^{11} & \Lambda^{12} \\ \Lambda^{21} & \Lambda^{22} \end{bmatrix},$$

em que  $\Lambda^{-1}$  é a inversa de  $\Lambda$ . Sejam  $\Lambda^{11-1} = ((\lambda_{1w'j'}))$ ,  $\sigma^{jr} = \lambda^{jr} - \lambda^{jw'} \lambda_{1w'j'} \lambda^{j'r}$ ,  $\tau^{jj'} = \lambda^{jw'} \lambda_{1w'j'}$ ,  $\sigma_{suvw}^{(1)} = \sigma^{su} \sigma^{vw} [3]$ ,  $\lambda_{j'r's'u'}^{(1)} = \lambda^{j'r'} \lambda^{s'u'} [3]$ , e  $\lambda_{j'r's'u'v'w'}^{(2)} = \lambda^{j'r'} \lambda^{s'u'} \lambda^{v'w'} [15]$ , em que  $[\cdot]$  denota um somatório com o número entre colchetes indicando o número de termos obtidos pela permutação dos índices. Por exemplo,  $\sigma^{su} \sigma^{vw} [3] = \sigma^{su} \sigma^{vw} + \sigma^{sv} \sigma^{uw} + \sigma^{sw} \sigma^{uv}$ . Seja  $\pi_j = D_j \pi(\theta)$ ,  $\pi_{jr} = D_j D_r \pi(\theta)$ ,  $\widehat{\pi} = \pi(\widehat{\theta})$ ,  $\widehat{\pi}_j = \pi_j(\widehat{\theta})$  e  $\widehat{\pi}_{jr} = \pi_{jr}(\widehat{\theta})$ . Uma expansão para a função característica a posteriori de  $S$  sob a densidade a priori  $\pi(\cdot)$  para  $\theta$  até ordem  $o(n^{-1})$  é determinada na Seção 2.6.1 do Capítulo 1 e é dada por

$$M_\pi(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^3 H_i (1 - 2\xi)^{-i} \right\} + o_p(n^{-1}), \quad (4.4)$$

em que  $H_0 = -(H_1 + H_2 + H_3)$ ,

$$H_1 = \frac{9}{8} \Psi_{j'r's'}^{(3)} \Psi_{u'v'w'}^{(3)} \lambda_{j'r's'u'v'w'}^{(2)} + \Gamma_{j'r'}^{(2)} \lambda^{j'r'} + \lambda_{j'r's'u'}^{(1)} \left\{ 2(\Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'}^{(3)} \Psi_{u'}^{(1)}) + \frac{3}{2} \Psi_{j'r's'}^{(3)} \Gamma_{u'}^{(1)} \right\},$$

$$H_2 = -\frac{3}{4} \Psi_{j'r's'}^{(3)} \Psi_{u'v'w'}^{(3)} \lambda_{j'r's'u'v'w'}^{(2)} + \lambda_{j'r's'u'}^{(1)} \left\{ \Gamma_{j'r's'u'}^{(4)} - 2(\Psi_{j'r's'u'}^{(4)} - \Psi_{j'r's'}^{(3)} \Psi_{u'}^{(1)}) - \frac{3}{2} \Psi_{j'r's'}^{(3)} \Gamma_{u'}^{(1)} \right\},$$

$$H_3 = \frac{1}{8} \Psi_{j'r's'}^{(3)} \Psi_{u'v'w'}^{(3)} \lambda_{j'r's'u'v'w'}^{(2)},$$

em que  $\xi = it$  e  $\Psi_{u'}^{(1)}$ ,  $\Psi_{j'r's'}^{(3)}$ ,  $\Psi_{j'r's'u'}^{(4)}$ ,  $\Gamma_{u'}^{(1)}$ ,  $\Gamma_{j'r'}^{(2)}$  e  $\Gamma_{j'r's'u'}^{(4)}$  são quantidades de ordem  $O_p(1)$ , dadas no apêndice.

Sejam  $\Upsilon_j(\cdot)$  e  $v(\cdot)$  a função distribuição acumulada e a densidade de uma variável aleatória qui-quadrado com  $j$  graus de liberdade ( $j = 1, 2, \dots$ ) respectivamente e  $z_{1-\alpha}^2$  o quantil de ordem  $1 - \alpha$  de

uma distribuição qui-quadrado com  $q$  graus de liberdade. Por demonstração análoga à feita em Datta e Mukerjee (2003) (ver Lema 4.1 na Seção 4.7) pode ser demonstrado que, se

$$k_{1-\alpha}(\pi, \mathbf{x}) = z_{1-\alpha}^2 - \frac{1}{nv_q(z^2)} \sum_{j=0}^3 H_j \Upsilon_{q+2j}(z^2), \quad (4.5)$$

então (4.3) é satisfeita.

De posse da função característica a posteriori aproximada da estatística gradiente sob a densidade a priori  $\pi(\cdot)$ , aplicamos os três passos do argumento de encolhimento e determinamos a probabilidade de cobertura frequentista  $P_\theta[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{x})]$  com erro de ordem  $o(n^{-1})$ .

**Passo 1.** Consideramos uma densidade a priori auxiliar  $\bar{\pi}(\cdot)$  satisfazendo as condições de Bickel e Ghosh (1990), a saber,  $\bar{\pi}(\cdot)$  tem suporte compacto no espaço paramétrico e se anula bem como suas duas primeiras derivadas parciais, na fronteira do seu suporte. Assim, escrevemos  $M_{\bar{\pi}}(t)$  que é a função característica posteriori aproximada (4.4) sob a densidade a priori  $\bar{\pi}(\cdot)$ , substituindo  $\pi(\cdot)$  por  $\bar{\pi}(\cdot)$ , ou seja,

$$M_{\bar{\pi}}(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^3 \bar{H}_i (1 - 2\xi)^{-i} \right\} + o_p(n^{-1}), \quad (4.6)$$

em que  $\bar{H}_i$  é a contraparte de  $H_i$  substituindo  $\pi(\cdot)$  por  $\bar{\pi}$ . Podemos observar que  $\bar{H}_3 = H_3$ , pois  $H_3$  independe de densidade priori, e que várias quantidades dadas em  $H_1$ ,  $H_2$  e  $H_0$  ficam intactas, já que também independem de densidade a priori. Assim, pela inversão de (4.6), temos  $P^{\bar{\pi}}[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = P^{\bar{\pi}}[S \leq k_{1-\alpha}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}]$ , em que

$$P^{\bar{\pi}}[S \leq k_{1-\alpha}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = 1 - \alpha + n^{-1} \sum_{j=0}^3 (\bar{H}_j - H_j) \Upsilon_{q+2j}(z^2) + o(n^{-1}).$$

Utilizando as relações  $\Upsilon_m(z^2) - \Upsilon_{m+2}(z^2) = 2v_{m+2}(z^2) = 2m^{-1}z^2v_m(z^2)$ , e após alguma álgebra, temos

$$P^{\bar{\pi}}[S \leq k_{1-\alpha}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = 1 - \alpha + \frac{2z^2v_q(z^2)}{nq} \left\{ -(\hat{\Xi}_1 + \hat{\Xi}_2) + (\hat{\Xi}_3 - \hat{\Xi}_2) \frac{z^2}{q+2} \right\} + o(n^{-1}), \quad (4.7)$$

em que

$$\begin{aligned}\widehat{\Xi}_1 &= \frac{1}{2} \left( \frac{\widehat{\pi}_{jr}}{\widehat{\pi}} - \frac{\widehat{\pi}_{jr}}{\widehat{\pi}} \right) \tau^{jj'} \tau^{rr'} \lambda^{j'r'} + \frac{1}{2} \psi_{jrs} \left( \frac{\widehat{\pi}_u}{\widehat{\pi}} - \frac{\widehat{\pi}_u}{\widehat{\pi}} \right) \sigma^{su} \tau^{jj'} \tau^{rr'} \lambda^{j'r'} \\ &\quad + \frac{1}{2} \psi_{jsu} \left( \frac{\widehat{\pi}_r}{\widehat{\pi}} - \frac{\widehat{\pi}_r}{\widehat{\pi}} \right) \sigma^{su} \tau^{jj'} \tau^{rr'} \lambda^{j'r'},\end{aligned}$$

$$\widehat{\Xi}_2 = \frac{1}{6} \psi_{jrs} \left( \frac{\widehat{\pi}_u}{\widehat{\pi}} - \frac{\widehat{\pi}_u}{\widehat{\pi}} \right) \tau^{jj'} \tau^{rr'} \tau^{ss'} \tau^{uu'} \lambda_{j'r's'u'}^{(1)},$$

$$\widehat{\Xi}_3 = \frac{1}{4} \psi_{jrs} \left( \frac{\widehat{\pi}_j}{\widehat{\pi}} - \frac{\widehat{\pi}_j}{\widehat{\pi}} \right) \tau^{jj'} \tau^{rr'} \tau^{ss'} \tau^{uu'} \lambda_{j'r's'u'}^{(1)}.$$

**Passo 2.** Sejam  $U_j = D_j \ell(\boldsymbol{\theta})$ ,  $U_{jr} = D_j D_r \ell(\boldsymbol{\theta})$ ,  $U_{jrs} = D_j D_r D_s \ell(\boldsymbol{\theta})$ ,  $U_{jr su} = D_j D_r D_s D_u \ell(\boldsymbol{\theta})$ ,  $\kappa_{jr} = E(U_{jr})$ ,  $\kappa_{jrs} = E(U_{jrs})$ ,  $\kappa_{jr su} = E(U_{jr su})$  e  $\mathbf{K}$  a matriz de informação de Fisher por observação,

$$\mathbf{K} = -((\kappa_{jr})) = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix},$$

em que  $\mathbf{K}^{-1} = -((\kappa^{jr}))$  é a sua inversa. Definimos as matrizes

$$\mathbf{A} = ((a^{jr})) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{22}^{-1} \end{bmatrix}, \quad (4.8)$$

$$\mathbf{M} = ((m^{jr})) = \mathbf{K}^{-1} - \mathbf{A}. \quad (4.9)$$

Por (4.7) e após alguma álgebra, temos

$$\begin{aligned}T(\boldsymbol{\theta}) &= E_{\boldsymbol{\theta}} \{ P^{\bar{\pi}} [S \leq k_{1-\alpha}(\bar{\pi}, \mathbf{X}) | \mathbf{X} = \mathbf{x}] \} = 1 - \alpha + \frac{2z^2 v_q(z^2)}{nq} \left\{ \Xi_1 \right. \\ &\quad \left. + \left[ 1 - \frac{z^2}{2(q+2)} \right] \Xi_2 \right\} + o(n^{-1}),\end{aligned} \quad (4.10)$$

em que

$$\begin{aligned}\Xi_1 &= \frac{1}{2} m^{jr} \left( \frac{\bar{\pi}_{jr}}{\bar{\pi}} - \frac{\pi_{jr}}{\pi} \right) + \frac{1}{2} \kappa_{jrs} \left( \frac{\bar{\pi}_u}{\bar{\pi}} - \frac{\pi_u}{\pi} \right) a^{su} m^{jr} \\ &\quad + \frac{1}{2} \kappa_{jsu} \left( \frac{\bar{\pi}_r}{\bar{\pi}} - \frac{\pi_r}{\pi} \right) a^{su} m^{jr},\end{aligned}$$

$$\Xi_2 = \frac{1}{6} \kappa_{jrs} \left( \frac{\bar{\pi}_u}{\bar{\pi}} - \frac{\pi_u}{\pi} \right) (m^{jr} m^{su} [3])$$

**Passo 3.** Agora, supomos que o suporte de  $\bar{\pi}(\cdot)$  contém o verdadeiro valor de  $\theta$ . Utilizando as condições de Bickel e Ghosh (1990), integramos (4.10) por partes em relação a  $\bar{\pi}(\cdot)$ , ou seja,

$$\int T(\theta) \bar{\pi}(\theta) d\theta,$$

e fazemos  $\bar{\pi}(\cdot)$  convergir em distribuição para a densidade a priori degenerada no verdadeiro  $\theta$ . Assim,

$$\begin{aligned} P_{\theta}[\theta_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{x})] &= 1 - \alpha + \frac{z^2 v_q(z^2)}{nq\pi(\theta)} \left\{ \Delta_1(\pi, \theta) \right. \\ &\quad \left. + \frac{1}{3} \left[ 1 - \frac{z^2}{2(q+2)} \right] \Delta_2(\pi, \theta) \right\} + o(n^{-1}), \end{aligned} \quad (4.11)$$

em que

$$\begin{aligned} \Delta_1(\pi, \theta) &= m^{jr} \pi_{jr}(\theta) - D_j D_r (m^{jr}) \pi(\theta) + D_u [\kappa_{jrs} m^{jr} a^{su} \pi(\theta)] \\ &\quad + D_r [\kappa_{jsu} m^{jr} a^{su} \pi(\theta)], \end{aligned} \quad (4.12)$$

$$\Delta_2(\pi, \theta) = D_u [\kappa_{jrs} m^{jr} m^{su} \pi(\theta)]. \quad (4.13)$$

Desta forma, o lado esquerdo de (4.11) se iguala a  $1 - \alpha + o(n^{-1})$  para todo  $\alpha$  e todo  $\theta$  se e somente se  $\Delta_1(\pi, \theta) = 0$  e  $\Delta_2(\pi, \theta) = 0$ . Com isso, caracterizamos as matching priors associadas à estatística gradiente, e enunciamos o seguinte resultado.

**Teorema 4.1** *Uma densidade a priori  $\pi(\cdot)$  assegura validação frequentista com margem de erro de ordem  $o(n^{-1})$ , em que  $n$  é o tamanho da amostra, de regiões de credibilidade para o parâmetro de interesse  $\theta_1$  dadas pela inversão da estatística gradiente  $S$  se e somente se satisfaz o seguinte sistema de equações diferenciais parciais:*

$$\Delta_1(\pi, \theta) = 0, \quad (4.14)$$

e

$$\Delta_2(\pi, \theta) = 0 \quad (4.15)$$

em que  $\Delta_1(\pi, \theta)$  e  $\Delta_2(\pi, \theta)$  são dados por (4.12) e (4.13) respectivamente.

□

Na ausência de parâmetros de perturbação, temos que  $q = p$ ,  $\mathbf{A} = \mathbf{0}$  e  $\mathbf{M} = \mathbf{K}^{-1}$ . Assim, podemos enunciar o seguinte corolário.

**Corolário 4.1** *Uma densidade a priori  $\pi(\cdot)$  assegura validação frequentista com margem de erro de ordem  $o(n^{-1})$  de regiões de credibilidade para o parâmetro  $\boldsymbol{\theta}$ , dada pela inversão da estatística gradiente  $S$  se e somente se satisfaz o seguinte sistema de equações diferenciais parciais*

$$\delta_1(\pi, \boldsymbol{\theta}) = D_j D_r [\kappa^{jr} \pi(\boldsymbol{\theta})] - 2D_r [\kappa^{jr} \pi_j(\boldsymbol{\theta})] = 0 \quad (4.16)$$

e

$$\delta_2(\pi, \boldsymbol{\theta}) = D_u [\kappa_{jrs} \kappa^{jr} \kappa^{su} \pi(\boldsymbol{\theta})] = 0. \quad (4.17)$$

□

Do Corolário 4.1 observamos que as prioris de correspondência associadas à inversão da estatística gradiente na ausência de parâmetros de perturbação são exatamente as mesmas para o caso da estatística score e de Wald (Rao e Mukerjee, 1995). Conseqüentemente, se  $\pi(\cdot)$  satisfaz as condições do Corolário 4.1, então elas também satisfazem a condição para a obtenção de prioris de correspondência associadas à estatística da razão de verossimilhanças (Ghosh e Mukerjee, 1991), que é  $\delta_1(\pi, \boldsymbol{\theta}) + \delta_2(\pi, \boldsymbol{\theta}) = 0$ . Uma observação importante é que a *matching prior* pode não ser própria.

### 4.3 O caso uniparamétrico

Aqui, consideramos os modelos indexados por um parâmetro unidimensional  $\phi$ , desconhecido. Do Corolário 4.1, as prioris de correspondência associadas à inversão da estatística gradiente satisfazem as equações diferenciais ordinárias

$$-\kappa^{\phi\phi} \pi_{\phi\phi}(\phi) - (\kappa^{\phi\phi})^2 D_\phi(\kappa_{\phi\phi}) \pi(\phi) = \text{constante}, \quad (4.18)$$

$$\kappa_{\phi\phi\phi} (\kappa^{\phi\phi})^2 \pi(\phi) = \text{constante}, \quad (4.19)$$

ou seja, as mesmas condições para o caso da estatística score em Datta e Mukerjee (2003) e Rao e Mukerjee (1995).



**Exemplo 4.1** (*Distribuição exponencial*)

Aqui,  $x_1, \dots, x_n$  é uma amostra aleatória da distribuição exponencial com densidade

$$f(x; \phi) = \frac{1}{\phi} e^{-x/\phi}, \quad x > 0, \quad \phi > 0.$$

Temos  $\kappa_{\phi\phi} = -\phi^{-2}$ ,  $\kappa_{\phi\phi\phi} = 4\phi^{-3}$  e  $\kappa_{\phi\phi\phi\phi} = -18\phi^{-4}$ . A estatística gradiente é dada por  $S = n(\bar{x} - \phi)^2/\phi^2$ , em que  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ . Aqui,  $\pi(\phi) \propto \phi^{-1}$  satisfaz (4.18) e (4.19), e é a densidade a priori imprópria de Jeffreys (Jeffreys, 1961). Assim o intervalo de credibilidade para  $\phi$  baseado em  $\pi(\cdot)$  e na inversão da estatística gradiente é

$$\left( \frac{\bar{x}}{1 + \sqrt{(1/n)k_{1-\alpha}}}, \frac{\bar{x}}{1 - \sqrt{(1/n)k_{1-\alpha}}} \right),$$

em que

$$k_{1-\alpha} = z_{1-\alpha}^2 - \frac{1}{nv_1(z_{1-\alpha}^2)} \left[ -\frac{1}{12} \Upsilon_1(z_{1-\alpha}^2) + \Upsilon_3(z_{1-\alpha}^2) - \frac{7}{4} \Upsilon_5(z_{1-\alpha}^2) + \frac{5}{6} \Upsilon_7(z_{1-\alpha}^2) \right], \quad (4.20)$$

$z_{1-\alpha}^2$  é o quantil de ordem  $1 - \alpha$  da distribuição qui-quadrado com um grau de liberdade,  $v_1(\cdot)$  é a densidade da distribuição qui-quadrado com 1 grau de liberdade e  $\Upsilon_j(\cdot)$  é a distribuição acumulada qui-quadrado com  $j$  graus de liberdade.

Na Figura 4.1, mostramos as probabilidades de cobertura de diferentes tipos de estimação intervalar para o parâmetro  $\phi$  da distribuição exponencial para o nível  $\alpha = 0,05$ . A linha cheia indica a probabilidade de cobertura frequentista para os intervalos de credibilidade para  $\phi$  baseando-se na *matching prior* determinada acima, utilizando o quantil  $k_{1-\alpha}$  como referência, a linha tracejada indica a probabilidade de cobertura para intervalos de confiança baseados na inversão da estatística gradiente tomando a distribuição  $\chi_1^2$  como referência, e a linha pontilhada, a inversão da estatística de Wald. O cálculo dessas probabilidades é exato, pois  $(n\bar{x})/\phi$  tem distribuição Gama( $n, 1$ ).

Podemos observar que os intervalos de credibilidade baseados na inversão da estatística gradiente utilizando a abordagem de prioris de correspondência tem probabilidade de cobertura mais próxima de 0,95 principalmente em pequenas amostras em relação aos intervalos de confiança baseados na inversão das estatísticas gradiente e de Wald usando os quantis da  $\chi_1^2$  como referência.

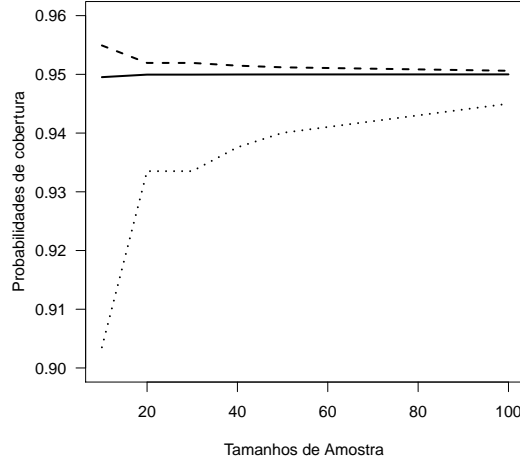


Figura 4.1: Probabilidades de cobertura frequentista para intervalos de credibilidade obtidos pela inversão do teste gradiente utilizando como referência o quantil (4.20) (linha cheia); probabilidades de cobertura dos intervalos de confiança baseados na inversão da estatística gradiente (tracejada) e de Wald (pontilhada) para  $\phi$  utilizando os quantis da distribuição  $\chi_1^2$  como referência

## 4.4 O caso biparamétrico

O caso biparamétrico corresponde à situação na qual estamos interessados em um parâmetro de interesse  $\phi$  e o modelo envolve ainda um parâmetro de perturbação  $\beta$ . De (4.9) e (4.8), temos que  $m^{\phi\phi} = -\kappa^{\phi\phi} = C$ ,  $m^{\phi\beta} = m^{\beta\phi} = -m^{\beta\beta}/E = -\kappa^{\beta\phi} = -CE$ ,  $-\kappa^{\beta\beta} = (C\kappa^{\phi\phi})/\kappa^{\beta\beta}$ ,  $a^{\phi\phi} = a^{\phi\beta} = a^{\beta\phi} = 0$ ,  $a^{\beta\beta} = -\kappa_{\beta\beta}^{-1}$ , em que  $C = (-\kappa_{\phi\phi} + \kappa_{\beta\beta}^{-1}\kappa_{\beta\phi}^2)^{-1}$  e  $E = \kappa_{\phi\beta}/\kappa_{\beta\beta}$ . Assim, do Teorema 4.1, as equações diferenciais parciais que nos dão as densidade prioris  $\pi(\cdot)$  desejadas são

$$\begin{aligned}
& C\pi_{\phi\phi}(\phi, \beta) - 2CE\pi_{\phi\beta}(\phi, \beta) + E^2C\pi_{\beta\beta}(\phi, \beta) - \pi(\phi, \beta)D_{\phi}^2(C) \\
& + 2\pi(\phi, \beta)D_{\phi}D_{\beta}(CE) - \pi(\phi, \beta)D_{\beta}^2(CE^2) - D_{\phi}[C\kappa_{\beta\beta}^{-1}(\kappa_{\phi\beta\beta} + \kappa_{\beta\beta\beta})\pi(\phi, \beta)] \\
& - D_{\beta}[\pi(\phi, \beta)C\kappa_{\beta\beta}^{-1}(\kappa_{\phi\phi\beta} - 3E\kappa_{\beta\beta\phi} + 2E^2\kappa_{\beta\beta\beta})] = 0,
\end{aligned} \tag{4.21}$$

e

$$D_\phi[\pi(\phi, \beta)C^2(\kappa_{\phi\phi\phi} - 3E\kappa_{\phi\phi\beta} + 3E^2\kappa_{\phi\beta\beta} - E^3\kappa_{\beta\beta\beta})] \quad (4.22)$$

$$+ D_\beta[\pi(\phi, \beta)C^2E(\kappa_{\phi\phi\phi} - 3E\kappa_{\phi\phi\beta} + 3E^2\kappa_{\phi\beta\beta} - E^3\kappa_{\beta\beta\beta})] = 0.$$

**Exemplo 4.2** (*Modelo de locação-escala*)

Como ilustração, tomamos um modelo simples de locação-escala (Datta e Mukerjee, 2003; DiCiccio, Kuffner e Young, 2012) dado por

$$g(x; \beta, \phi) = \frac{1}{\beta} f\left(\frac{x - \phi}{\beta}\right), \quad (4.23)$$

em que  $x, \phi \in \mathbb{R}$  e  $\beta > 0$ . Aqui o suporte de  $f(\cdot)$  é  $\mathbb{R}$ . Suponha que o parâmetro de locação  $\phi$  é o parâmetro de interesse e o parâmetro de escala  $\beta$  é um parâmetro de perturbação. Temos  $\kappa_{jr} \propto -\beta^{-2}$  e  $\kappa_{jrs} \propto \beta^{-3}$ , para todo  $j, r$  e  $s$  representando permutações dos parâmetros  $\beta$  e  $\phi$ . Assim,  $C \propto \beta^2$  e  $E = \text{constante}$ . Então a densidade a priori imprópria  $\pi(\beta, \phi) \propto \beta^{-1}$ , que é a densidade a priori de Jeffreys, satisfaz (4.21) e (4.22).

## 4.5 Parâmetro de interesse unidimensional ortogonal aos parâmetros de perturbação

Aqui, discutimos o caso em que o parâmetro de interesse  $\phi$  é escalar e ortogonal ao parâmetro de perturbação (Cox e Reid, 1987), que tem dimensão  $p - 1$ , ou seja,  $\boldsymbol{\theta} = (\phi, \theta_2, \dots, \theta_p)$ . Temos  $\kappa_{\phi j} = 0$  ( $j = 2, \dots, p$ ), em que o índice  $j \geq 2$  representa o parâmetro  $\theta_j$ . Do Teorema 4.1, uma densidade a priori  $\pi(\cdot)$  nos dá a validação frequentista aproximada para regiões de credibilidade a posteriori para  $\phi$ , baseadas na inversão da estatística gradiente satisfaz as seguintes equações:

$$- \kappa^{\phi\phi} \pi_{\phi\phi}(\boldsymbol{\theta}) + \pi(\boldsymbol{\theta}) D_\phi^2(\kappa^{\phi\phi}) + \sum_{s=2}^p \sum_{u=2}^p D_\phi[\kappa_{us\phi} \kappa^{us} \kappa^{\phi\phi} \pi(\boldsymbol{\theta})] + \quad (4.24)$$

$$+ \sum_{s=2}^p \sum_{u=2}^p D_u[\kappa_{\phi\phi s} \kappa^{su} \kappa^{\phi\phi} \pi(\boldsymbol{\theta})] = 0,$$

e

$$D_\phi[\kappa_{\phi\phi\phi}(\kappa^{\phi\phi})^2\pi(\boldsymbol{\theta})] = 0. \quad (4.25)$$

**Exemplo 4.3** (*Regressão linear simples*)

Consideramos aqui o modelo de regressão linear simples homoscedástico sem intercepto, ou seja,

$$y_i = \beta x_i + \epsilon_i, \quad i = 1, \dots, n,$$

em que  $\beta \in \mathbb{R}$  é um parâmetro desconhecido,  $x_i$  para  $i = 1, \dots, n$ , são constantes conhecidas e  $\epsilon_i$  tem distribuição  $N(0, \phi)$  para  $i = 1, \dots, n$  e são independentes. Para esse modelo, o logaritmo da função de verossimilhança é

$$l(\beta, \phi) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \phi - \frac{1}{2n\phi} \sum_{i=1}^n (y_i - \beta x_i)^2.$$

A informação esperada é dada por

$$\mathbf{K}(\beta, \phi) = \begin{bmatrix} \frac{\sum x_i^2}{n\phi} & 0 \\ 0 & \frac{1}{\phi^2} \end{bmatrix}$$

e  $\kappa_{\beta\beta\beta} = \kappa_{\beta\phi\phi} = \kappa_{\phi\phi\beta} = 0$ ,  $\kappa_{\beta\beta\phi} = \kappa_{\phi\beta\beta} = \sum x_i^2/n\phi^2$ ,  $\kappa_{\phi\phi\phi} = 2/\phi^3$ . Se  $\beta$  é o parâmetro de interesse, (4.24) se reduz a

$$\frac{n\phi}{\sum x_i^2} \pi_{\beta\beta}(\beta, \phi) + D_\phi[\phi\pi(\beta, \phi)] = 0.$$

Qualquer densidade a priori  $\pi(\beta, \phi)$  satisfaz (4.25), pois  $\kappa_{\beta\beta\beta} = 0$ . Assim, a densidade a priori  $\pi(\beta, \phi) \propto \phi^{-1}$  (imprópria) é solução o sistema de equações (4.24)–(4.25).

Se  $\phi$  é o parâmetro de interesse, (4.24) e (4.25) ficam dadas respectivamente por

$$\phi^2 \pi_{\phi\phi}(\beta, \phi) - 2\pi(\phi, \beta) + D_\phi[\phi\pi(\phi, \beta)] = 0,$$

e

$$D_\phi[\phi\pi(\phi, \beta)] = 0.$$

Portanto, a densidade a priori  $\pi(\beta, \phi) \propto \phi^{-1}$  é uma solução, que corresponde à densidade a priori não-informativa de Jeffreys.

**Exemplo 4.4** (*Distribuição normal bivariada*)

Consideremos a distribuição normal bivariada (Ghosh, Santra e Kim, 2008) cuja função densidade é dada por

$$f(x_1, x_2, \mu_1, \mu_2, \boldsymbol{\nu}) = (2\pi\theta)^{-1} \exp \left\{ -\frac{1}{2} \left[ \frac{(x_2 - \mu_2 - \beta(x_1 - \mu_1))^2}{\theta\eta} + \frac{\eta(x_1 - \mu_1)^2}{\theta} \right] \right\}, \quad (4.26)$$

em que  $\boldsymbol{\nu} = (\beta, \theta, \eta)$ ,  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $\beta = \rho\sigma_2\sigma_1$ ,  $\theta = \sigma_1\sigma_2(1 - \rho^2)^{\frac{1}{2}}$  e  $\eta = \sigma_2(1 - \rho^2)^{\frac{1}{2}}/\sigma_1$ , com  $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+$  e  $-1 \leq \rho \leq 1$ . Aqui,  $(x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})$  é uma amostra aleatória, em que cada  $(x_{1i}, x_{2i})$  possui distribuição normal bivariada com médias  $\mu_1$  e  $\mu_2$ , variâncias  $\sigma_1^2$  e  $\sigma_2^2$ , e coeficiente de correlação  $\rho$ . O vetor de médias  $(\mu_1, \mu_2)$  é ortogonal ao vetor  $(\beta, \theta, \eta)$  (Cox e Reid, 1987). Estamos interessados em prioris do tipo  $\pi(\mu_1, \mu_2, \beta, \theta, \eta) = \pi(\beta, \theta, \eta)$ . A matriz de informação esperada é

$$\mathbf{K}(\mu_1, \mu_2, \beta, \theta, \eta) = \begin{bmatrix} \mathbf{A}(\beta, \theta, \eta) & \mathbf{0} \\ \mathbf{0} & \text{diag}(\eta^{-2}, \theta^{-2}, \eta^{-2}) \end{bmatrix},$$

em que  $\mathbf{A}(\beta, \theta, \eta)$  é uma matriz  $2 \times 2$  não-nula. Aqui  $\kappa_{\beta\beta\beta} = \kappa_{\beta\theta\theta} = \kappa_{\beta\eta\eta} = \kappa_{\theta\eta\eta} = 0$ ,  $\kappa_{\beta\beta\theta} = \kappa_{\theta\eta\eta} = (\theta\eta^2)^{-1}$ ,  $\kappa_{\theta\theta\theta} = 4\theta^{-3}$ ,  $\kappa_{\beta\beta\eta} = \eta^{-3}$ ,  $\kappa_{\eta\eta\eta} = 3\eta^{-3}$ . Se considerarmos prioris da forma  $\pi(\beta, \theta, \eta) \propto \theta^m \eta^p$ , tanto quando  $\eta$  ou  $\theta$  ou  $\beta$  é o parâmetro de interesse, a priori  $\pi(\beta, \theta, \eta) \propto (\theta\eta)^{-1}$  é uma *matching prior*, associada à inversão da estatística gradiente.

## 4.6 Discussão

Rao e Mukerjee (1995) determinaram condições para validação frequentista até ordem  $o(n^{-1})$  de regiões de credibilidade a posteriori baseadas na estatística escore no caso multiparamétrico na ausência de parâmetros de perturbação, utilizando uma rota Bayesiana baseada no argumento de encolhimento. Os autores concluíram que as condições são as mesmas para a estatística de Wald, complementando resultados similares para a estatística da razão de verossimilhanças (Ghosh e Mukerjee, 1991, 1992). Como aplicação dos resultados de Ghosh e Mukerjee (1991), Ghosh, Santra e Kim (2008) determinaram prioris de correspondência associadas à estatística da razão de verossimilhanças para parâmetros de modelos normais bivariados, caracterizando regiões de credibilidade a posteriori com probabilidades de cobertura frequentista acuradas para esses parâmetros.

Neste capítulo, complementamos os resultados de Rao e Mukerjee (1995) obtendo prioris de correspondência para o caso da estatística gradiente (Terrell, 2002) na presença de possíveis parâmetros de perturbação. Concluimos que, na ausência de parâmetros de perturbação, as condições para determinação de prioris de correspondência associadas à estatística gradiente são exatamente as mesmas para o caso da estatística score e de Wald. Em um contexto geral, o argumento de encolhimento foi uma técnica crucial para derivar as equações diferenciais parciais cujas soluções são as prioris de correspondência associadas à inversão do teste gradiente. Obtidas essas densidades a priori, temos a expressão explícita do quantil ajustado que serve de referência para a determinação de regiões de credibilidade com probabilidades de cobertura frequentistas acuradas. Como os resultados obtidos aqui não estão vinculados a modelos específicos, podemos aplicá-los em várias classes de modelos paramétricos e fazer o estudo do desempenho das regiões de credibilidade baseadas na inversão do teste gradiente para esses modelos.

## 4.7 Detalhes técnicos

### 4.7.1 Quantidades dadas em (4.4)

$$\begin{aligned}\Psi_{j'}^{(1)} &= \frac{1}{2}\psi_{jrs}\sigma^{rs}\tau^{jj'}, \\ \Psi_{j'r's'}^{(3)} &= \frac{1}{6}\psi_{jrs}\tau^{jj'}\tau^{rr'}\tau^{ss'}, \\ \Psi_{j'r's'u'}^{(4)} &= \frac{1}{24}\{\psi_{jrsu} + \sigma^{vw}(2\psi_{jrs}\psi_{uvw} + 3\psi_{jrv}\psi_{suw})\}\tau^{jj'}\tau^{rr'}\tau^{ss'}\tau^{uu'}, \\ \Psi_{j'r'}^{(2)} &= \left\{ \frac{\widehat{\pi}_{jr}}{2\widehat{\pi}} + \frac{1}{4}\psi_{jrsu}\sigma^{su} + \frac{1}{24}(2\psi_{jrs}\psi_{uvw} + 3\psi_{jsu}\psi_{rvw})\sigma_{suw}^{(1)} \right\} \tau^{jj'}\tau^{rr'}, \\ \Gamma_{j'}^{(1)} &= \Psi_{j'}^{(1)} + \frac{\widehat{\pi}_j}{\widehat{\pi}}\tau^{jj'} \quad \Gamma_{j'r'}^{(2)} = \Psi_{j'r'}^{(2)} + \frac{1}{2\widehat{\pi}}(\psi_{jrs}\widehat{\pi}_u + \psi_{jsu}\widehat{\pi}_r)\sigma^{su}\tau^{jj'}\tau^{rr'}, \\ \Gamma_{j'r's'u'}^{(4)} &= \Psi_{j'r's'u'}^{(4)} + \frac{\widehat{\pi}_u}{6\widehat{\pi}}\psi_{jrs}\tau^{jj'}\tau^{rr'}\tau^{ss'}\tau^{uu'}.\end{aligned}$$

### 4.7.2 Lema auxiliar

**Lema 4.1** *Sejam  $\Upsilon_j(\cdot)$  e  $v_j(\cdot)$  respectivamente, a função distribuição acumulada e a densidade qui-quadrado com  $j$  graus de liberdade, e  $z_{1-\alpha}^2$  o quantil de ordem  $1 - \alpha$  de uma distribuição qui-quadrado com  $q$  graus de liberdade. Então a validade de (4.5) implica a validade de (4.3).*

*Prova.* Como

$$P^\pi[\boldsymbol{\theta}_1 \in R_S^{(1-\alpha)}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = P^\pi[S \leq k_{1-\alpha}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}],$$

e utilizando a expansão (4.4), temos

$$P^\pi[S \leq k_{1-\alpha}(\pi, \mathbf{X}) | \mathbf{X} = \mathbf{x}] = \Upsilon_q(k_{1-\alpha}(\pi, \mathbf{X})) + \frac{1}{n} \sum_{j=0}^3 H_j \Upsilon_{q+2j}(k_{1-\alpha}(\pi, \mathbf{X})) + o(n^{-1}). \quad (4.27)$$

Abrindo-se o somatório e expandindo termo a termo, temos

$$\begin{aligned}\Upsilon_q(k_{1-\alpha}(\pi, \mathbf{X})) &= 1 - \alpha - \frac{Z}{n} + o(n^{-1}), \\ \Upsilon_{q+2}(k_{1-\alpha}(\pi, \mathbf{X})) &= \Upsilon_{q+2}(z_{1-\alpha}^2) - \frac{z_{1-\alpha}^2 Z}{nq} + o(n^{-1}), \\ \Upsilon_{q+4}(k_{1-\alpha}(\pi, \mathbf{X})) &= \Upsilon_{q+4}(z_{1-\alpha}^2) - \frac{(z_{1-\alpha}^2)^2 Z}{nq(q+2)} + o(n^{-1}),\end{aligned}$$

$$\Upsilon_{q+6}(k_{1-\alpha}(\pi, \mathbf{X})) = \Upsilon_{q+6}(z_{1-\alpha}^2) - \frac{(z_{1-\alpha}^2)^3 Z}{nq(q+2)(q+4)} + o(n^{-1}),$$

em que

$$Z = \sum_{j=0}^3 H_j \Upsilon_{q+2j}(z_{1-\alpha}^2).$$

Substituindo as expansões acima em (4.27), o resultado é imediato. □



# Capítulo 5

## Considerações finais e pesquisas futuras

Basicamente, os resultados desta tese resumem-se nos seguintes pontos:

- Obtivemos uma expansão assintótica sob a hipótese nula para a função distribuição acumulada da estatística gradiente para teste de hipóteses na possível presença de parâmetros de perturbação;
- A partir dessa expansão propusemos uma estatística gradiente corrigida por um fator de correção tipo-Bartlett que possui distribuição qui-quadrado com erro de ordem  $o(n^{-1})$  sob a hipótese nula;
- Derivamos fórmulas matriciais e algébricas para o cálculo dos fatores de correção tipo-Bartlett para a estatística gradiente para testar parâmetros de MLG com dispersão conhecida e desconhecida;
- Derivamos condições para a obtenção de prioris de correspondência associadas à inversão da estatística gradiente na possível presença de parâmetros de perturbação. Propusemos um quantil ajustado que depende dessas prioris e dos dados, com o qual é possível a construção de regiões de credibilidade a posteriori baseados na inversão da estatística, que possuem probabilidades de cobertura frequentista acuradas, caracterizando assim a validação frequentista aproximada dessas regiões de credibilidade.

---

Vários trabalhos poderão ser desenvolvidos a partir de idéias e resultados dessa tese. Citamos alguns deles:

- Cálculo dos fatores de Bartlett para a estatística gradiente em MLG com dispersão heterogênea;
- Desenvolvimento de fórmulas matriciais e algébricas para os fatores de correção da estatística gradiente em modelos não-lineares da família exponencial e modelos de dispersão;
- Obtenção de regiões de credibilidade baseadas na estatística gradiente em modelos específicos, com estudos de simulação e aplicações de interesse prático;
- Estudo de poder de testes estatísticos usuais utilizando a abordagem do argumento de encolhimento.

Esperamos que através dos resultados dessa tese tenhamos contribuído de forma significativa na pesquisa de aperfeiçoamento de métodos assintóticos, e que vários trabalhos futuros sejam fruto das idéias aqui discutidas.

# Referências Bibliográficas

- Atkinson, A.C.(1982) Regression, diagnostics, transformations and constructed variables (with discussion). *Journal of the Royal Statistical Society B* **44**, 1–36.
- Bai, P. (2009). Sphericity test in a GMANOVA-MANOVA model with normal error. *Journal of Multivariate Analysis* **100**, 2305–2312.
- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A* **160**, 268–282.
- Bartlett, M.S. (1947). Multivariate Analysis. *Journal of the Royal Statistical Society (supplement)* **9**, 179–197.
- Bartlett, M.S. (1953a). Approximate confidence intervals. *Biometrika* **40**, 12–19.
- Bartlett, M.S. (1953b). Approximate confidence intervals, II. More than one unknown parameter. *Biometrika* **40**, 306–317.
- Bartlett, M.S. (1954). A note on the multiplying factors for various  $\chi^2$  approximations. *Journal of the Royal Statistical Society B* **16**, 296–298.
- Bickel, P.J., Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction – a Bayesian argument. *Annals of Statistics* **18**, 1070–1090.
- Birnbaum, Z.W., Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327.

- Cabras, S., Castellanos, M.E., Racugno, W., Ventura, L. (2012). A matching prior for the shape parameter of the skew-normal distribution. *Scandinavian Journal of Statistics* **39**, 236–247.
- Chandra, T.K. e Mukerjee, R. (1991). Bartlett-type modification for Rao's efficient score statistic. *Journal of Multivariate Analysis* **36**, 103–112.
- Chang, H.I., Mukerjee, R. (2010). Highest posterior density regions with approximate frequentist validity: the role of data-dependent priors. *Statistics and Probability Letters* **80**, 1791–1797.
- Chang, H.I., Mukerjee, R. (2011). Data-dependent probability matching priors for likelihood ratio and adjusted likelihood ratio statistics. *Statistics*. A aparecer, DOI:10.1080/02331888.2011.587880.
- Cook, R.D, Weisberg, S. (1983). Diagnostics for heterocedasticity in regression. *Biometrika* **70**, 1–10.
- Cordeiro, G.M., Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order  $n^{-1}$ . *Biometrika* **78**, 573–582.
- Cordeiro, G.M., Ferrari, S.L.P., Paula, G.A. (1993) Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629–643.
- Cordeiro, G.M., Cribari-Neto, F. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339–367.
- Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B* **45**, 404–413.
- Cordeiro, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika* **74**, 265–274.
- Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* **40**, 1–39.
- Cribari-Neto, F., Ferrari, S.L.P. (1995). Second order asymptotics for score tests in generalised linear models. *Biometrika* **82**, 426–432.

- Datta, G.S., Mukerjee, R. (2003). *Probability Matching Priors: Higher Order Asymptotics*. Springer-Verlag: New York.
- Davison, A., Tsai, C.L. (1990). Regression model diagnostics. *International Statistical Review* **60**, 337–353.
- DiCiccio, T.J., Kuffner, T.A., Young, G.A. (2012). Objective Bayes, conditional inference and the signed root likelihood ratio statistic. *Biometrika* **99**, 675–686.
- Doornik, J.A.(2008). *Ox: An Object-Oriented Matrix Programming Language*. Timberlake Consultants and Oxford, London, 5th ed.
- Ferrari, S.L.P., Lucâmbio, F., Cribari-Neto, F. (2005) Improved profile likelihood inference. *Journal of Statistical Planning and Inference* **134**, 373–391.
- Freund, R.J. (1983). Regression with SAS with emphasis on PROC REG. *Eighth Annual SAS Users Group International Conference* 16–19, New Orleans, Louisiana.
- Ghosh, J.K., Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *Journal of Multivariate Analysis* **38**, 385–393.
- Ghosh, J.K., Mukerjee, R. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *Journal of Royal Statistical Society B* **54**, 867–875.
- Ghosh, J.K., Mukerjee, R. (1994). Higher order comparison of tests: A Bayesian route. *Essays on Probability and Statistics in Honor of Professor A.K. Bhattacharyya*, Eds. S. P. Mukerjee, A. Chaudhuri and S.K. Basu, 92–101, Presidency College, Calcutta.
- Ghosh, M., Santra, U., Kim, D.(2008). Probability matching priors for some parameters of the bivariate normal distribution. *IMS Collections Publishing the Limits of Contemporary Statistics* **3**,71–81.
- Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* **72**, 653–659.

- Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* **29**, 359–378.
- Hill, G.W., Davis, A.W. (1968). Generalized asymptotic expansions of Cornish–Fisher type. *The Annals of Mathematical Statistics* **39**, 1264–73.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London.
- Lagos, B.M., Morettin, P.A. (2004). Improvement of the likelihood ratio test statistic in ARMA models. *Journal of Time Series Analysis* **25**, 83–101.
- Lagos, B.M., Morettin, P.A., Barroso, L.P. (2010). Some corrections of the score test statistic for Gaussian ARMA models. *Brazilian Journal of Probability and Statistics* **24**, 434–456.
- Larsen, P.V., Jupp, P.E. (2003). Parametrization–invariant Wald tests. *Bernoulli* **9**, 167–182.
- Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303.
- Lemonte, A.J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference* **141**, 1981–1989.
- Lemonte, A.J., Ferrari, S.L.P. (2011a). Size and power properties of some tests in the Birnbaum–Saunders regression model. *Computational Statistics and Data Analysis* **55**, 1109–1117.
- Lemonte, A.J. (2012). Local power properties of some asymptotic tests in symmetric linear regression models. *Journal of Statistical Planning and Inference* **142**, 1178–1188.
- Lemonte, A.J. (2012a). Nonnull asymptotic distributions of the LR, Wald, score and gradient statistics in generalized linear models with dispersion covariates. *Statistics. A* a aparecer, DOI:10.1080/02331888.2012.672422.
- Lemonte, A.J., Ferrari, S.L.P. (2012b). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics* **64**, 373–381.

- Lemonte, A.J., Ferrari, S.L.P. (2012c). A note on the local power of the LR, Wald, score and gradient tests. *Electronic Journal of Statistics* **6**, 421–434.
- Lemonte, A.J., Ferrari, S.L.P. (2012d). Local power and size properties of the LR, Wald, score and gradient tests in dispersion models. *Statistical Methodology* **9**, 537–554.
- McCullagh, P., Nelder, J.A. (1989) Generalized Linear Models. Chapman and Hall, 511p. *Mono-graphs on Statistics and Applied Probability*.
- Min, X., Sun, D. (2012). A matching prior based on the modified profile likelihood in a generalized Weibull stress-strength model. *The Canadian Journal of Statistics*. A aparecer. DOI: 10.1002/cjs.11164.
- Mukerjee, R., Reid, N. (2000). On the Bayesian approach for frequentist computations. *Brazilian Journal of Probability and Statistics* **14**, 159–166.
- Myers, R.H.(1990). *Classical and Modern Regression with Applications*, 2ed. Duxbury Press: Belmont.
- Myers, R.H.(1976). *Response surface methodology*, 1ed. Wiley: River Street, Hoboken.
- Nelder, J.A., Wedderburn, R.W.M. (1972). Generalised linear models *Journal of Royal Statistical Society A* **135**, 370–384.
- Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* **30**, 3304–3312.
- Pregibon, D.(1982). Score tests in GLIM with applications. *GILCHRIST, R., ed. GLIM 82: Proceedings of the International Conference on Generalized Linear* **1**, 87–97.
- Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50–57.

- Rao, C.R. (2005). Score test: historical review and recent developments. Em *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, N. Balakrishnan, N. Kannan and H. N. Nagaraja, eds. Birkhuser, Boston.
- Rao, C.R., Mukerjee, R. (2005). On posterior credible set based on the score statistic. *Statistica Sinica*, **5**, 781–791.
- Taniguchi, M. (1991). Third-order asymptotic properties of a class of tests statistics under a local alternative. *Journal of Multivariate Analysis*, **37**, 223–238.
- Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* **34**, 206–215.
- Tu, D., Chen, J., Shi, P., Wu, Y. (2005). A Bartlett type correction for Rao's score test in Cox regression model. *Sankhya* **67**, 722–735.
- van Giersbergen, N.P.A. (2009). Bartlett correction in the stable AR(1) model with intercept and trend. *Econometric Theory* **25**, 857–872.
- Ventura, L., Cabras, S., Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *Journal of the American Statistical Association* **104**, 768–774.
- Ventura, L., Sartori, N., Racugno, W. (2013). Objective Bayesian higher-order asymptotics in models with nuisance parameters. *Computational Statistics & Data Analysis*, **60**, 90–96.
- Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482.
- Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypothesis. *Annals of Mathematical Statistics* **9**, 60–62.
- Zucker, D.M., Lieberman, O., Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society B* **62**, 827–838.