



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

Programa de Pós-Graduação em Estatística

**CORREÇÃO DE ALTA ORDEM DE ESTIMADORES DE
MÁXIMA VEROSSIMILHANÇA**

WALDEMAR ARAÚJO DE SANTA CRUZ OLIVEIRA JÚNIOR

Tese de Doutorado

Recife - PE

Fevereiro de 2013

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

**CORREÇÃO DE ALTA ORDEM DE ESTIMADORES DE
MÁXIMA VEROSSIMILHANÇA**

WALDEMAR ARAÚJO DE SANTA CRUZ OLIVEIRA JÚNIOR

Trabalho apresentado ao Programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Estatística.

Orientador:
Prof. Dr.
Klaus Leite Pinto Vasconcellos

Recife, fevereiro de 2013

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da Silva, CRB4-1217

Oliveira Júnior, Waldemar Araújo de Santa Cruz
Correção de alta ordem de estimadores de máxima verossimilhança / Waldemar Araújo de Santa Cruz Oliveira Júnior. - Recife: O Autor, 2013.
xi, 104 f.: il., fig., tab.

Orientador: Klaus Leite Pinto Vasconcellos.
Tese (Doutorado) - Universidade Federal de Pernambuco.
CCEN. Estatística, 2013.

Inclui bibliografia e apêndice.

1. Estatística matemática. 2. Teoria assintótica. I.
Vasconcellos, Klaus Leite Pinto (orientador). II. Título.

519.5

CDD (23. ed.)

MEI2013 – 032

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

Recife, 20 de fevereiro de 2013.

Nós recomendamos que a tese de doutorado de autoria de

Waldemar Araújo de Santa Cruz O. Júnior

Intitulada

“Correção de Alta Ordem de Estimadores de Máxima Verossimilhança”

Seja aceita como cumprimento parcial dos requerimentos para o grau de Doutor(a) em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Klaus Leite Pinto de Vasconcellos

Orientador / UFPE

Francisco Cribari Neto

UFPE

Gauss Moutinho Cordeiro

UFPE

Artur José Lemonte

USP

Denise Aparecida Botter

USP

Este documento será anexado à versão final da tese.

Dedico este trabalho a minha esposa Ana Karla, a meus filhos Antônio, Bianca e Larissa. E a minha mãe, Izabel Araújo, que aos 82 anos de idade está cursando Doutorado em Psicanálise, mostrando, assim, que a busca pelo conhecimento não tem início, meio ou fim.

Agradecimentos

A realização deste trabalho só foi possível devido à ajuda de muitos amigos, aos quais sou muito grato. Assim, agradeço:

- A Deus, que a cada dia me mostra como a vida é mais fácil quando temos fé.
- A minha mãe Izabel e *in memoriam* a meu pai Waldemar, pelo amor, criação e exemplo de vida.
- A minha doce e amada esposa Ana Karla, pela paciência, amor e dedicação.
- A meu filho Antônio, minhas filhas Bianca e Larissa, por serem a razão da minha vida.
- A minhas irmãs, pela presença na minha vida.
- A Bela, Lua, João e Biel, por serem sobrinhos maravilhosos.
- Ao Prof. Dr. Klaus Leite Pinto Vasconcellos, pela dedicação, disponibilidade, paciência e orientação.
- A Ir. Inácio Ferreira Dantas, pela orientação espiritual ao longo da minha vida e pela amizade.
- A todos os professores do Departamento de Estatística, pelos ensinamentos, apoio e convivência.
- A todos meus colegas de curso, em particular a Marcelo e Cícero, pelos choppes após as “fáceis” provas de estatística matemática, a Laércio, Luz e Fernanda, pela excelente convivência.
- A Valéria, pela disponibilidade e atenção.
- À Secretaria de Fazenda do Estado de Alagoas, pelo apoio na realização deste curso, em particular a Jacque Damasceno, Luiz Antônio, Telma Lobo e Guataçara Franco, pelas infinitas autorizações de trocas de plantão.
- A todos os meus amigos da Secretaria de Fazenda do Estado de Alagoas, pela excelente convivência.

O conhecimento da verdade por si só não basta; ao contrário, tal conhecimento precisa ser constantemente renovado com esforços incessante, se não se deseja perdê-lo.

Albert Einstein.

Resumo

A técnica de estimação por máxima verossimilhança é uma das metodologias mais utilizadas na área de Estatística. Em determinados modelos, esta técnica produz um estimador viesado ou assintoticamente não-viesado. No último caso, a ordem de magnitude dos vieses desses estimadores é em geral $O(n^{-1})$ e seu desvio padrão na ordem de $O(n^{-1/2})$. Por esse motivo, esses vieses não são levados em conta em amostras de tamanho grande. Porém, em pequenas amostras esse viés na estimação pode ter um significado importante. Assim, o estudo sobre diminuir o viés do estimador de máxima verossimilhança torna-se bastante relevante em diversas áreas, tais como, medicina, farmácia, biologia, entre outras, que necessitam de precisão e ao mesmo tempo trabalham com amostras pequenas.

Durante décadas, muitos artigos foram publicados na área de correção de viés, utilizando diversos tipos de modelos e técnicas de estimação. Neste trabalho, propomos uma técnica de correção de viés baseada em uma sequência de translações da função escore, de forma que a primeira translação é exatamente a que David Firth propôs, ver [18]. Para isso, usamos inicialmente a expansão de Taylor do estimador de máxima verossimilhança para realizar a primeira translação, o zero desta função transladada é o estimador θ_0^* , que é o estimador proposto por Firth. Com a expansão de Taylor deste estimador, realizamos outra translação na função escore já transladada, encontrando o estimador θ_1^* . Sob determinadas condições de regularidade, o viés deste novo estimador tem ordem de magnitude $O(n^{-3})$. Repetindo esse processo k -vezes, obtemos um estimador cujo viés tem ordem de magnitude $O(n^{-k})$, para $k = 1, 2, \dots$. Realizamos várias simulações de Monte Carlo em uma grande variedade de situações e de modelos estatísticos. No caso uniparamétrico, comparamos o desempenho do estimador θ_1^* com o estimador de máxima verossimilhança $\hat{\theta}$, com θ_0^* , com $\hat{\theta}_1$ visto na equação 2.18 e com o estimador $\hat{\theta}_2$ proposto por Ferrari et al [17], que pode ser visto na equação 2.19. No caso biparamétrico, comparamos o estimador θ_1^* com os estimadores $\hat{\theta}$ e θ_2^* . Os resultados das simulações mostram que esses estimadores, cuja proposta é de corrigir viés, são competitivos entre si, mas há uma leve superioridade dos estimadores θ_1^* e $\hat{\theta}_2$. No caso biparamétrico é mais evidente a superioridade do estimador θ_1^* , para n pequeno.

Abstract

The technique of maximum likelihood estimation is one of the most used methodologies in the field of Statistics. In certain models, this technique produces an estimator asymptotically biased or non-biased. In the latter case, the order of magnitude of the bias of these estimators is generally $O(n^{-1})$ and their standard deviation has order $O(n^{-\frac{1}{2}})$, when n is the sample size. For this reason, these biases are not taken into account for large sample size. However, in small samples this bias in the estimation may have important meaning. Thus, the study of the bias of the maximum likelihood estimator becomes very important in many fields such as medicine, pharmacy, biology, among others, requiring precision while working with small samples.

For decades, many articles have been published in the area of bias correction, using various types of models and estimation techniques. In this thesis, we propose a technique for bias correction based on a series of translations of the score function, where the first translation is exactly what David Firth proposed, see [18]. For this, we initially use the Taylor expansion of maximum likelihood estimator to perform the first translation, the zero of the modified score function is the estimator θ_0^* , that is the estimator proposed by Firth. With the Taylor expansion of this estimator, we performed another translation in score function already modified, finding the estimator θ_1^* . Under certain regularity conditions, the bias of the new estimator has order of magnitude $O(n^{-3})$. Repeating this process k -times, we obtain an estimator whose bias has order of magnitude $O(n^{-k})$, for $k = 1, 2, \dots$. We conducted several Monte Carlo simulations in a variety of situations and statistical models. In the one-parameter case, we compare the performance of the estimator θ_1^* with the maximum likelihood estimator $\hat{\theta}$, with θ_0^* , with $\hat{\theta}_1$ seen in equation 2.18 and with the estimator $\hat{\theta}_2$ proposed by Ferrari et al [17], which can be seen in equation 2.19. In the two-parameter case, we compare the estimator θ_1^* with the estimators $\hat{\theta}$ and θ_0^* . The simulation results show that these estimators, whose proposal is to correct bias, are competitive with each other, but there is a slight superiority of estimators θ_1^* and $\hat{\theta}_2$. In the two-parameter case it is most evident the superiority of the estimator θ_1^* , for small n .

Sumário

LISTA DE FIGURAS	x
LISTA DE TABELAS	xi
1 INTRODUÇÃO	1
2 CONCEITOS BÁSICOS	5
2.1 Introdução	5
2.2 Histórico	6
2.3 Conceitos Básicos	8
2.4 Equações de Bartlett	11
2.5 Resíduos de Cox & Snell	16
2.6 Correção de Viés Preventiva	22
2.7 Uma Redução de Viés de Terceira Ordem para a Família Exponencial Uniparamétrica	26
3 CORREÇÃO DE VIÉS DE ORDEM SUPERIOR	30
3.1 Introdução	30
3.2 Expressão de Bartlett Generalizada	31
3.3 Estimador de Máxima Verossimilhança com Viés de Ordem Reduzida	40
3.4 Correção de Viés Biparamétrica	48
3.5 Estimação Multiparamétrica	59
3.6 Considerações Finais	66
4 AVALIAÇÃO NUMÉRICA	68
4.1 Introdução	68

4.2	Modelos Uniparamétricos	69
4.3	Modelos Biparamétricos	76
4.4	Considerações Finais	81
5	CONCLUSÕES	82
REFERÊNCIAS BIBLIOGRÁFICAS		85
A	Expansão da Equação (3.3)	91
B	Valores Esperados de Potências da Função Escore	94
C	Derivadas da Função Inversa da Função Escore	97

Listas de Figuras

2.1	Funções escore $L'(\theta)$ e escore modificada $L'^*(\theta)$.	24
4.1	Densidades da distribuição beta com parâmetros (μ, ϕ) .	70
4.2	Densidades da distribuição gama.	73

Listas de Tabelas

3.1	Possíveis valores do expoente de n .	35
3.2	Expressão de Bartlett quando $k = 2$.	36
4.1	Vieses dos estimadores do parâmetro p da densidade beta, $n = 10$.	71
4.2	Vieses dos estimadores do parâmetro p da densidade beta, $n = 15$.	72
4.3	Vieses dos estimadores do parâmetro α da densidade gama, $n = 10$.	74
4.4	Vieses dos estimadores do parâmetro α da densidade gama, $n = 15$.	75
4.5	Vieses dos estimadores do parâmetro canônico da densidade de Poisson.	76
4.6	Vieses dos estimadores do parâmetro canônico da densidade binomial.	77
4.7	Vieses dos estimadores do parâmetro canônico da densidade geométrica.	77
4.8	Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 15$.	78
4.9	Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 20$.	79
4.10	Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 25$.	80

Capítulo 1

INTRODUÇÃO

Um modelo estatístico pode ser caracterizado por um ou mais parâmetros, assim um dos objetivos da inferência estatística é estimar esses parâmetros. Uma técnica de estimação muito utilizada é estimação por máxima verossimilhança, que consiste em maximizar a função de verossimilhança com relação aos seus parâmetros. O estimador obtido a partir desta técnica é denominado estimador de máxima verossimilhança (EMV), representado neste trabalho por $\hat{\theta}$. Em muitos casos esse estimador é viesado, com viés da ordem $O(n^{-1})$ e desvio padrão com ordem de magnitude $O(n^{-\frac{1}{2}})$. Em geral, quando a amostra tem um tamanho grande, esse viés não é levado em conta, porém em amostras pequenas o EMV pode produzir estimativas com valores significativamente diferentes do verdadeiro valor do parâmetro do modelo. Nesta conjutura, uma área de bastante interesse na estatística é a correção de viés e que é objeto de estudo deste trabalho.

O ponto de máximo da função de verossimilhança corresponde ao zero da função escore, então, uma técnica de correção de viés foi proposta por David Firth [18], que consiste em transladar a função escore, de forma que θ_0^* zero desta nova função transladada é o EMV com viés corrigido. Podemos dizer que é feita uma correção de viés preventiva, pois, antes de estimar o parâmetro é feita a correção de viés, diferentemente de muitas técnicas em que estima-se o viés e depois o subtrai do estimador de máxima verossimilhança. Por exemplo, o estimador de segunda ordem $\hat{\theta}_1$ e o estimador de terceira ordem proposto por Ferrari et al. [17], que podem ser vistos nas equações (2.18) e (2.19), respectivamente.

Dada a expansão de Taylor da função escore em torno de θ_0^* , a proposta de David Firth consiste basicamente em retirar o termo de primeira ordem do viés de Cox & Snell. Inspirados nessa ideia de translação da função escore e com base na expressão (2.4), apresentada por Bartlett [2], propomos no espaço uniparamétrico uma técnica nova de estimação, de modo que o estimador obtido tenha viés de ordem de magnitude $O(n^{-k})$, para $k = 2, 3, \dots$. Essa nova metodologia de estimação, proposta nesta tese, consiste basicamente em fazer sucessivas translações na função escore, sendo que a primeira translação é exatamente a que David Firth propôs. A segunda translação tem como objetivo retirar os termos de segunda ordem do viés do estimador θ_0^* . O zero da função escore transladada pela segunda vez é o estimador θ_1^* , cujo viés tem ordem de magnitude $O(n^{-3})$. Repetindo este processo de translação k vezes, obtemos uma sequência de estimadores, tal que os três primeiros elementos são $\widehat{\theta}$, θ_0^* e θ_1^* , que têm vieses com ordem de magnitude $O(n^{-1})$, $O(n^{-2})$ e $O(n^{-3})$, respectivamente; e o k -ésimo elemento da sequência sendo um estimador cujo viés é da ordem $O(n^{-k})$.

Generalizamos esse raciocínio para o espaço biparamétrico. Dada a função escore biparamétrica $L'(\theta)$, utilizamos a expansão de Taylor do EMV, em torno do ponto $L'(\widehat{\theta})$ e propomos uma translação na função escore. Provamos que essa translação é exatamente a translação proposta por David Firth. Assim, o zero desta função transladada é o estimador θ_0^* . A seguir, dada a função escore transladada $L_0(\theta)$, fazemos a expansão de Taylor do estimador θ_0^* , em torno do ponto $L_0(\theta_0^*)$ e propomos uma translação na função L_0 , cujo zero é o estimador θ_1^* , que tem viés com ordem de magnitude $O(n^{-3})$. Repetindo esse processo k vezes, obtemos, como no caso uniparamétrico, uma sequência de estimadores, cujo k -ésimo elemento é um estimador cujo viés tem ordem de magnitude $O(n^{-k})$, para $k = 1, 2, \dots$. Nestes termos, esta técnica foi estendida para o caso multiparamétrico.

Com o intuito de avaliar o desempenho do estimador θ_1^* , no caso uniparamétrico, comparamos este estimador com os estimadores $\widehat{\theta}$, $\widehat{\theta}_1$, θ_0^* e $\widetilde{\theta}_2$. Já para o caso biparamétrico, o comparamos apenas com os estimadores $\widehat{\theta}$ e θ_0^* . Para isso, realizamos simulações de Monte Carlo em uma grande variedade de modelos esta-

tísticos e em uma grande diversidade de situações. Em cada simulação de Monte Carlo, calculamos o viés estimado de cada estimador, ou seja, simulamos dez mil réplicas de Monte Carlo e calculamos a média dos vieses desses estimadores.

Concluímos no caso uniparamétrico que, apesar desses estimadores serem competitivos entre si, há uma leve superioridade dos estimadores θ_1^* e $\tilde{\theta}_2$, inclusive há uma grande similaridade entre estes dois estimadores. Percebemos, que em alguns casos, como nos modelos Poisson, binomial e geométrico, pode não fazer sentido a estimação corretiva, pois, dependendo do parâmetro a ser estimado, o viés pode ser menos infinito ou até mesmo nem existir. No caso biparamétrico, a superioridade do estimador θ_1^* torna-se mais evidente, apesar de θ_0^* e θ_1^* serem competitivos entre si.

Esta tese de doutorado está dividida em cinco capítulos, sendo que no Capítulo 2 estão os conceitos básicos abordados ao longo do texto, além de um resumo de trabalhos publicados em correção de viés. Neste capítulo demonstramos a expressão (2.4), proposta por Bartlett, mas não demonstrada em seu artigo, ver [2]. Demonstramos a expressão para o viés de Cox & Snell, assim como demonstramos as expressões para os estimadores $\hat{\theta}_1$, θ_0^* e $\tilde{\theta}_2$.

Em seguida, no Capítulo 3, generalizamos a expressão (2.4), apresentada por Bartlett. Ou seja, expandimos o estimador obtido para encontrar o zero da função escore transladada em parcelas com ordem de magnitude $O_p(n^{-\frac{i}{2}})$, até o termo residual ser da ordem $O_p(n^{-\frac{k+1}{2}})$, para $k = 2, 3, \dots$. Este resultado está apresentado no Teorema 3.2.1 e através dele propomos o estimador θ_1^* , cujo viés tem ordem de magnitude $O(n^{-3})$. No Teorema 3.3.1 apresentamos uma sequência de estimadores uniparamétricos, cujos vises têm ordem de magnitude $O(n^{-k})$, para $k = 1, 2, \dots$. Generalizamos essa sequência para o caso biparamétrico, que pode ser visto no Teorema 3.4.1, e para o caso multiparamétrico, ver o Teorema 3.5.1.

Os resultados das simulações de Monte Carlo estão apresentados no Capítulo 4. Já no Capítulo 5 temos as conclusões desta tese, assim como sugestões para trabalhos futuros.

No Apêndice A expandimos a equação generalizada de Bartlett até o quinto

termo. Nas implementações das simulações de Monte Carlo usamos expressões fechadas para o valor esperado de potências da função escore, estas expressões são dadas no Apêndice B. A expansão de Taylor dos estimadores $\hat{\theta}$ e θ_0^* foi feita a partir das derivadas da função inversa da escore. Estas derivadas podem ser calculadas por meio de derivação implícita, sem necessariamente encontrar a expressão algébrica da função inversa. Assim, colocamos no Apêndice C as derivadas da função inversa da função escore.

As simulações deste artigo foram realizadas usando a linguagem de programação matricial Ox, que pode ser obtida no endereço <http://www.doornik.com>. Os gráficos foram construídos usando o ambiente gráfico e de programação R, disponível gratuitamente no endereço <http://www.r-project.org>.

Capítulo 2

CONCEITOS BÁSICOS

2.1 Introdução

Em geral, o EMV tem viés de ordem $O(n^{-1})$ e desvio padrão de ordem $O(n^{-\frac{1}{2}})$. Assim, em amostras grandes, este viés normalmente não é considerado. Porém, no caso de amostras pequenas, o estimador pode divergir significativamente do verdadeiro valor do parâmetro. Neste contexto, o estudo sobre correção de viés do EMV torna-se uma área de pesquisa de bastante interesse, pois é uma técnica voltada para diminuir o erro de estimação.

Uma técnica de correção de viés foi proposta por Firth [18], que consiste em fazer uma correção preventiva e para isso ele sugere fazer uma translação na função escore. Inspirados na generalização da equação (2.4) proposta por Bartlett [2], usamos a expansão de Taylor da função inversa da função escore para propor uma nova técnica de estimação, que consiste em generalizar a ideia desenvolvida por Firth, no sentido de desenvolver sequências de estimadores com vieses de ordem de magnitude $O(n^{-k})$, para qualquer $k = 3, 4, \dots$.

Com o intuito de desenvolver estes novos estimadores, realizamos neste capítulo um resumo sobre a teoria de correção de viés, no qual damos um breve histórico dos casos estudados, definimos os principais conceitos sobre o assunto, realizamos um estudo aprofundado em dois estimadores cujos os vieses têm ordens de magnitude $O(n^{-2})$ e $O(n^{-3})$, propostos, respectivamente, por Firth [18] e por Ferrari et al. [17], e detalhamos uma equação desenvolvida por Bartlett [2],

que é base desta tese de doutorado.

Este capítulo está dividido da seguinte forma: na Seção 2.2 resumimos brevemente os casos já estudados sobre este tema; na Seção 2.3 estudamos a teoria sobre estimação por máxima verossimilhança e ordem de magnitude; já na Seção 2.4 demonstramos a equação de Bartlett, equação (2.4), sob determinadas condições; na Seção 2.5 discutimos os resíduos de Cox & Snell; por fim, nas Seções 2.6 e 2.7 apresentamos, respectivamente, o estimador proposto por Firth [18], baseado na translação da função escore, para o qual o viés é $O(n^{-\frac{3}{2}})$ e o estimador proposto por Ferrari, Botter, Cordeiro & Cribari-Neto [17], cujo viés tem ordem de magnitude $O(n^{-3})$.

2.2 Histórico

Nesta seção fazemos um breve histórico dos casos envolvendo correção de viés do EMV já estudados.

Começamos pelo ano de 1953, onde Bartlett [2] apresentou a equação (2.5), que é uma expressão para o viés de ordem $O(n^{-1})$ do EMV no caso uniparamétrico. Esta expressão é estudada na Seção 2.4, a qual demonstramos e colocamos condições de regularidade para que ela tenha validade.

Entre os anos de 1953 e 1956, foram apresentados por Haldane [22], Haldane & Smith [23], vieses de ordem $O(n^{-1})$ para o caso biparamétrico. O caso geral para a expressão do viés de segunda ordem do EMV foi apresentado em 1968 por Cox & Snell [13]. Na Seção 2.5 demonstramos com detalhes essa expressão e assim como para a equação de Bartlett colocamos condições de regularidade para validar essa expressão. Essas condições de regularidade são generalizações, para o caso multiparamétrico, das condições constantes na Seção 2.4.

Em 1971, Box [4] calculou o viés de ordem $O(n^{-1})$ em modelos não-lineares multivariados com matriz de covariâncias conhecida. Em 1980, Pike et al. [39] utilizaram modelos lineares logísticos no estudo da correção de viés; Young & Bakir [54] estudaram modelo de regressão log-gama generalizado; Cordeiro & McCullagh [9] obtiveram o viés de segunda ordem para modelos lineares generali-

zados; Ferrari et al. [17] restringiram-se ao caso de modelos na família exponencial uniparamétrica e obtiveram vieses de segunda e terceira ordem para o estimador de máxima verossimilhança. Nesse artigo, eles encontram fórmulas fechadas para os coeficientes desses vieses e de uma expansão assintótica para a variância do EMV. Um estudo detalhado desse artigo é realizado na Seção 2.7.

Nos anos de 1997 a 2002 várias pesquisas foram desenvolvidas em correções de viés para diferentes tipos de modelos e com diversos focos, por exemplo: modelos de regressão não-lineares normais multivariados, Cordeiro & Vasconcellos [10]; modelos de regressão heterocedásticos e modelos SUR não-lineares, Vasconcellos & Cordeiro [46] e [47]; modelos de regressão não-linear com erros independentes e distribuídos como t de Student com número de graus de liberdade conhecido [7]; Ferrari & Cribari-Neto [16] utilizaram expansões de Edgeworth; modelos de regressão de dados circulares com distribuição von Mises foram usados por Cordeiro & Vasconcellos [11]; Cordeiro et al. [8] usaram modelos não-lineares de regressão em que as observações têm distribuições independentes e simétricas; Vasconcellos & Cordeiro [48] e Vasconcellos et al. [50] generalizam resultados de Cordeiro et al. [7] para modelos de regressão com distribuição t de Student multivariada com número de graus de liberdade conhecido; Cordeiro & Botter [6] basearam seus estudos em modelos lineares generalizados com superdispersão. No ano de 2002, Cribari-Neto & Vasconcellos [14] publicaram alternativas para correções de viés com dados provenientes de uma distribuição beta e, em 2005, Vasconcellos & Cribari-Neto [49] utilizaram modelos de regressão beta. Vasconcellos et al. [51] aplicaram correções de viés na modelagem de dados de imagens de radar. Ainda, em 2005, Vasconcellos & Silva [53] estudaram o caso em que os dados têm distribuição t de Student com número de graus de liberdade desconhecido. Ospina et al. [37] utilizaram o método de bootstrap na estimação pontual e intervalar em um modelo de regressão beta. A distribuição Birnbaum-Saunders foi utilizada nas pesquisas sobre correção de viés por Lemonte et al. [32]. Vasconcellos & Dourado [52] pesquisaram a correção de viés nos modelos da família exponencial biparamétrica. Recentemente, Barreto-Souza & Vasconcellos [1] utilizaram modelos de regressão nos quais as observações têm distribuição usada para modelar

valores extremos.

2.3 Conceitos Básicos

Considere o contexto em que temos uma amostra de variáveis aleatórias de tamanho n , cuja representação é $X = (X_1, \dots, X_n)$, com densidade de probabilidade conjunta $f(X|\theta)$, em que o parâmetro $\theta \in \mathbb{R}^p$ caracteriza a distribuição dentro da família \mathcal{F} , a qual contém a densidade $f(X|\theta)$.

Um dos objetivos da inferência estatística é estimar o parâmetro θ e dentre as técnicas mais utilizadas destacamos a proposta de Fisher no começo da década de 20, quando cursava o seu terceiro ano da graduação, ver Fisher [19]. A ideia desta técnica é escolher θ para o qual a amostra X é mais provável de ter acontecido. Em outras palavras, escolhe-se θ que maximiza a função de verossimilhança $f(\theta|X)$. Observe que esta função é a própria densidade de probabilidade fixada no ponto X e tendo como argumento o parâmetro θ . O estimador obtido a partir desta técnica é chamado de EMV $\hat{\theta}$.

É bastante comum a amostra X ter suas componentes X_1, \dots, X_n mutuamente independentes. Assim, a função de verossimilhança reduz-se à expressão

$$f(\theta|X) = \prod_{i=1}^n f(\theta|X_i). \quad (2.1)$$

O primeiro passo para maximizar esta função é derivá-la em relação ao parâmetro θ . A derivada de um produto pode ser uma expressão grande, então, uma alternativa é maximizar a função log-verossimilhança $L(\theta) = \log[f(\theta|X)]$, pois esta transformação torna o produto da equação (2.1) em soma de logaritmos, o que torna muito mais simples calcular a derivada. A partir daí, definimos a função escore como

$$L'(\theta) = \frac{\partial L(\theta)}{\partial \theta},$$

que é um vetor de tamanho p . Portanto, o EMV é a solução da equação $L'(\theta) = 0$.

A matriz $I(\theta) = E[L'(\theta)L'^T(\theta)]$, de dimensão $p \times p$, em que $L'^T(\theta)$ denota o vetor $L'(\theta)$ transposto, é chamada matriz de informação. Esta matriz é uma importante quantidade relacionada a um estimador não viésado, pois ela determina

o seu nível de eficiência. Um estimador é dito não viesado para um parâmetro θ , se seu valor esperado é igual a θ . No caso uniparamétrico, a eficiência de um estimador não viesado $\hat{\theta}$ para um parâmetro θ é definida com o quociente entre o limite inferior das variâncias dos estimadores não viesados para θ e $\text{Var}[\hat{\theta}]$, ver [3]. Assim, quando este limite coincide com a variância, temos um estimador eficiente. O teorema da desigualdade da informação garante que, sob certas condições de regularidade [20] e [21], a variância de qualquer estimador não viesado satisfaz a desigualdade

$$\text{Var}[\hat{\theta}] \geq I^{-1}(\theta).$$

Ou seja, a matriz de informação nos diz o quanto um estimador $\hat{\theta}$ está próximo da eficiência. Na estimação multiparamétrica, este teorema garante que a matriz $\text{Var}[\hat{\theta}] - I^{-1}(\theta)$ é positiva semi-definida.

No presente trabalho não iremos abordar esses tipos de estimadores, pois vamos explorar estimadores viesados e sempre com intuito de diminuir o viés desses estimadores. Mas, poderemos ver uma outra importante utilidade da matriz de informação, que é na construção de estimadores com vieses corrigidos, como, por exemplo, o estimador corrigido de Cox & Snell, visto na Seção 2.5.

Em geral, o cálculo do viés de um estimador envolve expressões complicadas, porém uma ferramenta que facilita esses cálculos é a função geradora de momentos $M_Y(t)$ de uma variável aleatória Y , que é definida como o valor esperado da função exponencial calculada no ponto tY . Assim, $M_Y(t) = \text{E}[\exp(tY)]$. Ver James [26], Magalhães [34].

Se a função geradora de momentos existe em uma vizinhança da origem, então todos os momentos da variável aleatória Y existem e podemos expandir esta função em uma série de potências, ver Lehmann & Casella [31]. Logo, podemos expressar a função geradora de momentos na forma

$$M_Y(t) = 1 + \sum_{j=1}^{\infty} \mu^{(j)} \frac{(t)^j}{j!}, \quad (2.2)$$

em que $\mu^{(j)} = \text{E}[Y^j]$ é o j -ésimo momento da variável Y . Com a função geradora de momentos $M_Y(t)$, definimos uma outra função importante, $\mathcal{K}_Y(t) =$

$\log [M_Y(t)]$, cujo desenvolvimento em série de potências é representado por

$$\mathcal{K}_Y(t) = \sum_{j=1}^{\infty} \kappa^{(j)} \frac{(t)^j}{j!}. \quad (2.3)$$

A função $\mathcal{K}_Y(t)$ é chamada de função geradora de cumulantes e os coeficientes $\kappa^{(j)}$ são denominados de cumulantes. A partir das equações definidas em (2.2) e (2.3), podemos expressar os cumulantes $\kappa^{(j)}$ em função dos momentos $\mu^{(j)}$. Por exemplo, por uma simples comparação entre as duas séries, temos que $\kappa^{(1)} = \mu^{(1)}$ e $\kappa^{(2)} = \mu^{(2)} - (\mu^{(1)})^2$, ver Lehmann & Casella [31], Cramér [15] e Stuart & Ord [44].

Se uma variável aleatória Y pertence à família exponencial p -paramétrica, então, sua densidade está na forma canônica da família exponencial se estiver escrita da seguinte forma:

$$f(Y|\theta) = \exp \left(\sum_{i=1}^p \theta_i T_i(Y) - A(\theta) \right),$$

em que $\theta = (\theta_1, \dots, \theta_p)^T$ caracteriza a densidade de probabilidade da variável. Neste caso, a função geradora de momentos e a função geradora de cumulantes são dadas por

$$\begin{aligned} M_Y(t) &= \frac{\exp(A(\theta + t))}{\exp(A(\theta))}, \\ \mathcal{K}_Y(t) &= A(\theta + t) - A(\theta). \end{aligned}$$

Outro conceito que iremos abordar ao longo do texto é o de ordem de magnitude. Dadas duas sequências de números reais a_n e b_n , dizemos que a_n tem ordem de magnitude menor que b_n , se $\frac{a_n}{b_n} \rightarrow 0$, quando $n \rightarrow \infty$ e usamos a notação $a_n = o(b_n)$. No caso em que a sequência $|\frac{a_n}{b_n}|$ é menor que uma constante fixa para todo n suficientemente grande, dizemos que a_n tem ordem de magnitude no máximo igual à de b_n e escrevemos $a_n = O(b_n)$.

Podemos estender o conceito de ordem de magnitude para o caso de sequências de variáveis aleatórias utilizando o conceito de convergência em probabilidade. Dizemos que uma sequência de variáveis aleatórias Y_n converge em probabilidade para uma variável aleatória Y , se $\forall \epsilon > 0$, temos que $\lim_{n \rightarrow \infty} P(|Y_n - Y| > \epsilon) = 0$. Indicamos isto por $Y_n \xrightarrow{P} Y$. Assim, dizemos que $Y_n = o_p(a_n)$, se $\frac{Y_n}{a_n} \xrightarrow{P} 0$. Ou

seja, se $\frac{Y_n}{a_n}$ converge em probabilidade para zero. Similarmente, $Y_n = O_p(a_n)$, se a sequência $|\frac{Y_n}{a_n}|$ é limitada em probabilidade para n suficientemente grande. Em outras palavras, $\forall \epsilon > 0$ existe uma constante M , tal que $P(|\frac{Y_n}{a_n}| < M) > 1 - \epsilon$, para n suficientemente grande.

Um tipo particular de sequências de números reais que utilizamos para medir a precisão de um estimador é a sequência do tipo n^k , para k pertencente ao conjunto dos inteiros.

As ordens de magnitudes gozam de várias propriedades; dentre elas, destacamos duas muito utilizadas ao longo do texto. São elas:

- P1: Se $Y_n = O_p(n^{k_1})$ e $Z_n = O_p(n^{k_2})$, então, $Y_n + Z_n = O_p(n^{\max\{k_1, k_2\}})$.
- P2: Se $Y_n = O_p(n^{k_1})$ e $Z_n = O_p(n^{k_2})$, então, $Y_n Z_n = O_p(n^{k_1+k_2})$.

Uma boa abordagem sobre este tema pode ser vista em Cordeiro [5].

2.4 Equações de Bartlett

Seguindo artigo de Vasconcellos [45], o objetivo desta seção é realizar um estudo aprofundado da expansão truncada,

$$\widehat{\theta} - \theta \sim \frac{L'(\theta)}{I(\theta)} + \left\{ \frac{L'(\theta) [L''(\theta) + I(\theta)]}{I(\theta)^2} + \frac{(L'(\theta))^2 L'''(\theta)}{2I(\theta)^3} \right\}, \quad (2.4)$$

em que $\widehat{\theta}$ é o EMV de θ . Da equação (2.4), temos

$$E[\widehat{\theta}] = \theta - \left\{ \frac{1}{2} E[L'''(\theta)] + \frac{\partial I(\theta)}{\partial \theta} \right\} / I(\theta)^2 + \dots \quad (2.5)$$

Estas equações foram apresentadas por Bartlett [2] em 1953, porém, não demonstradas. Assim, é aqui provado, sob certas condições de regularidade, que os termos da expansão $\widehat{\theta} - \theta$ definidos em (2.4) são, respectivamente, de ordens $O_p(n^{-\frac{1}{2}})$ e $O_p(n^{-1})$. Já o restante da expressão tem ordem $O_p(n^{-\frac{3}{2}})$. Também é demonstrado que o termo do viés da equação (2.5) tem ordem $O(n^{-1})$.

Considerando apenas o caso uniparamétrico, iremos supor que a amostra X tem suas componentes independentes e identicamente distribuídas. Portanto, a

função log-verossimilhança resume-se a $L(\theta) = \sum_{i=1}^n \ell_i(\theta)$, sendo que $\ell_i(\theta) = \log[f(\theta|X_i)]$. Por simplicidade do caso uniparamétrico, simplificaremos a notação das derivadas das funções com relação ao parâmetro: $f' = \frac{\partial}{\partial \theta} f(\theta|X)$, $G'' = \frac{\partial^2}{\partial \theta^2} G(\theta|X)$, ..., $L^{(k)} = \frac{\partial^k}{\partial \theta^k} L(\theta|X)$, etc.

Ao longo do texto iremos considerar, a princípio, as seguintes condições de regularidade:

Condições de Regularidade Uniparamétricas:

- (a): L tem suas derivadas contínuas.
- (b): L' tem exatamente uma raiz $\hat{\theta}$ e $L''(\hat{\theta}) < 0$.
- (c): L'' não se anula em uma vizinhança compacta de θ à qual pertence $\hat{\theta}$.
- (d): $E[L'] = 0$.
- (e): $E[(L')^2] = E[-L'''] < \infty$.
- (f): $I' = -E[L'''] - E[L''L']$.
- (g): $L' = O_p(n^{\frac{1}{2}})$.
- (h): $(L'' + I) = O_p(n^{\frac{1}{2}})$.
- (i): $(L''' - K) = O_p(n^{\frac{1}{2}})$, em que $K = E[L''']$ é finita.
- (j): As derivadas de segunda e terceira ordens ℓ'' e ℓ''' têm esperanças uniformemente limitadas, no sentido de que existem constantes positivas C_1, C_2, C_3 e C_4 , tais que $C_1 < |E[\ell''_i]| \leq E[|\ell''_i|] < C_2$, $E[|\ell'''_i|] < C_3$ e $E[|\ell'''_i|] < C_4$, para todo $i = 1, \dots, n$.
- (k): A função n/L'' é $O_p(1)$ em uma vizinhança de θ que inclui $\hat{\theta}$.
- (l): As funções L'' , L''' e $L^{(iv)}$ são $O_p(n)$ em uma vizinhança de θ que inclui $\hat{\theta}$.

As condições (a) a (c) garantem a existência de um máximo global e que L' é invertível na vizinhança de θ citada. Já a condição (d) é válida, por exemplo, se podemos derivar dentro da integral, ou seja,

$$\frac{\partial}{\partial \theta} \int L f dx = \int \frac{\partial}{\partial \theta} [L f] dx,$$

sendo $f = f(X|\theta)$. Assim, temos que $E[L'] = \int f' dx = \frac{\partial}{\partial \theta} \int f dx = 0$. Derivando mais uma vez dentro da integral temos a condição (e) satisfeita, pois,

$$\begin{aligned} E[L''] &= \int \frac{\partial}{\partial \theta} \left[\frac{f'}{f} \right] f dx = \int \left[\frac{f''}{f} - \frac{f'^2}{f^2} \right] f dx \\ &= \frac{\partial^2}{\partial \theta^2} \int f dx - E[L'^2] = -E[L'^2]. \end{aligned}$$

A condição (f) pode ser obtida se pudermos derivar dentro da integral na expressão da informação I , pois $I' = -\frac{\partial}{\partial \theta} \int L'' f dx = -\int L''' f dx - \int L'' f' dx = -K - E[L'L'']$. As condições (g) a (i) serão válidas se assumirmos, por exemplo, que os ℓ_i 's satisfazem algum teorema central do limite. Por exemplo,

$$\left(\frac{L^{(k)} - E[L^{(k)}]}{\sqrt{\text{Var}[L^{(k)}]}} \right) \xrightarrow{D} \mathcal{N}(0, 1).$$

Como $\text{Var}[L^{(k)}] = O(n)$, teremos que $L^{(k)} - E[L^{(k)}] = O_p(n^{\frac{1}{2}})$. A condição (j) é importante para garantir o comportamento do termo residual da expressão (2.5). As condições (g) e (l) garantem que produtos da forma $(L')^{e_1} (L'')^{e_2} (L''')^{e_3}$ têm ordens de magnitudes iguais a $O_p(n^{\frac{e_1}{2} + e_2 + e_3})$ e a condição (j) garante que os valores esperados desses produtos têm ordens de magnitudes iguais a $O(n^{\frac{e_1}{2} + e_2 + e_3})$. Esta condição também dá sentido a condição (k), pois teremos certeza de que a informação esperada é positiva. Então, como veremos, o comportamento do termo residual do viés será, o valor esperado de quocientes da forma $\frac{(L')^{e_1} (L'')^{e_2} (L''')^{e_3}}{I^{e_4}}$ e terá ordem de magnitude $O(n^{\frac{e_1}{2} + e_2 + e_3 - e_4})$. Por fim, as condições (k) e (l) são frequentemente satisfeitas, por exemplo, no caso em que a amostra é independente e identicamente distribuída ou no caso da amostra pertencer à família exponencial.

O teorema a seguir estabelece que a aproximação vista na equação (2.4) tem ordem $O_p(n^{-\frac{3}{2}})$. Na Seção 2.4, generalizamos esta expressão de forma que, dados $k > i > 0$, o i -ésimo termo da expressão tenha ordem de magnitude $O_p(n^{-\frac{i}{2}})$ e o termo residual seja $O_p(n^{-\frac{k+1}{2}})$.

Teorema 2.4.1 *Dadas as condições (a) a (l), temos que*

$$\widehat{\theta} - \theta = \frac{L'}{I} + \left\{ \frac{L' [L'' + I]}{I^2} + \frac{(L')^2 L'''}{2I^3} \right\} + O_p(n^{-3/2}),$$

em que o primeiro e segundo termos são de ordens $O_p(n^{-1/2})$ e $O_p(n^{-1})$, respectivamente.

Este teorema pode ser demonstrado utilizando a expansão de Taylor da função inversa de L' . Considere H tal que $H(L'(\theta)) = \theta$. Então,

$$\widehat{\theta} = \theta - \frac{L'}{L''} - \frac{(L')^2 L'''}{2(L'')^3} - \frac{(L')^3 [3(L'')^2 - L^{(iv)} L'']}{6(L'')^5} + r(\theta^*), \quad (2.6)$$

sendo $r(\theta^*) = \frac{\partial^{(4)}}{\partial L^{(4)}} H(L'(\theta^*)) L'(\theta^*)^4$ e θ^* tal que $|L'(\theta^*)| < |L'(\theta)|$. As condições (g) a (l) garantem que o penúltimo termo desta expansão é de ordem $O_p(n^{-3/2})$.

A validade das condições (c), (e), (h), (j) e (k) permite concluirmos que

$$\begin{aligned} \frac{1}{(L'')^5} &= \left(\frac{L'' + I}{L''} - \frac{1}{I} \right)^5 \\ &= O_p(n^{-\frac{13}{2}}) - \frac{10}{I^5} \left(\frac{L'' + I}{L''} \right)^2 + \frac{5}{I^5} \left(\frac{L'' + I}{L''} \right) - \frac{1}{I^5}. \end{aligned}$$

Substituindo esta igualdade na expressão (2.6) e encontrando o denominador comum desta expressão, temos que

$$\begin{aligned} \widehat{\theta} &= \theta + \frac{1}{6} \left[-6L'(L'')^4 - 3(L')^2(L'')^2L''' - (L')^3[3(L'')^2 - L''L^{(iv)}] \right] \\ &\quad \times \left[-\frac{10}{I^5} \left(\frac{L'' + I}{L''} \right)^2 + \frac{5}{I^5} \left(\frac{L'' + I}{L''} \right) - \frac{1}{I^5} + O_p(n^{-\frac{13}{2}}) \right] + O_p(n^{-2}). \end{aligned}$$

Com um pouco de álgebra temos que $(L'')^4 = I^4 - 4I^3(L'' + I) + 6I^2(L'' + I)^2 + O_p(n^{\frac{5}{2}})$; da mesma forma $(L'')^3 = -I^3 + 3I^2(L'' + I) + O_p(n^2)$; de maneira análoga $(L'')^2 = I^2 - 2I(L'' + I) + O_p(n)$, que também é igual $I^2 + O_p(n^{\frac{3}{2}})$ e por fim $L'' = -I + O_p(n^{\frac{1}{2}})$. Concluimos que a expressão (2.6) pode ser escrita da seguinte forma:

$$\begin{aligned} \widehat{\theta} &= \theta + \frac{L'}{I} + \left\{ \frac{L' (L'' + I)}{I^2} + \frac{(L')^2 L'''}{2I^3} \right\} + \frac{L' (L'' + I)^2}{I^3} + \\ &\quad + \frac{3(L')^2 L''' (L'' + I)}{2I^4} + \frac{(L')^3 [3(L'')^2 L'' - L'' L^{(iv)}]}{6I^5} + O_p(n^{-2}). \quad (2.7) \end{aligned}$$

O segundo termo desta expressão tem ordem $O_p(n^{-\frac{1}{2}})$, o terceiro e o quarto termo têm ordem $O_p(n^{-1})$, já do quinto termo até o sétimo a ordem de magnitude é

$O_p(n^{-\frac{3}{2}})$. Ou seja, esta é exatamente a expressão (2.4) apresentada por Bartlett, acrescida do termo $O_p(n^{-\frac{3}{2}})$.

O viés de segunda ordem de $\hat{\theta}$ é definido como o primeiro termo da expansão

$$E[\hat{\theta} - \theta] = -\frac{1}{I(\theta)^2} \left(\frac{K}{2} + I'(\theta) \right) + E[R_n],$$

em que $K = E[L'''(\theta)]$ e $R_n = O_p(n^{-\frac{3}{2}})$. Apesar da ordem de R_n , sob as condições (a) a (l), temos que $E[R_n] = O(n^{-2})$.

Dada a Afirmação 2.4.1 abaixo, temos que tanto este viés, como a ordem de magnitude de R_n , podem ser obtidos diretamente da equação (2.7).

Afirmação 2.4.1 *Seja P uma função tal que $\frac{\partial}{\partial \theta} E[P]$ e $E[P']$ são $O(n^k)$, então, $E[L'P] = O(n^k)$.*

Dem.

$$\begin{aligned} E[L'P] &= \int f' P dx = \int \frac{\partial}{\partial \theta} [Pf] dx - \int P' f dx \\ &= \frac{\partial}{\partial \theta} E[P] - E[P'] = O(n^k). \blacksquare \end{aligned}$$

Trivialmente temos que as esperanças do quinto ao sétimo termo da equação (2.7) têm ordem de magnitude $O(n^{-2})$, pois, dada a Afirmação 2.4.1, basta no quinto termo desta equação definir $P = (L'' + I)^2$, no sexto termo $P = L'L'''(L'' + I)$ e por fim no sétimo termo fazer $P = (L')^2[3(L''')^2 - L''L^{(iv)}]$.

Para obter o viés de segunda ordem do EMV observe que

$$\begin{aligned} E\left[\frac{(L')^2 L'''}{2I^3}\right] &= E\left[\frac{(L')^2 (L''' - K + K)}{2I^3}\right] \\ &= \frac{KE[(L')^2]}{2I^3} + E\left[\frac{(L')^2 (L''' - K)}{2I^3}\right] \quad (\text{Pela Afirmação 2.4.1}) \\ &= \frac{K}{2I^2} + O(n^{-2}). \end{aligned}$$

Como $E[L'L''] = -I' - K$, temos que o viés de segunda ordem do EMV pode ser obtido diretamente da equação de Bartlett:

$$E[\hat{\theta} - \theta] = E\left[\frac{L'}{I} + \frac{L'(L'' + I)}{I^2} + \frac{(L')^2 L'''}{2I^3}\right] + O(n^{-2})$$

$$\begin{aligned}
&= \frac{-I' - K}{I^2} + \frac{K}{2I^2} + O(n^{-2}) \\
&= -\frac{1}{I^2} \left(\frac{K}{2} + I' \right).
\end{aligned}$$

Em Vasconcellos [45] é possível ver a ordem de magnitude das esperanças das seguintes potências de $\hat{\theta} - \theta$:

- $E[(\hat{\theta} - \theta)^2] = I^{-1} + O(n^{-3/2})$,
- $E[(\hat{\theta} - \theta)^3] = O(n^{-3/2})$,
- $E[(\hat{\theta} - \theta)^4] = O(n^{-2})$.

Também é possível generalizar estes resultados para o caso de uma reparametrização $\eta = g(\theta)$ do parâmetro θ .

2.5 Resíduos de Cox & Snell

Na Seção 2.4, a equação (2.4), proposta por Bartlett [2], nos conduzia a um viés do EMV de ordem n^{-1} , porém, restrito ao caso em que temos apenas modelos uniparamétricos. O caso multiparamétrico foi tratado por Cox & Snell [13] em seu artigo “*A general definition of residuals (with discussion)*”, em que os autores definiram resíduos de ordem n^{-1} em modelos lineares. Nesta seção, detalhamos com precisão as fórmulas para o viés do estimador de máxima verossimilhança com ordem de magnitude $O(n^{-1})$ para o caso multiparamétrico, proposto por Cox & Snell [13]. Este será chamado de viés de Cox & Snell.

Seguindo a ideia desenvolvida por Vasconcellos em [45] para demonstrar essas fórmulas, considere no caso uniparamétrico a expansão de Taylor da função escore calculada no ponto $\hat{\theta}$ e em torno desta função calculada no ponto θ , ou seja,

$$L'(\hat{\theta}) = L'(\theta) + L''(\theta)(\hat{\theta} - \theta) + \frac{1}{2}L'''(\theta^*)(\hat{\theta} - \theta^*)^2, \quad (2.8)$$

sendo que θ^* pertence ao intervalo formado por $\hat{\theta}$ e θ . Dadas as condições de regularidades vista na Seção 2.4 e supondo que $\hat{\theta} - \theta$ tem ordem de magnitude

$O_p(n^{-\frac{1}{2}})$, podemos observar que

$$\begin{aligned} -L' &= (L'' + I - I)(\hat{\theta} - \theta) + O_p(1) \\ &= -I(\hat{\theta} - \theta) + O_p(1). \end{aligned}$$

Portanto,

$$\hat{\theta} - \theta = \frac{L'}{I} + O_p(n^{-1}), \quad (2.9)$$

$$(\hat{\theta} - \theta)^2 = \frac{(L')^2}{I^2} + O_p(n^{-\frac{3}{2}}). \quad (2.10)$$

Da definição de covariância de duas variáveis aleatórias, temos que $E[L''(\hat{\theta} - \theta)] = -IE[\hat{\theta} - \theta] + \text{Cov}[L'', \hat{\theta} - \theta]$. Substituindo o resultado da equação (2.9) na expressão da covariância é fácil observar que $\text{Cov}[L'', \hat{\theta} - \theta] = \frac{1}{I}E[L'L''] + \text{Cov}[L'', O_p(n^{-1})]$. Seja $L^{(k)}$ a k -ésima derivada da função log-verossimilhança e suponha que $L^{(k)} - E[L^{(k)}] = O_p(n^{\frac{1}{2}})$. Então, podemos supor que a ordem de magnitude da covariância entre $L^{(k)}$ e os termos de ordem $O_p(n^{-1})$ da expansão de Taylor da função escore L' é $O(n^{-\frac{1}{2}})$, pois,

$$\begin{aligned} \text{Cov}[L^{(k)}, O_p(n^{-1})] &= \text{Cov}[L^{(k)} - E[L^{(k)}] + E[L^{(k)}], O_p(n^{-1})] \\ &= \text{Cov}[L^{(k)} - E[L^{(k)}], O_p(n^{-1})] + \text{Cov}[E[L^{(k)}], O_p(n^{-1})] \\ &= E[(L^{(k)} - E[L^{(k)}])O_p(n^{-1})] \\ &= O(n^{-\frac{1}{2}}). \end{aligned}$$

Dessa forma temos a seguinte igualdade:

$$E[L''(\hat{\theta} - \theta)] = -IE[\hat{\theta} - \theta] + \frac{1}{I}E[L'L''] + O(n^{-\frac{1}{2}}). \quad (2.11)$$

Com raciocínio análogo e usando a equação (2.10) chegamos, também, à seguinte expressão:

$$\begin{aligned} E\left[L'''(\hat{\theta} - \theta)^2\right] &= E[L'''] E[(\hat{\theta} - \theta)^2] + \text{Cov}\left[L''', (\hat{\theta} - \theta)^2\right] \\ &= E[L'''] E[(\hat{\theta} - \theta)^2] + O(n^{-\frac{1}{2}}) \\ &= \frac{K}{I} + O(n^{-\frac{1}{2}}), \end{aligned} \quad (2.12)$$

sendo $K = E[L''']$.

O quarto termo da expansão de Taylor da equação (2.8) é $\frac{1}{6}L^{(iv)}(\hat{\theta} - \theta)^3$, que tem ordem de magnitude $O_p(n^{-\frac{1}{2}})$. Assim, levando em consideração que a função escore é zero no ponto $\hat{\theta}$ e que seu valor esperado em qualquer ponto θ também é zero, temos que o valor esperado desta expansão é

$$0 = E[L''(\hat{\theta} - \theta)] + \frac{1}{2}E[L'''(\hat{\theta} - \theta)^2] + O(n^{-\frac{1}{2}}).$$

Então, com um pequeno desenvolvimento algébrico e dadas as equações (2.11) e (2.12), chegamos à expressão

$$\begin{aligned} E[\hat{\theta} - \theta] &= \frac{E[L'L'']}{I^2} + \frac{K}{2I^2} + O(n^{-\frac{3}{2}}) \\ &= \frac{-K - I'}{I^2} + \frac{K}{2I^2} + O(n^{-\frac{3}{2}}) \\ &= -\frac{1}{I^2} \left(\frac{K}{2} + I' \right) + O(n^{-\frac{3}{2}}). \end{aligned}$$

Esta é exatamente a expressão do viés de segunda ordem do EMV no caso uniparamétrico vista na Seção 2.4.

Consideraremos agora o caso multiparamétrico, ou seja, o parâmetro caracterizador da distribuição tem a forma $\theta = (\theta_1, \dots, \theta_p)^T$. Deste modo a função escore é o vetor gradiente da função log-verossimilhança. Utilizaremos neste caso multiparamétrico a notação introduzida em Lawley [30]. Assim, as derivadas parciais da log-verossimilhança são representadas por $U_r = \partial L / \partial \theta_r$, $U_{rs} = \partial^2 L / \partial \theta_r \partial \theta_s$, $U_{rst} = \partial^3 L / \partial \theta_r \partial \theta_s \partial \theta_t, \dots$. Os momentos são representados por κ , assim, $\kappa_{rs} = E[U_r U_s]$, $\kappa_{rs} = E[U_{rs}]$, $\kappa_{rs,t} = E[U_{rs} U_t], \dots$. E as derivadas dos momentos são denotadas por $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \theta_t$. Uma representação muito utilizada na literatura é $\kappa^{r,s}$, que significa o elemento de posição (r, s) na matriz inversa I^{-1} , sendo $I = E[L'L'^T]$.

Para o caso multiparamétrico definimos as seguintes matrizes, $L'' = [\partial L'' / \partial \theta_1, \dots, \partial L'' / \partial \theta_p], \dots, L^{(k)} = [\partial L^{(k-1)} / \partial \theta_1, \dots, \partial L^{(k-1)} / \partial \theta_p]$.

Suponha as seguintes condições de regularidades para o caso multiparamétrico,

Condições de Regularidade Multiparamétricas:

- (a): As derivadas parciais de L são contínuas.

- (b): L' tem exatamente uma raiz $\hat{\theta}$ e a matriz $L''(\hat{\theta})$ é negativa definida.
- (c): $E[L'] = 0$.
- (d): $E[L'L'^T] = E[-L'']$ e tem todos os elementos finitos.
- (e): $I' = -E[L'''] - E[L''L']$.
- (f): $L' = O_p(n^{\frac{1}{2}})$.
- (g): Qualquer derivada parcial menos seu valor esperado é $O_p(n^{\frac{1}{2}})$. Ou seja, $U_{rst\dots} - \kappa_{rst\dots} = O_p(n^{\frac{1}{2}})$.
- (h): Se uma função polinomial das derivadas parciais $f(U_r U_{rs} U_{rst} \dots)$ tem ordem de magnitude $O_p(n^k)$, então, $E[f(U_r U_{rs} U_{rst} \dots)] = O(n^k)$.
- (i): Os elementos da matriz $(L'')^{-1}$ são $O_p(n^{-1})$ em uma vizinhança de θ que inclui $\hat{\theta}$.
- (j): Os elementos das matrizes L'' , L''' e $L^{(iv)}$ são $O_p(n)$ em uma vizinhança de θ que inclui $\hat{\theta}$.
- (k): O EMV satisfaz as igualdades $\hat{\theta} - \theta = O_p(n^{-\frac{1}{2}})$ e $E[\hat{\theta} - \theta] = O(n^{-1})$.

Estas condições de regularidades são generalizações das condições dadas na Seção 2.4. As condições (a) a (d) garantem a existência do EMV e a invertibilidade da função L' . É bastante comum nos modelos estatísticos podermos passar a derivada para dentro do integrando, assim, as condições (d) e (e) são na maioria das vezes satisfeitas. A condição (g), (i) e (j) na grande parte dos modelos também são satisfeitas, por exemplo, nas amostras independentes, indenticamente distribuídas e com momentos finitos. Já a condição (h) é importante para garantir que a ordem de magnitude do valor esperado dos termos da expansão de Taylor continue a mesma, no sentido que $E[O_p(n^{-k})] = O(n^{-k})$. Por fim, a condição (k) garante que estamos trabalhando com estimadores pelo menos assintoticamente não viesado. Assim como vimos naquela seção, essas suposições são comuns de acontecer, ou seja, não há uma restrição muito forte nas classes dos modelos estatístico que satisfazem essas condições.

A demonstração do viés de Cox & Snell para o caso p -dimensional é apenas uma generalização do que foi visto no modelo uniparamétrico.

O produto de Kronecker [25] e [35], representado pelo símbolo \otimes , é definido da seguinte forma, se A e B são matrizes de dimensões $m_1 \times n_1$ e $m_2 \times n_2$, respectivamente, então,

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n_1}B \\ \vdots & \ddots & \vdots \\ \vdots & \dots & \vdots \\ \vdots & \dots & \vdots \\ a_{m_11}B & \dots & a_{m_1n_1}B \end{bmatrix}.$$

Observe que a matriz $A \otimes B$ tem dimensão $m_1m_2 \times n_1n_2$. Em particular, temos que

$$(\widehat{\theta} - \theta) \otimes (\widehat{\theta} - \theta) = \begin{bmatrix} (\widehat{\theta}_1 - \theta_1)(\widehat{\theta}_1 - \theta_1) \\ \vdots \\ (\widehat{\theta}_p - \theta_p)(\widehat{\theta}_1 - \theta_1) \\ \vdots \\ (\widehat{\theta}_1 - \theta_1)(\widehat{\theta}_p - \theta_p) \\ \vdots \\ (\widehat{\theta}_p - \theta_p)(\widehat{\theta}_p - \theta_p) \end{bmatrix}.$$

A expansão de Taylor da função escore calculada no EMV e em torno do ponto θ é dada por

$$L'(\widehat{\theta}) = L' + L''(\widehat{\theta} - \theta) + \frac{1}{2}L'''((\widehat{\theta} - \theta) \otimes (\widehat{\theta} - \theta)) + r(\theta^*), \quad (2.13)$$

sendo que θ^* está no retângulo p -dimensional definido por θ e $\widehat{\theta}$. Dadas as condições de regularidades multiparamétrica, temos que o termo $r(\theta^*)$ tem ordem $O_p(n^{-\frac{1}{2}})$.

Podemos esperar dessa espansão que $L' + (L'' + I - I)(\widehat{\theta} - \theta) = O_p(1)$ e, portanto, $\widehat{\theta} - \theta = I^{-1}L' + O_p(n^{-1})$. Assim, o t -ézimo termo deste vetor é dado por

$$\widehat{\theta}_t - \theta_t = \sum_{u=1}^p \kappa^{t,u} U_u + O_p(n^{-1}).$$

E do produto $(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^T$ concluímos que

$$\mathbb{E} \left[(\widehat{\theta}_t - \theta_t)(\widehat{\theta}_u - \theta_u) \right] = \kappa^{t,u} + O(n^{-\frac{3}{2}}). \quad (2.14)$$

Note ainda que $\text{Cov}[U_{rts\dots}, O_p(n^{-1})] = \text{Cov}[U_{rtu\dots} - \kappa_{rtu\dots}, O_p(n^{-1})]$. Em particular, temos que $\text{Cov}[U_{rt}, O_p(n^{-1})] = O(n^{-\frac{1}{2}})$ e $\text{Cov}[U_{rtu}, (\widehat{\theta}_u - \theta_u)^2] = O(n^{-\frac{1}{2}})$.

Assim, temos que

$$\begin{aligned} \text{Cov}(U_{rt}, \widehat{\theta}_t - \theta_t) &= \text{Cov} \left[U_{rt}, \sum_{u=1}^p \kappa^{t,u} U_u + O_p(n^{-1}) \right] \\ &= \sum_{u=1}^p \kappa^{t,u} \text{Cov}[U_{rt}, U_u] + O(n^{-\frac{1}{2}}) \\ &= \sum_{u=1}^p \kappa^{t,u} \kappa_{rt,u} + O(n^{-\frac{1}{2}}). \end{aligned} \quad (2.15)$$

Dada a expansão de Taylor na equação (2.13), observamos que a r -ésima coordenada do vetor gradiente $L'(\widehat{\theta})$ é dada por

$$0 = U_r + \sum_{s=1}^p U_{rs}(\widehat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{t=1}^p \sum_{u=1}^p U_{rtu}(\widehat{\theta}_t - \theta_t)(\widehat{\theta}_u - \theta_u) + O_p(n^{-\frac{1}{2}}).$$

Usando o fato de que as covariâncias vista acima são $O(n^{-\frac{1}{2}})$ e usando as igualdades vista nas equações (2.14) e (2.15) temos que o valor esperado desta última expressão é dado por,

$$\sum_{s=1}^p \kappa_{rs} \mathbb{E} \left[\widehat{\theta}_s - \theta_s \right] + \sum_{t=1}^p \sum_{u=1}^p \kappa^{t,u} \kappa_{rt,u} + \frac{1}{2} \sum_{t=1}^p \sum_{u=1}^p \kappa_{rtu} \kappa^{t,u} = O(n^{-\frac{1}{2}}).$$

Este resultado foi encontrado para r -ésima linha do vetor expandido na equação de Taylor, então, usando o fato que $\kappa_{rs} = -\kappa_{r,s}$, podemos representar esta igualdade

usando notação matricial,

$$IE \left[\widehat{\theta} - \theta \right] = \begin{bmatrix} \sum_{t=1}^p \sum_{u=1}^p \kappa^{t,u} (\kappa_{1t,u} + \frac{1}{2} \kappa_{1tu}) + O(n^{-\frac{1}{2}}) \\ \vdots \\ \vdots \\ \sum_{t=1}^p \sum_{u=1}^p \kappa^{t,u} (\kappa_{pt,u} + \frac{1}{2} \kappa_{ptu}) + O(n^{-\frac{1}{2}}) \end{bmatrix}.$$

Multiplicando o lado esquerdo desta igualdade pela matriz I^{-1} é fácil verificar que a r -ésima linha deste produto é dada por

$$E \left[\widehat{\theta}_r - \theta_r \right] = \sum_{s=1}^p \kappa^{r,s} \sum_{t=1}^p \sum_{u=1}^p \kappa^{t,u} \left(\kappa_{st,u} + \frac{1}{2} \kappa_{stu} \right) + O(n^{-\frac{3}{2}}). \quad (2.16)$$

Este igualdade é extamente o viés de Cox & Snell.

2.6 Correção de Viés Preventiva

A ideia central desta seção é mostrar uma técnica proposta por Firth [18], na qual se pode remover o termo de primeira ordem do viés de $\widehat{\theta}$ em modelos paramétricos a partir da função escore. Para este intuito é feita uma modificação na função escore de forma que o estimador obtido igualando a zero esta nova função reduza o viés.

A técnica proposta por Firth [18] considera estimação multiparamétrica e consiste em fazer a seguinte modificação na função escore:

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta),$$

em que A_r pode depender dos dados e é $O_p(1)$.

Seja θ^* , tal que, $U^*(\theta^*) = 0$, então, utilizando uma forma fechada para $\widehat{\gamma} = n^{1/2}(\theta^* - \theta)$ vista em McCullagh [33], baseada na expansão de Taylor de $U_r^*(\theta)$ em torno de θ , temos que o viés de θ^* é dado por

$$E[n^{-1/2} \widehat{\gamma}^r] = n^{-1} \kappa^{r,s} \left\{ -\kappa^{t,u} (\kappa_{s,t,u} + \kappa_{s,tu}) / 2 + \alpha_s \right\} + O(n^{-\frac{3}{2}}), \quad (2.17)$$

em que α_s é a esperança de A_s .

Esta ideia é base do desenvolvimento da presente tese, pois, no Capítulo 3 iremos generalizar este raciocínio para encontrar o viés com ordem de magnitude $O(n^{-k})$, para qualquer $k = 3, 4, \dots$. Em princípio, esta generalização se dará apenas no caso uniparamétrico, em seguida apresentamos a ideia para espaços biparamétricos.

Um candidato natural para remover o viés de segunda ordem é escolher

$$\alpha_s = \kappa_{r,s} \kappa^{r,s} \left\{ \kappa^{t,u} (\kappa_{s,t,u} + \kappa_{s,tu}) / 2 \right\}.$$

Assim, observamos que para remover o viés de segunda ordem basta escolher A_r , tal que $E[A_r] = -I(\theta)b_1(\theta)/n + O(n^{-1})$, sendo b_1 o termo de primeira ordem do viés de Cox & Snell, visto na equação (2.16).

No caso uniparamétrico em que a amostra tem uma distribuição exponencial com parâmetro θ , ou seja, $X \sim \exp(\theta)$, temos que $\kappa_{s,tu} = 0$ e $\kappa_{s,t,u} = \kappa_{stu} = n^{-1}E[L'''(\theta)]$. Logo, se tomarmos

$$\alpha = -\frac{n^{-1}E[L'''(\theta)]}{n^{-1}2I(\theta)},$$

teremos que o viés da equação (2.17) tem ordem de magnitude $O(n^{-\frac{3}{2}})$. Ou seja, a função escore modificada é dada por

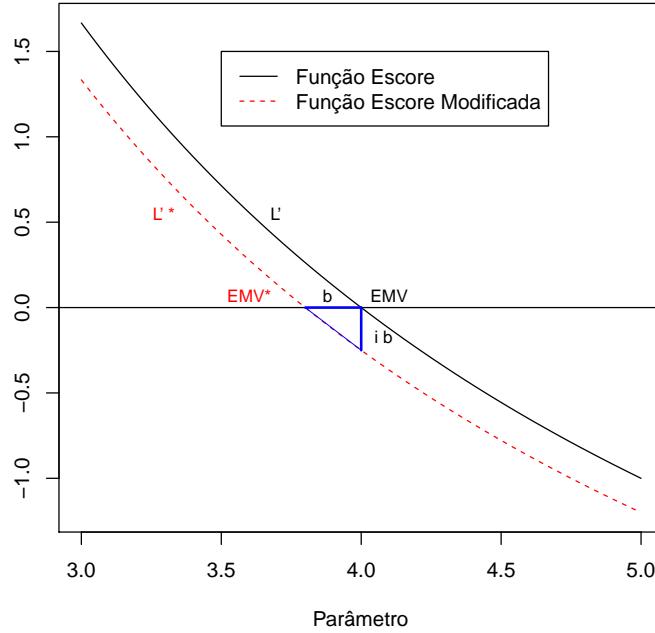
$$L'^*(\theta) = L'(\theta) - \frac{E[L'''(\theta)]}{2I(\theta)},$$

que na verdade é apenas uma translação da função escore $L'(\theta)$. Neste caso, podemos encontrar o estimador $\hat{\theta}^*$ proposto por Firth [18] igualando a zero esta nova função escore L'^* . Assim, temos

$$\hat{\theta}^* = \frac{n-1}{\sum_{i=1}^n x_i} = \hat{\theta} \left(1 - \frac{1}{n} \right).$$

Este novo estimador, também, é viesado, porém tem viés com ordem de magnitude menor.

O gráfico da Figura 2.1, que é uma adaptação do gráfico do artigo de Firth [18], representa as funções escore $L'(\theta)$ e sua modificação $L'^*(\theta)$, para o caso particular de uma amostra de tamanho $n = 20$ e $\sum_{i=1}^{20} x_i = 5$. A reta tangente à curva $L'(\theta)$ no ponto $\hat{\theta}$ tem como coeficiente angular a derivada da função escore

Figura 2.1. Funções escore $L'(\theta)$ e escore modificada $L'^*(\theta)$.

calculada neste ponto. Então, o valor esperado deste coeficiente é o simétrico da informação de Fisher estimada $I(\hat{\theta})$. Assim, o triângulo retângulo formado pela reta, que além de passar pelo ponto $(\theta^*, 0)$, tem como coeficiente angular $I(\hat{\theta})$, e o segmento b do eixo das abscissas delimitado pelas duas curvas tem o terceiro lado de tamanho aproximado ib , com $i = I(\hat{\theta})$ e $b = b(\hat{\theta})$. Portanto, a ideia é que deslocando-se a função escore desta forma, o ponto zero dela estará mais próximo do verdadeiro valor do parâmetro, encontrando-se assim um estimador θ^* cujo viés é menor que o viés de $\hat{\theta}$.

Um caso bastante interessante observado por David Firth acontece na família exponencial. Se θ é o parâmetro canônico da família exponencial, então, o estimador corrigido coincide com o estimador obtido pela função de verossimilhança penalizada, que pode ser vista como a priori invariante de Jeffreys [28], [27] e [38], já que a função a ser maximizada é proporcional à $|I(\theta)|^{\frac{1}{2}}$.

Na família exponencial canônica temos que $\kappa_{r,st} = 0$, para todos r, s, t , por-

tanto, podemos considerar

$$A = \frac{1}{2} \text{tr} \left\{ I(\theta)^{-1} \left(\frac{\partial I(\theta)}{\partial \theta} \right) \right\} = \frac{1}{2} \frac{\partial}{\partial \theta} \{ \log [I(\theta)] \}.$$

Assim, a função a ser maximizada é

$$L^*(\theta) = L(\theta) + \frac{1}{2} \log [I(\theta)].$$

A distribuição normal, com média μ e variância σ^2 desconhecidas, é um exemplo clássico apresentado no artigo de David Firth. Suponha que o interesse esteja em estimar o parâmetro canônico desta distribuição, ou seja, estimar

$$\theta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right).$$

Neste caso é fácil verificar que a matriz de informação é

$$I(\theta) = \begin{bmatrix} n\sigma^2 & 2n\mu\sigma^2 \\ n\mu\sigma^2 & n(4\mu^2\sigma^2 + 2\sigma^4) \end{bmatrix}.$$

Como o determinante da matriz de informação é $\log[2n^2\sigma^6]$, a função log-verossimilhança a ser maximizada é

$$L^*(\theta) = -\frac{n}{2} \log[2\pi\sigma^2] - \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} + \frac{1}{2} \log [2n^2\sigma^6].$$

Sejam $\theta_1 = \frac{\mu}{\sigma^2}$ e $\theta_2 = -\frac{1}{2\sigma^2}$. Então, derivando esta função em relação ao parâmetro canônico temos o vetor gradiente

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i + \frac{n\alpha}{2\theta_2} \\ -\frac{n}{2\theta_2} + \sum_{i=1}^n x_i^2 - \frac{n\theta_1^2}{4\theta_2^2} + \frac{3}{2\theta_2} \end{bmatrix}.$$

Igualando a zero este vetor e resolvendo o sistema de equações, chegamos ao estimador de máxima verossimilhança corrigido (EMVC):

$$\theta^* = \left(\frac{(n-3)\bar{x}}{S_x^2}, -\frac{n-3}{2S_x^2} \right),$$

sendo $S_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$. Dado que $\frac{S_x^2}{\sigma^2} \sim \chi_{(n-1)}^2$, então,

$$E[S_x^r] = \frac{\sigma^r 2^{\frac{r}{2}} \Gamma\left(\frac{n-1+r}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

para $r > 1 - n$. Assim, fazendo-se $r = -2$, temos que $E\left[-\frac{n-3}{2S_x^2}\right] = \theta_2$. O teorema de Basu garante que \bar{x} e S_x^2 são independentes na distribuição normal [24] e [36]. Logo,

$$E\left[\frac{(n-3)\bar{x}}{S_x^2}\right] = E\left[\frac{n-3}{S_x^2}\right]E[\bar{x}] = \frac{\mu}{\sigma^2} = \theta_2.$$

Portanto, o estimador θ^* é não viesado para o parâmetro θ .

Observe que esta técnica garante que o viés do estimador é pelo menos de ordem de magnitude $O(n^{-\frac{3}{2}})$, mas, nada impede que este viés tenha uma ordem de magnitude menor ainda, como neste exemplo em que ele é não viesado, para $n > 3$.

2.7 Uma Redução de Viés de Terceira Ordem para a Família Exponencial Uniparamétrica

Vimos na Seção 2.6 uma correção de viés preventiva, em que fazemos uma translação na função escore, de tal forma que ao encontrar o zero desta função transladada já obtemos um estimador de máxima verossimilhança com viés corrigido. Nesta seção, iremos discutir um estimador de máxima verossimilhança, cujo viés tem ordem de magnitude $O(n^{-3})$. Este estimador foi proposto por Ferrari, Botter, Cordeiro & Cribari-Neto [17], apenas para o caso do modelo pertencer à família exponencial uniparamétrica, utilizando o viés de Cox & Snell estimado. Nesta técnica, estima-se o viés e depois corrige o estimador, diferentemente da técnica de estimação preventiva.

Considere que a amostra x_1, \dots, x_n pertence a família exponencial uniparamétrica, então, baseado nas expansões do viés e da variância do EMV até a segunda ordem, $B(\theta) = \frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2} + O(n^{-3})$ e $V(\theta) = \frac{V_1(\theta)}{n} + \frac{V_2(\theta)}{n^2} + O(n^{-3})$, sendo $B(\theta) = E\left[\hat{\theta} - \theta\right]$, são definidos três estimadores:

$$\begin{aligned}\hat{\theta}_1 &= \hat{\theta} - \frac{B_1(\hat{\theta})}{n}, \\ \hat{\theta}_2 &= \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2(\hat{\theta})}{n^2},\end{aligned}\tag{2.18}$$

$$\tilde{\theta}_2 = \hat{\theta} - \frac{B(\hat{\theta})}{n} - \frac{B^*(\hat{\theta})}{n^2}, \quad (2.19)$$

em que $B_2^*(\theta) = B_2(\theta) - B_1(\theta)B_1'(\theta) - \frac{1}{2}B_1''(\theta)V_1(\theta)$.

Se seguíssemos a intuição, comentaríamos um erro em afirmar que o estimador $\hat{\theta}_2$ tem viés de ordem de magnitude $O(n^{-2})$. É demonstrado que tanto $\hat{\theta}_1$, como $\hat{\theta}_2$, são $O(n^{-1})$ e que apenas o estimador $\tilde{\theta}_2$ é não viesado até ordem $O(n^{-2})$.

Para estabelecer as ordens de magnitudes de $\hat{\theta}_1$, $\hat{\theta}_2$ e $\tilde{\theta}_2$ consideramos a expansão de Taylor de da função $H(\theta) = \theta - \frac{B_1(\theta)}{n}$ em torno de θ ,

$$\hat{\theta}_1 - \frac{B_1(\hat{\theta}_1)}{n} = \theta - \frac{B_1(\theta)}{n} + \left(1 - \frac{B_1'(\theta)}{n}\right)(\hat{\theta}_1 - \theta) + \frac{1}{2} \frac{B_1''(\theta)}{n}(\hat{\theta}_1 - \theta)^2$$

Como $B(\hat{\theta}) = 0$ e considerando que $\frac{B_1(\theta)}{n} = B(\theta) - \frac{B_2(\theta)}{n^2} + O(n^{-3})$, temos que

$$E[\hat{\theta}_1 - \theta] = \frac{B_2^*(\theta)}{n^2} + O(n^{-3}).$$

Se fizermos a expansão de Taylor em torno de θ da função $H(\theta) = \theta - \frac{B_1(\theta)}{n} - \frac{B_2(\theta)}{n^2}$, chegamos à conclusão que

$$E[\hat{\theta}_2 - \theta] = \frac{B_2^*(\theta) - B_2(\theta)}{n^2} + O(n^{-3}).$$

Similarmente, fazendo-se a expansão da função $H(\theta) = \theta - \frac{B_1(\theta)}{n} - \frac{B_2^*(\theta)}{n^2}$ e calculando o valor esperado temos que

$$\begin{aligned} E[\tilde{\theta}_2 - \theta] &= \frac{B_2(\theta)}{n^2} - \frac{B_2^*(\theta)}{n^2} + \left(1 - \frac{B_1'(\theta)}{n} - \frac{B_2^{**}(\theta)}{n^2}\right) \left(\frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2}\right) \\ &+ \frac{1}{2} \left(\frac{B_1''(\theta)}{n} - \frac{B_2^{***}(\theta)}{n^2}\right) \left(\frac{V_1(\theta)}{n} + \frac{V_2(\theta)}{n^2}\right) + O(n^{-3}) \\ &= O(n^{-3}). \end{aligned}$$

Ou seja, apenas o viés do estimador $\tilde{\theta}_2$ tem ordem de magnitude $O(n^{-3})$.

A partir de resultados de Shenton & Bowman (1977, pp.44-47) [43] é possível determinar expressões fechadas para estes vieses, a saber:

$$\begin{aligned} B_1(\theta) &= -\frac{1}{2\kappa_{\theta\theta}^2} \left(\kappa_{\theta\theta\theta} - 2\kappa_{\theta\theta}^{(\theta)}\right), \\ B_2(\theta) &= -\frac{1}{8\kappa_{\theta\theta}^3} (\kappa_{\theta\theta\theta\theta\theta} + 12\kappa_{\theta\theta,\theta\theta\theta} + 4\kappa_{\theta,\theta\theta\theta\theta} + 8\kappa_{\theta,\theta\theta,\theta\theta} + 4\kappa_{\theta,\theta,\theta\theta\theta}) \\ &+ \frac{1}{12\kappa_{\theta\theta}^4} \left(13\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta\theta} - 18\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta\theta} - 18\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta}^{(\theta)}\right) \end{aligned}$$

$$\begin{aligned}
& - 36\kappa_{\theta\theta\theta}\kappa_{\theta\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta}^{(\theta)} + 18\kappa_{\theta\theta\theta}\kappa_{\theta\theta}^{(\theta)} \Big) - \frac{1}{8\kappa_{\theta\theta}^5} \left[11\kappa_{\theta\theta\theta}^3 \right. \\
& \left. - 48\kappa_{\theta\theta\theta}^2\kappa_{\theta\theta}^{(\theta)} + 48\kappa_{\theta\theta\theta}(\kappa_{\theta\theta}^{(\theta)})^2 \right], \\
V_1(\theta) & = -\frac{1}{\kappa_{\theta\theta}}, \\
B_2^*(\theta) & = B_2(\theta) - B_1(\theta)B_1'(\theta) - \frac{1}{2}B_1''(\theta)V_1(\theta).
\end{aligned}$$

Um caso particular ocorre quando a matriz de informação não depende mais dos dados, por exemplo na família exponencial canônica, então temos que: $\kappa_{\theta,\theta\theta\dots\theta} = 0$ e $\kappa_{\theta\theta}^{(\theta)} = \kappa_{\theta\theta\theta}$. Logo, as expressões acima reduz-se a

$$\begin{aligned}
B_1(\theta) & = \frac{L'''}{2(L'')^2}, \\
B_1'(\theta) & = \frac{L^{iv}}{2(L'')^2} - \frac{(L''')^2}{(L'')^3}, \\
B_1''(\theta) & = \frac{L^v}{2(L'')^2} - \frac{3L''''L^{iv}}{(L'')^3} + \frac{3(L''')^3}{(L'')^4}, \\
B_2(\theta) & = -\frac{(L^v + 12L''L''')}{8(L'')^3} + \frac{(13L^{iv}L'''' + 18(L'')^2L''')}{12(L'')^4} - \frac{11(L''')^3}{8(L'')^5}, \\
V_1(\theta) & = -\frac{1}{L''}.
\end{aligned}$$

Considere agora uma amostra $x = (x_1, \dots, x_n)$, independente e identicamente distribuída segundo uma Poisson com parâmetro λ . Ou seja, $x_i \sim \mathcal{P}(\lambda)$, para todo $i = 1, \dots, n$. Suponha que estamos interessados em estimar o parâmetro canônico $\theta = \log(\lambda)$. A função de verossimilhança e a função de log-verossimilhança são,

$$\begin{aligned}
f(x|\lambda) & = \frac{\lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda)}{\prod_{i=1}^n x_i!}, \\
f(x|\theta) & = \frac{\exp(\theta \sum_{i=1}^n x_i) \exp[-n \exp(\theta)]}{\prod_{i=1}^n x_i!}, \\
L(\theta) & = \theta \sum_{i=1}^n x_i - n \exp(\theta) - \log \left(\prod_{i=1}^n x_i! \right).
\end{aligned}$$

A função escore $L'(\theta)$ e suas derivadas são dadas, respectivamente, por

$$L'(\theta) = \sum_{i=1}^n x_i - n \exp(\theta),$$

$$L^k(\theta) = -n \exp(\theta), \forall k \geq 2.$$

Portanto, a informação de Fisher é $I(\theta) = n \exp(\theta)$.

Sabemos que o EMV de θ é obtido por igualar a zero a função escore, logo é dado por,

$$\hat{\theta} = \log(\bar{x}).$$

Já o estimador com viés corrigido preventivamente, proposto por Firth [18], é obtido por transladar a função escore $L'(\theta)$ com a matriz A_0 , sendo $A_0 = -\frac{L'''}{2I}$. Assim,

$$\theta_0^* = \log\left(\bar{x} + \frac{1}{2n}\right).$$

Para encontrarmos o estimador proposto por Ferrari et al. [17], precisamos calcular

$$B_1(\hat{\theta}) = \frac{L'''}{2(L'')^2} = -\frac{1}{2n\bar{x}},$$

$$B'_1(\hat{\theta}) = \frac{L^{iv}}{2(L'')^2} - \frac{(L''')^2}{(L'')^3} = \frac{1}{2n\bar{x}},$$

$$B''_1(\hat{\theta}) = \frac{L^v}{2(L'')^2} - \frac{3L'''L^{iv}}{(L'')^3} + \frac{3(L''')^3}{(L'')^4} = -\frac{1}{2n\bar{x}},$$

$$B_2(\hat{\theta}) = -\frac{L^v + 12L''L'''}{8(L'')^3} + \frac{13L^{iv}L''' + 18(L'')^2L'''}{12(L'')^4} - \frac{11(L''')^3}{8(L'')^5} = -\frac{5}{12n^2\bar{x}^2},$$

$$V_1(\hat{\theta}) = -\frac{1}{L''} = \frac{1}{n\bar{x}},$$

$$B_2^*(\hat{\theta}) = B_2(\hat{\theta}) - B_1(\hat{\theta})B'_1(\hat{\theta}) - \frac{1}{2}B''_1(\hat{\theta})V_1(\hat{\theta}) = \frac{5}{12n^2\bar{x}^2}.$$

Assim, o estimador cujo viés tem ordem de magnitude $O(n^{-3})$ é

$$\tilde{\theta}_2 = \hat{\theta} - B_1(\hat{\theta}) - B_2^*(\hat{\theta}) = \log(\bar{x}) + \frac{1}{2n\bar{x}} - \frac{5}{12n^2\bar{x}^2}.$$

Capítulo 3

CORREÇÃO DE VIÉS DE ORDEM SUPERIOR

3.1 Introdução

Neste capítulo propomos uma técnica de estimação que é uma variação da estimação por máxima verossimilhança, baseada em sucessivas translações da função escore. O estimador obtido a partir dessa nova técnica corrige de forma preventiva o viés do EMV até a ordem de magnitude $O(n^{-(h+2)})$, em que $h = 0, 1, \dots$ e está relacionado com o número de translações realizadas na função escore.

Muitas técnicas de correção de viés consistem em estimar o viés, para depois proceder à correção do estimador, por exemplo, [12], [17]. Diferentemente, a técnica de estimação proposta nesta tese de doutorado é preventiva.

Pioneiramente, em meados dos anos 90, David Firth propôs um EMVC preventivamente. Ele observou que, na prática, a estimativa realizada para maximizar a função verossimilhança consiste basicamente em encontrar o zero da função escore. Assim, foi proposto realizar uma translação nessa função com o intuito de que a nova função transladada tivesse seu zero mais próximo do verdadeiro valor do parâmetro a ser estimado e, consequentemente, diminuísse o erro de estimativa. Em geral, esta metodologia produz um estimador não viesado até ordem $O(n^{-1})$.

No capítulo anterior vimos que essa ideia consiste em fazer uma translação

na função escore, de forma a anular o viés de Cox & Snell (2.16). A técnica de estimação apresentada neste trabalho é uma generalização desta proposta; para isso, utilizamos a função inversa da função escore. Com a equação de Bartlett (2.4) foi possível mostrar na Seção 2.4, utilizando a função inversa da função escore, como se dá a correção de viés uniparamétrica preventiva proposta por Firth. Com o intuito de expandir essa técnica, no sentido de produzir estimadores com vieses cada vez menores, generalizamos a expressão de Bartlett. A partir dessa expressão generalizada, que pode ser vista na equação (3.2), propomos uma sequência de translações na função escore, de tal forma que o zero da função transladada h vezes é um EMVC, cujo viés tem ordem de magnitude $O(n^{-(h+2)})$.

Utilizando a expansão de Taylor da função inversa da função escore, expandimos essa metodologia para o caso de estimação biparamétrica, logo em seguida para o espaço multiparamétrico. Assim, neste capítulo, é proposta uma metodologia de estimação uniparamétrica, biparamétrica e multiparamétrica, que produz estimadores não viesados até a ordem de magnitude desejada.

O presente capítulo está dividido da seguinte forma: na Seção 3.2 generalizamos a equação de Bartlett, e por meio dessa generalização mostramos como se comporta a função escore transladada; na Seção 3.3 propomos uma técnica nova de estimação, que é uma extensão da proposta por David Firth; esta metodologia foi estendida para o caso biparamétrico na Seção 3.4 e para o caso multiparamétrico na Seção 3.5; por fim, ofertamos as considerações finais na Seção 3.6.

3.2 Expressão de Bartlett Generalizada

Na Seção 2.4, vimos que se a função escore L' satisfaz determinadas condições de regularidade, então, $\widehat{\theta}$, zero desta função, pode ser expresso por meio da equação de Bartlett vista em (2.4), em que o termo residual tem ordem de magnitude $O_p(n^{-\frac{3}{2}})$. Nesta seção iremos generalizar essa expressão de Bartlett, no sentido de que para qualquer função G que satisfaça as suposições apresentadas nesta seção, teremos que θ^* zero da função G' , também, poderá ser escrito como uma expressão em que o termo residual tem ordem de magnitude $O_p(n^{-\frac{k+1}{2}})$, para

qualquer $k \geq 2$. Para isso, dado $k \geq 2$, considere uma função $G(\theta)$ que satisfaz as seguintes condições:

- **S1** G tem as k primeiras derivadas contínuas.
- **S2** G' tem exatamente uma raiz θ^* e G'' é negativa nesta raiz.
- **S3** G'' não se anula em uma vizinhança compacta de θ à qual pertence θ^* .
- **S4** G' é de ordem $O_p(n^{\frac{1}{2}})$.
- **S5** $G'' + I$ e $G''' - E[G''']$ são de ordem $O_p(n^{\frac{1}{2}})$.
- **S6** $I^{(k)} = \frac{\partial^k I}{\partial \theta^k}$ e $G^{(k)} = \frac{\partial^k G}{\partial \theta^k}$ são de ordem $O_p(n)$, para qualquer k -ésima derivada de G ou de I , em que $k \geq 2$.
- **S7** Tanto $(G'')^{-1}$ como I^{-1} são de ordem $O_p(n^{-1})$.
- **S8** Se uma função polinomial das derivadas $f(G', G'', \dots, G^{(i)})$ tem ordem de magnitude $O_p(n^k)$, então, tanto $E[f(G', G'', \dots, G^{(i)})]$ quanto a derivada dessa esperança são $O(n^k)$.

Consideraremos que a função escore L' satisfaz as condições de regularidade dadas na Seção 2.4; em particular, satisfaz essas suposições também.

As suposições **S1** a **S3** garantem que G' é invertível em uma vizinhança compacta de θ^* . Seja H tal função inversa, ou seja, $H(G'(\theta)) = \theta$ para todo θ pertencente a tal vizinhança de θ^* . Assim, podemos usar a expansão de Taylor da função H em torno do ponto $G'(\theta)$, ou seja,

$$\theta^* = \theta - H'G' + H''\frac{G'^2}{2!} + \dots + \frac{\partial^{(k)} H}{\partial G'^{(k)}} \frac{(-G')^k}{k!} + r,$$

sendo $G' = G'(\theta)$.

O Lema 3.2.1 estabelece como se comportam as derivadas da função H e qual a ordem de magnitude dessas derivadas.

Lema 3.2.1 A i -ésima derivada da função H é da forma $\frac{\partial^{(i)} H}{\partial G'^{(i)}} = \frac{H_i}{(G'')^{2i-1}}$, em que $H_1 = 1$, $H_i = H'_{i-1}G'' - (2i-3)G'''H_{i-1}$, para todo $i > 1$, com H'_i sendo a derivada em relação a θ . Ainda, tanto H_i quanto H'_i têm ordem $O_p(n^{i-1})$.

Dem. Vamos usar indução matemática. Para $i=1$ a afirmação vale, pois $H' = \frac{1}{G''}$. Para $i = 2$, temos que $\frac{\partial^{(2)} H}{\partial G'^{(2)}} = \frac{-G'''}{(G'')^3}$, que claramente satisfaz as hipóteses de indução. Suponha que a afirmação é verdadeira para um determinado i . Usando a regra da cadeia, temos que

$$\begin{aligned} \frac{\partial^{(i+1)} H}{\partial G'^{(i+1)}} &= \frac{\partial}{\partial \theta} \left[\frac{\partial^{(i)} H}{\partial G'^{(i)}} \right] \frac{1}{G''} && \text{(usando a regra da cadeia)} \\ &= \frac{\partial}{\partial \theta} \left[\frac{H_i}{(G'')^{2i-1}} \right] \frac{1}{G''} && \text{(por hipótese de indução)} \\ &= \frac{H'_i(G'')^{2i-1} - (2i-1)(G'')^{2i-2}G'''H_i}{(G'')^{4i-1}} \\ &= \frac{H'_iG'' - (2i-1)G'''H_i}{(G'')^{2i+1}}. \end{aligned}$$

Observe que o termo no numerador tem ordem $O_p(n^i)$, pois G' satisfaz a suposição **S6** e pela hipótese de indução temos que $H_i = O_p(n^{i-1})$ e $H'_i = O_p(n^{i-1})$. ■

Dado o Lema 3.2.1, podemos reescrever a expansão de Taylor da função H em torno do ponto zero da seguinte forma:

$$\theta^* - \theta = \frac{H_1}{G''} \frac{(-G')}{1!} + \frac{H_2}{(G'')^3} \frac{(-G')^2}{2!} + \cdots + \frac{H_k}{(G'')^{2k-1}} \frac{(-G')^k}{k!} + r.$$

Desta forma, fica fácil verificar que o i -ésimo termo da série tem ordem $O_p(n^{-\frac{i}{2}})$, para $i = 1, \dots, k$ e o último termo tem ordem $O_p(n^{-\frac{k+1}{2}})$.

Assim, a menos dos termos de ordem menor ou igual a $O_p(n^{-\frac{k+1}{2}})$, temos que

$$\theta^* - \theta = \frac{a_1(-G')(G'')^{2k-2} + a_2(-G')^2(G'')^{2k-4}H_2 + \cdots + a_k(-G')^kH_k}{(G'')^{2k-1}}, \quad (3.1)$$

sendo $a_i = \frac{1}{i!}$, para $i = 1, \dots, k$.

Usando a identidade $\frac{1}{(G'')^{2k-1}} = \sum_{j=0}^{2k-1} \mathcal{C}_j^{2k-1} \left(-\frac{1}{I}\right)^j \left(\frac{G''+I}{G''I}\right)^{2k-1-j}$, em que $\mathcal{C}_c^d = \frac{d!}{(d-c)!c!}$, temos que o produto do j -ésimo termo deste somatório pelo i -ésimo termo do numerador da soma dada pela equação (3.1) tem ordem $O_p(n^{\frac{-6k+3+j}{2}})O_p(n^{2k-1-\frac{i}{2}}) = O_p(n^{\frac{-2k+1-i+j}{2}})$. Logo, a soma dos primeiros $k-2+i$ termos deste produto tem ordem $O_p(n^{-\frac{k+1}{2}})$. Então, a menos dos termos de ordem

menor ou igual a $O_p(n^{-\frac{k+1}{2}})$, o i -ésimo termo da equação (3.1) pode ser escrito da forma

$$\begin{aligned} \frac{a_i(-G')^i(G'')^{2(k-i)}H_i}{(G'')^{2k-1}} &= a_i(-G')^i(G'')^{2(k-i)}H_i \sum_{j=0}^{2k-1} \mathcal{C}_j^{2k-1} \left(-\frac{1}{I}\right)^j \left(\frac{G''+I}{G''I}\right)^{2k-1-j} \\ &= a_i(-G')^i(G'')^{2(k-i)}H_i \sum_{j=k-1+i}^{2k-1} \mathcal{C}_j^{2k-1} \left(-\frac{1}{I}\right)^j \left(\frac{G''+I}{G''I}\right)^{2k-1-j} \\ &= \frac{a_i(-G')^i}{I^{2k-1}} H_i \sum_{j=k-1+i}^{2k-1} (-1)^j \mathcal{C}_j^{2k-1} (G'')^{(j+1-2i)} (G''+I)^{2k-1-j}. \end{aligned}$$

Como $j \geq 2i-1$, temos que $j+1-2i \geq 0$. Seja o par (i, j) tal que $1 \leq i \leq k$ e $k-1+i \leq j \leq 2k-1$. Defina

$$H_{ij} = \frac{(-1)^j \mathcal{C}_j^{2k-1} a_i (-G')^i H_i (G''+I)^{2k-1-j}}{I^{2k-1}}.$$

Considere $k_1 = j+1-2i$ e a identidade $(G'')^{k_1} = \sum_{l=0}^{k_1} \mathcal{C}_l^{k_1} (G''+I)^{k_1-l} (-I)^l$. Então, temos que o produto de H_{ij} pelo l -ésimo termo do somatório desta identidade tem ordem $O_p(n^{\frac{-2k-1+3i-j+k_1+l}{2}})$. Desta forma, a soma dos primeiros $k-1-i$ termos deste somatório tem ordem $O_p(n^{-\frac{k+1}{2}})$. Então,

$$\begin{aligned} H_{ij}(G'')^{(j+1-2i)} &= H_{ij} \sum_{l=0}^{k_1} \mathcal{C}_l^{k_1} (G''+I)^{k_1-l} (-I)^l \\ &= H_{ij} \sum_{l=k-i}^{k_1} \mathcal{C}_l^{k_1} (G''+I)^{k_1-l} (-I)^l + O_p(n^{-\frac{k+1}{2}}). \end{aligned}$$

A equação abaixo, a menos dos termos de ordem menor ou igual a $O_p(n^{-\frac{k+1}{2}})$, é a generalização da expressão dada por Bartlett,

$$\theta^* - \theta = \sum_{i=1}^k \sum_{j=k-1+i}^{2k-1} \sum_{l=k-i}^{j+1-2i} \frac{\mathcal{C}_{ijlk}(G')^i H_i (G''+I)^{2k-2i-l}}{I^{2k-1-l}}, \quad (3.2)$$

sendo $\mathcal{C}_{ijlk} = (-1)^{i+j+l} \mathcal{C}_j^{2k-1} \mathcal{C}_l^{j+1-2i} a_i$.

O termo da parcela correspondente aos índices i, j e l tem ordem $O_p(n^{\frac{-2k+i+l}{2}})$.

A Tabela 3.1 mostra os possíveis valores para o expoente de n .

Assim, podemos expandir θ^* , tal que $G'(\theta^*) = 0$, em parcelas da seguinte forma:

$$\theta^* - \theta = \sum_{i=1}^k b_i + O_p(n^{-\frac{k+1}{2}}), \quad (3.3)$$

Tabela 3.1. Possíveis valores do expoente de n .

	$k - 1 + i \leq j \leq 2k - 1$	$k - i \leq l \leq j + 1 - 2i$	$\frac{-2k+i+l}{2}$
i	j	l	
	$k - 1 + i$	$k - i$	$\frac{-k}{2}$
		$\mathbf{k - i \leq l \leq k - i + 1}$	
	$k + i$	$k - i$	$\frac{-k}{2}$
		$k - i + 1$	$\frac{-k+1}{2}$
		$\mathbf{k - i \leq l \leq k - i + 2}$	
	$k - i + 1$	$k - i$	$\frac{-k}{2}$
		$k - i + 1$	$\frac{-k+1}{2}$
		$k - i + 2$	$\frac{-k+2}{2}$
i	.	.	.
			.
			.
		$\mathbf{k - i \leq l \leq 2k - 2i}$	
	$2k - 1$	$k - i$	$\frac{-k}{2}$
		$k - i + 1$	$\frac{-k+1}{2}$
		$k - i + 2$	$\frac{-k+2}{2}$
		.	.
		.	.
		$2k - (2i - 2)$	$-\frac{i-2}{2}$
		$2k - (2i - 1)$	$-\frac{i-1}{2}$
		$2k - 2i$	$-\frac{i}{2}$

com $b_i = O_p(n^{-\frac{i}{2}})$, para $i = 1, \dots, k$.

É fácil verificar que se a função escore satisfaz essas suposições, então, podemos expandir o EMV em parcelas com essas características. Este caso particular

está dado na Tabela 3.2, pois, fazendo-se $k = 2$ e $G' = L'$, então, como $H_1 = 1$ e $H_2 = -L''$, temos que a soma dos termos da quarta coluna da Tabela 3.2 é exatamente igual à expressão (2.4).

Tabela 3.2. Expressão de Bartlett quando $k = 2$.

i	j	l	Equação (3.2)
1	2	1	$\frac{3L'(L''+I)}{I^2}$
	1	1	$\frac{-2L'(L''+I)}{I^2}$
	3	2	$\frac{L'}{I}$
2	3	0	$\frac{(L')^2 L'''}{2I^3}$

A ideia é fazer sucessivas translações na função escore de forma que a i -ésima parcela da equação (3.3) seja particionada em sub-parcelas com ordem de magnitude menor que $O_p(n^{-\frac{i}{2}})$. Assim, dada uma função G que satisfaz as suposições **S1** a **S8**, permitiremos fazer uma translação nela de forma que essas suposições continuem a valer na função transladada. Dessa forma acrescentamos a seguinte suposição:

- **S9** Sejam G' uma função que satisfaz as suposições **S1** a **S8** e θ^* zero desta função. Considere a expansão de θ^* como na equação (3.3) e defina $EO(h+1)$ como os elementos dessa expansão cuja soma tem valor esperado

com ordem de magnitude $O(n^{-(h+1)})$. Então, a função $G'_h = G' + A_h$, em que $A_h = -IE[EO(h+1)]$, satisfaz as suposições **S1** a **S8**. Além disso, tanto A_h como todas suas derivadas são de ordem $O(n^{-h})$, com $h \geq 0$.

O Lema 3.2.2 estabelece como se comporta a função inversa de G' transladada.

Lema 3.2.2 *Sejam G'_h , A_h como em **S9**, H_i como no Lema 3.2.1 e \mathbf{H}_h a função inversa de G'_h , então, a i -ésima derivada da função \mathbf{H}_h é da forma*

$$\frac{\partial^i \mathbf{H}_h}{\partial G'_h^i} = \frac{\mathbf{H}_h}{G'_h^{2i-1}},$$

em que $\mathbf{H}_h = H_i + \mathcal{R}_{ih}(A_h)$, com $\mathcal{R}_{1h} = 0$, $\mathcal{R}_{2h} = -A_h''$ e para todo $i \geq 2$ temos que:

$$i) \mathcal{R}_{ih}(A_h) = H'_{i-1}A'_h - (2i-3)H_{i-1}A''_h + \mathcal{R}'_{(i-1)h}(A_h)(G'' + A'_h) +$$

$$-(2i-3)\mathcal{R}_{(i-1)h}(A_h)(G''' + A''_h).$$

$$ii) \mathcal{R}_{(i-1)h}(A_h) = O_p(n^{i-2-h}).$$

$$iii) \mathcal{R}'_{(i-1)h}(A_h) = O_p(n^{i-2-h}).$$

Dem. Vamos usar o princípio da indução matemática nesta demonstração. Do Lema 3.2.1 temos que $\mathbf{H}_h = 1$ e $\mathbf{H}_2 = -(G''' + A''_h)$. Como por hipótese A''_h e A'''_h são $O_p(n^{-h})$, temos que a hipótese de indução é satisfeita para $i = 1$ e $i = 2$. Suponha as hipóteses de indução válidas para um certo $i \geq 2$. Pelo Lema 3.2.1 temos que

$$\begin{aligned} \mathbf{H}_{(i+1)} &= \mathbf{H}'_{i+1}G''_h - (2i-1)\mathbf{H}_iG'''_h \\ &= (H'_i + \mathcal{R}'_{ih}(A_h))(G'' + A'_h) - (2i-1)(H_i + \mathcal{R}_{ih}(A_h))(G''' + A''_h) \\ &= H'_i(G'' + A'_h) - (2i-1)H_i(G''' + A''_h) \\ &\quad + \mathcal{R}'_{ih}(A_h)(G'' + A'_h) - (2i-1)\mathcal{R}_{ih}(A_h)(G''' + A''_h) \\ &= \underbrace{H'_iG'' - (2i-1)H_iG''' + H'_iA'_h}_{H_{i+1}} - (2i-1)H_iA''_h \\ &\quad + \mathcal{R}'_{ih}(A_h)(G'' + A'_h) - (2i-1)\mathcal{R}_{ih}(A_h)(G''' + A''_h) \\ &= H_{i+1} + \mathcal{R}_{(i+1)h}(A_h). \end{aligned}$$

Observe que temos uma soma de quatro parcelas, sendo cada uma $O_p(n^{i-1-h})$. Logo, $\mathcal{R}(A_h)_{ih} = O_p(n^{i-1-h})$. A derivada de $\mathcal{R}_{(i+1)h}(A_h)$ é a soma de derivadas de ordem $O_p(n^{i-1-h})$ por hipótese, logo $\mathcal{R}'_{(i+1)h}(A_h)$, também, é $O_p(n^{i-1-h})$. ■

Com o lema anterior podemos provar o importante Teorema 3.2.1, que não só afirma que é possível particionar a i -ésima parcela da equação (3.3) em sub-parcelas de ordem de magnitude menor, como mostra a forma dessas sub-parcelas.

Teorema 3.2.1 *Considere as funções G' , G'_h e A_h , tal como no Lema 3.2.2 e seja θ^* tal que $G'_h(\theta^*) = 0$. Então,*

$$\theta^* - \theta = \sum_{i=1}^k \sum_{j=k-1+i}^{2k-1} \sum_{l=k-i}^{j+1-2i} \left[\frac{\mathcal{C}_{ijkl}(G')^i H_i (G'' + I)^{2k-2i-l}}{I^{2k-1-l}} + B_{ijkl} \right],$$

em que B_{ijkl} tem ordem $O_p(n^{\frac{-2k+i+l-1-2h}{2}})$ e \mathcal{C}_{ijkl} é definido como na equação (3.2).

Dem. Observe que

$$\begin{aligned} & (G'_h)^i H_i (G''_h + I)^{2k-2i-l} \\ &= (G' + A_h)^i (H_i + \mathcal{R}_{ih}(A_h)) (G'' + A'_h + I)^{2k-2i-l} \\ &= \left[(G')^i + \sum_{i_1=1}^i \mathcal{C}_{i_1}^i (G')^{i-i_1} A_h^{i_1} \right] [H_i + \mathcal{R}_{ih}(A_h)] \sum_{\kappa_1=0}^{\kappa} \mathcal{C}_{\kappa_1}^{\kappa} (G'' + I)^{\kappa-\kappa_1} (A'_h)^{\kappa_1}, \end{aligned}$$

sendo $\kappa = 2k - 2i - l$. Decompondo este produto de somas em cinco parcelas, temos:

$$\begin{aligned} P1 &= (G')^i H_i (G'' + I)^{\kappa}, \\ P2 &= (G')^i H_i \sum_{\kappa_1=1}^{\kappa} \mathcal{C}_{\kappa_1}^{\kappa} (G'' + I)^{\kappa-\kappa_1} (A'_h)^{\kappa_1}, \\ P3 &= (G')^i \mathcal{R}_{ih}(A_h) \sum_{\kappa_1=0}^{\kappa} \mathcal{C}_{\kappa_1}^{\kappa} (G'' + I)^{\kappa-\kappa_1} (A'_h)^{\kappa_1}, \\ P4 &= \left[\sum_{i_1=1}^i \mathcal{C}_{i_1}^i (G')^{i-i_1} A_h^{i_1} \right] H_i \sum_{\kappa_1=0}^{\kappa} \mathcal{C}_{\kappa_1}^{\kappa} (G'' + I)^{\kappa-\kappa_1} (A'_h)^{\kappa_1}, \\ P5 &= \left[\sum_{i_1=1}^i \mathcal{C}_{i_1}^i (G')^{i-i_1} A_h^{i_1} \right] \mathcal{R}_{ih}(A_h) \sum_{\kappa_1=0}^{\kappa} \mathcal{C}_{\kappa_1}^{\kappa} (G'' + I)^{\kappa-\kappa_1} (A'_h)^{\kappa_1}. \end{aligned}$$

Desta forma, cada parcela dividida por I^{2k-1-l} tem as seguintes ordens:

$$\begin{aligned} P1 &= O_p(n^{\frac{-2k+i+l}{2}}), \quad P2 = O_p(n^{\frac{-2k+i+l-1-2h}{2}}), \quad P3 = O_p(n^{\frac{-2k+i+l-4-2h}{2}}), \\ P4 &= O_p(n^{\frac{-2k+i+l-1-2h}{2}}), \quad P5 = O_p(n^{\frac{-2k+i+l-3-4h}{2}}). \end{aligned}$$

Então, defina

$$B_{ijkl}(A_h) = \mathcal{C}_{ijkl} \frac{P2 + P3 + P4 + P5}{I^{2k-1-l}},$$

que tem ordem $O_p(n^{\frac{-2k+i+l-1-2h}{2}})$. ■

Assim, para cada $ijkl$, temos que $\frac{\mathcal{C}_{ijkl}(G')^i H_i (G'' + I)^{2k-2i-l}}{I^{2k-1-l}} = O_p(n^{\frac{-2k+l+i}{2}})$. Isto quer dizer que se θ^* e θ_h^* são tais que $G'(\theta^*) = 0$ e $G'_h(\theta_h^*) = 0$, então, podemos escrever

$$\theta_h^* - \theta = \sum_{i=1}^k b_i^h + O_p(n^{-\frac{k+1}{2}}), \quad (3.4)$$

em que $b_i^h = b_i + B_i^h$, b_i é dado pela equação (3.3) e $B_i^h = O_p(n^{-\frac{i+1+2h}{2}})$. Ou seja, cada B_i^h tem a ordem de b_i com o expoente diminuído de $\frac{1+2h}{2}$. Em outras palavras, ao fazermos a translação $G'_h = G' + A_h$, conseguimos ainda expandir $\theta_h^* - \theta$ usando a equação (3.3) com G'_h , de forma que essa expansão é a mesma se usássemos G' , porém, cada parcela fica acrescida de um termo com ordem de magnitude menor.

Considere o modelo de Poisson apresentado na Seção 2.7. Defina $L'_0 = L' + \frac{1}{2}$. Então, $L''_0 = L''$ e $L'' + I = 0$. Pela equação (3.3) temos que

$$\begin{aligned} \theta_0^* &= \theta + \frac{L'_0}{I} + \frac{L'_0(L''_0 + I)}{I^2} + \frac{(L'_0)^2 L'''_0}{2I^3} + O_p(n^{-\frac{3}{2}}) \\ &= \theta + \frac{L'}{I} + \frac{1}{2I} + \frac{(L')^2 L'''}{2I^3} + \frac{(4L' + 1)L'''}{8I^3} + O_p(n^{-\frac{3}{2}}). \end{aligned}$$

De fato, como $h = 0$, temos que $B_1^0 = \frac{1}{2I}$ e $B_2^0 = \frac{(4L'+1)L'''}{8I^3}$ são $O_p(n^{-1})$ e $O_p(n^{-\frac{3}{2}})$, respectivamente. Continuando a sequência podemos definir $L'_1 = L'_0 + \frac{1}{24I}$, assim, $L_1^{(k)} = L^{(k)} + (-1)^{k+1} \frac{1}{24I}$, para $k = 2, 3, \dots$. Logo, expandindo a equação (3.4) temos

$$\theta_1^* = \theta + b_1 + B_1^0 + B_1^1 + b_2 + B_2^0 + b_3 + B_3^0 + b_4 + O_p(n^{-\frac{5}{2}}),$$

sendo

$$B_3^0 = -\frac{(L')^2}{I^3}, \quad B_1^1 = \frac{1}{24I^2},$$

os quais têm ordem de magnitude $O_p(n^{-2})$. Para encontrarmos esses termos, além de desprezar os elementos com ordem de magnitude menor que $O_p(n^{-\frac{5}{2}})$, usamos a expressão de b_3 , apresentada no Apêndice A.

A escolha do termo $\frac{1}{24I}$ não foi aleatória, o propósito da sua escolha é que no modelo de Poisson, no qual desejamos estimar o parâmetro canônico, veremos na seção seguinte que θ_1^* zero da função $L' + \frac{1}{2} + \frac{1}{24I}$ tem viés de ordem $O(n^{-3})$.

O que queremos ilustrar aqui é que podemos fazer sucessivas translações na função escore, de modo que cada parcela da expansão da equação (3.3) seja particionada em várias subparcelas B_i^k , com ordem de magnitude menor. Este fato é de muita relevância para o desenvolvimento do algoritmo proposto na seção seguinte, na qual generalizamos, no caso unidimensional, a ideia proposta por Firth [18].

3.3 Estimador de Máxima Verossimilhança com Viés de Ordem Reduzida

O Teorema 3.2.1 é uma poderosa ferramenta para expandir o EMV em parcelas tendo um certo controle sobre a ordem de magnitude de cada parcela. Usando este teorema, generalizamos a ideia desenvolvida por Firth [18]. Propomos uma técnica nova de estimação que produz um estimador com viés de ordem de magnitude $O(n^{-k})$, para qualquer $k = 2, 3, \dots$. A ideia é basicamente fazer sucessivas translações na função escore, sendo que a primeira translação é exatamente a sugerida por David Firth.

Considere a expansão do EMV $\hat{\theta}$ como na equação (3.2). Seja A_0 definido como na Seção 2.6, ou seja,

$$A_0 = -I\mathbb{E}[b_2]$$

$$= -IE \left[\frac{L'(L'' + I)}{I^2} + \frac{(L')^2 L'''}{2I^3} \right].$$

O Teorema 3.2.1 afirma que θ_0^* , tal que $L'_0(\theta_0^*) = 0$, com $L'_0 = L' - IE[b_2]$, pode ser escrito como

$$\begin{aligned} \theta_0^* - \theta &= b_1^0 + b_2^0 + b_3^0 + O_p(n^{-2}) \\ &= b_1^0 + b_2^0 + b_3 + B_3^0 + O_p(n^{-2}) \\ &= \frac{L'_0}{I} + \frac{L'_0(L'' + I)}{I^2} + \frac{(L'_0)^2 L'''}{2I^3} + b_3 + O_p(n^{-2}) \\ &= \frac{L' + A_0}{I} + \frac{(L' + A_0)(L'' + A'_0 + I)}{I^2} + \frac{(L' + A_0)^2 (L''' + A''_0)}{2I^3} + b_3 + O_p(n^{-2}) \\ &= \frac{L' + A_0}{I} + \frac{L'(L'' + I) + L'A'_0 + A_0(L'' + A'_0 + I)}{I^2} + \\ &\quad + \frac{(L')^2 L''' + (L')^2 A''_0 + 2A_0 L' L''' + 2L' A_0 A''_0 + A_0^2 L''' + A_0^2 A''_0}{2I^3} + b_3 + O_p(n^{-2}) \\ &= \frac{L' - IE[b_2]}{I} + \underbrace{\frac{L'(L'' + I)}{I^2} + \frac{(L')^2 L'''}{2I^3}}_{b_2} + b_3 + \frac{L'A'_0 + A_0(L'' + A'_0 + I)}{I^2} + \\ &\quad + \frac{(L')^2 A''_0 + 2A_0 L' L''' + 2L' A_0 A''_0 + A_0^2 L''' + A_0^2 A''_0}{2I^3} + O_p(n^{-2}). \end{aligned}$$

Observe que pela Afirmação 2.4.1, o valor esperado de b_3 é

$$\begin{aligned} E[b_3] &= E \left[\frac{L'(L'' + I)^2}{I^3} + \frac{3(L')^2 L'''(L'' + I)}{2I^4} + \frac{(L')^3 [3(L''')^2 - L'' L_h^{(iv)}]}{6I^5} \right] \\ &= O(n^{-2}). \end{aligned}$$

Como **S8** vale, $E[L'] = 0$ e $E[(L')^2] = -E[L''] = I$, temos que

$$\begin{aligned} E[\theta_0^* - \theta] &= E \left[\frac{L'_0}{I} + \frac{L'_0(L'' + I)}{I^2} + \frac{(L'_0)^2 L'''}{2I^3} + b_3 \right] + O(n^{-2}) \\ &= \frac{A_0 A'_0}{I^2} + \frac{I A''_0 + 2A_0 E[L' L'''] + A_0^2 E[L'''] + A_0^2 A''_0}{2I^3} + O(n^{-2}) \\ &= O(n^{-2}). \end{aligned}$$

Observe que $E[L' L'''] = O(n^{-1})$, pois $E[L' L'''] = E[L'(L''' - K) + L' K]$, sendo $K = E[L''']$. Assim, θ^* tal que $L'(\theta^*) - I(\theta^*)E[b_2|\theta^*] = 0$ tem viés de ordem $O(n^{-2})$.

A partir desse viés, definimos a seguinte quantidade,

$$\mathcal{R}e_0 = \frac{A_0 A'_0}{I^2} + \frac{I A''_0 + 2A_0 E[L' L'''] + A_0^2 E[L''']}{2I^3}. \quad (3.5)$$

De fato, existem na literatura diversos estimadores cujo viés tem ordem de magnitude $O(n^{-2})$, tais como [12], [17], [40], [41], entre outros. Porém, este estimador θ_0^* tem características muito atrativas, já que ele não restringe muito as classes de distribuições que podem ser usadas nessa técnica; precisamos apenas encontrar um zero de uma função, o que demanda baixo custo computacional; e principalmente, a técnica inspira e dá ferramentas para encontrarmos estimadores com ordem de magnitudes menores que $O_p(n^{-2})$.

Faremos agora uma nova translação na função escore transladada L'_0 com a função A_1 , que além de ser $O_p(n^{-1})$, tem a característica de anular o termo de ordem $O(n^{-2})$ do viés do estimador θ_0^* usado na primeira translação de L' .

Seja θ_1^* zero da função $L'_1 = L'_0 + A_1$, com $A_1 = -I\{\mathcal{R}e_0 + E[b_3^0 + b_4^0 + b_5]\}$, b_3^0 , b_4^0 , b_5 como na equação (3.4) e $\mathcal{R}e_0$ dado por (3.5). Como vimos, $\mathcal{R}e_0$, b_3^0 , b_4^0 e b_5 são de ordem $O_p(n^{-2})$, então, de fato $h = 1$, ou seja, $A_1 = O(n^{-1})$ ¹. O Teorema 3.2.1 e a equação (3.4) garantem que os termos B_3^1 , B_4^1 , B_5^0 e B_5^1 têm ordens menores ou iguais à $O_p(n^{-3})$. Assim, podemos escrever

$$\begin{aligned}
\theta_1^* - \theta &= b_1^1 + b_2^1 + b_3^1 + b_4^1 + b_5^1 + O_p(n^{-3}) \\
&= b_1^1 + b_2^1 + b_3^0 + B_3^1 + b_4^0 + B_4^1 + b_5^0 + B_5^1 + O_p(n^{-3}) \\
&= b_1^1 + b_2^1 + b_3^0 + b_4^0 + b_5 + O_p(n^{-3}) \\
&= \frac{L'_1}{I} + \frac{L'_1(L''_1 + I)}{I^2} + \frac{(L'_1)^2 L'''_1}{2I^3} + b_3^0 + b_4^0 + b_5 + O_p(n^{-3}) \\
&= \frac{(L'_0 + A_1)}{I} + \frac{(L'_0 + A_1)(L''_0 + A'_1 + I)}{I^2} + \frac{(L'_0 + A_1)^2 (L'''_0 + A''_1)}{2I^3} \\
&\quad + b_3^0 + b_4^0 + b_5 + O_p(n^{-3}).
\end{aligned}$$

Ou seja,

$$\begin{aligned}
\theta_1^* - \theta &= \frac{L'_0}{I} + \frac{L'_0(L''_0 + I)}{I^2} + \frac{(L'_0)^2 L'''_0}{2I^3} + \frac{A_1}{I} \\
&\quad + \frac{L'_0 A'_1 + A_1 (L''_0 + A'_1 + I)}{I^2} \\
&\quad + \frac{(L'_0)^2 A''_1 + 2A_1 L'_0 L'''_0 + 2L'_0 A_1 A''_1 + A_1^2 L'''_0 + A_1^2 A''_1}{2I^3} \\
&\quad + b_3^0 + b_4^0 + b_5 + O_p(n^{-3}).
\end{aligned}$$

¹Observe que neste caso, A não é uma variável aleatória.

Assim, o valor esperado da expressão acima é dado por

$$\begin{aligned}
\mathbb{E}[\theta_1^* - \theta] &= \mathbb{E} \left[\underbrace{\frac{L'_0}{I} + \frac{L'_0(L''_0 + I)}{I^2} + \frac{(L'_0)^2 L'''_0}{2I^3}}_{\mathcal{R}e_0 - \frac{A_0^2 A''_0}{2I^3}} \right] + \\
&+ \mathbb{E} \left[\frac{-I\{\mathcal{R}e + \mathbb{E}[b_3^0 + b_4^0 + b_5]\}}{I} \right] + \\
&+ \frac{A'_1 \mathbb{E}[L'_0] + A_1(\mathbb{E}[L''_0] + A'_1 + I)}{I^2} + \\
&+ \frac{A''_1 \mathbb{E}[L''_0] + 2A_1 \mathbb{E}[L'_0 L'''_0] + 2A_1 A''_1 \mathbb{E}[L'_0] + A_1^2 \mathbb{E}[L'''_0] + A_1^2 A''_1}{2I^3} + \\
&+ \mathbb{E}[b_3^0 + b_4^0 + b_5] + O(n^{-3}).
\end{aligned}$$

Como $\mathbb{E}[(L'_0)^2] = I + A_0^2$, $\mathbb{E}[L''_0] = A'_0 - I$ e $\mathbb{E}[L'_0 L'''_0] = \mathbb{E}[L'(L''' - K)] + A_0 A''_0 + A_0 K$, temos que θ_1^* , que é zero da função $L' - I\{\mathbb{E}[\mathcal{R}e_0 + b_3^0 + b_4^0 + b_5]\}$, tem viés dado por

$$\begin{aligned}
\mathbb{E}[\theta_1^* - \theta] &= \frac{A'_1 A_0 + A_1(A'_0 + A'_1)}{I^2} + \frac{A''_1(A'_0 - I)}{2I^3} + \\
&+ \frac{2A_1\{\mathbb{E}[L'(L''' - K)] + A_0 A''_0 + A_0 K\} + 2A_0 A_1 A''_1}{2I^3} + \\
&+ \frac{A_1^2 \mathbb{E}[L'''_0] + A_1^2 A''_1 + A_1^2(K + A''_0) + A_1^2 A''_1}{2I^3} + O(n^{-3}) \\
&= O(n^{-3}).
\end{aligned}$$

Lembrando que tanto A_h , como suas derivadas são $O_p(n^{-h})$, então, o viés de θ_1^* tem ordem menor ou igual a $O(n^{-3})$, como pode ser verificado na equação acima.

Pelas suposições **S4** e **S8**, temos que $\mathbb{E}[(L')^{2k}] = O(n^k)$, mas quando este expoente não é par, podemos melhorar a ordem de magnitude de potências da função escore, como podemos observar na afirmação seguinte.

Afirmiação 3.3.1 *Seja uma função $H = O_p(n^l)$ tal que sua derivada H' também é $O_p(n^l)$, então $\mathbb{E}[(L')^{2k+1} H] = O(n^{k+l})$.*

Dem. Novamente a prova se dá por indução matemática. Quando $k = 0$, a igualdade é válida devido à Afirmiação 2.4.1. Basta no referido lema fazer $P = H$. Se $k = 1$, tomemos $P = (L')^2 H$ na Afirmiação 2.4.1, pelas suposições **S4** e **S8** temos $\frac{\partial \mathbb{E}[P]}{\partial \theta} = O(n^{l+1})$. E de sorte $\mathbb{E}[P']$ também é $O(n^{l+1})$, basta aplicar novamente a Afirmiação 2.4.1 em P' . Portanto, novamente pela mesma

afirmação temos $E[(L')^3 H] = O(n^{l+1})$. Suponha por hipótese de indução que $E[(L')^{2k+1} H] = O(n^{k+l})$, para k inteiro positivo. Para sabermos a ordem de magnitude de $E[(L')^{2k+3} H]$, recorremos novamente à Afirmação 2.4.1. Assim, defina $P = (L')^{2k+2} H$. Logo, $P = O_p(n^{k+1+l})$, consequentemente seu valor esperado também tem a mesma ordem de magnitude devido à suposição **S8**. A quantidade $P' = (L')^{2k+1} L'' H$ tem seu valor esperado com ordem de magnitude $O_p(n^{k+1+l})$, por hipótese de indução. Portanto, devido à Afirmação 2.4.1, temos $E[(L')^{2k+3} H] = O(n^{k+1+l})$. ■

Observe que não é necessário utilizar todos os termos de b_3^0 , b_4^0 e b_5 , basta considerar apenas os termos que têm ordem de magnitude de seus valores esperados maior que $O(n^{-3})$. Por exemplo, no caso particular em que L'' não depende dos dados, como $L'' + I = 0$, temos que

$$\begin{aligned} E[b_5] &= E \left[\frac{L'(L'' + I)^4}{I^5} + \frac{10(L')^2(-L''')(L'' + I)^3}{2I^6} + \frac{15(L')^3H_3(L'' + I)^2}{6I^7} \right. \\ &\quad \left. + \frac{7(L')^4H_4(L'' + I)}{24I^8} + \frac{(L')^5H_5}{120I^9} \right] \quad (\text{Pela Afirmação 3.3.1}) \\ &= O(n^{-3}), \end{aligned}$$

em que H_3 , H_4 e H_5 são dados pelo Lema 3.2.1, neste caso não dependendo dos dados, e $H_5 = O(n^4)$.

Neste caso particular ainda há uma simplificação na expressão da função escore transladada L'_1 , que é

$$A_0 = -\frac{L'''}{2I}, \quad A_1 = \frac{L^v}{8I^2} + \frac{5L'''L^{iv}}{12I^3} + \frac{(L''')^3}{4I^4}.$$

Por exemplo, no modelo de Poisson descrito na Seção 3.2, em que desejamos estimar o parâmetro canônico, temos que o valor de A_1 pode ser escrito da seguinte forma:

$$A_1 = 0 + \frac{1}{8I} + 0 - \frac{5}{6I} + \frac{18}{24I} = \frac{1}{24I}.$$

Assim, temos que o estimador θ_1^* , obtido ao encontrar o zero da função escore transladada

$$L'(\theta) = \sum_{i=1}^n x_i - n \exp(\theta) + \frac{1}{2} + \frac{1}{24I},$$

é não viesado até a ordem $O(n^{-3})$.

Neste modelo, temos expressões fechadas para os cinco estimadores:

$$\begin{aligned}\hat{\theta} &= \log(\bar{x}), \quad \hat{\theta}_1 = \log(\bar{x}) + \frac{1}{2n\bar{x}}, \quad \theta_0^* = \log\left(\bar{x} + \frac{1}{2n}\right), \\ \theta_1^* &= \log\left(\frac{\bar{x}}{2} + \frac{3 + ((6n\bar{x} + 3)^2 + 6)^{\frac{1}{2}}}{12n}\right), \\ \tilde{\theta}_2 &= \log(\bar{x}) + \frac{1}{2n\bar{x}} - \frac{5}{n^2\bar{x}^2}.\end{aligned}$$

Observe, contudo, que

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \mathbb{E}[\log(\bar{x})] = \mathbb{E}\left[\log\left(\frac{x_1 + \dots + x_n}{n}\right)\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{\exp(-n\lambda)(n\lambda)^k \log(k)}{k!} - \log(n)\right] \\ &= -\infty \exp(-n\lambda) + \sum_{k=2}^{\infty} \frac{\exp(-n\lambda)(n\lambda)^k \log(k)}{k!} - \log(n).\end{aligned}$$

Ou seja, a probabilidade de \bar{x} ser zero é positiva e como o somatório acima converge, temos que o viés do estimador $\hat{\theta}$ é menos infinito. No caso dos estimadores $\hat{\theta}_1$ e $\tilde{\theta}_2$, o viés não é definido. Portanto, neste caso, a técnica de estimação preventiva é uma excelente ferramenta para se produzir estimadores com vieses de ordem de magnitude pequena.

Um outro exemplo interessante é o modelo gaussiano com média zero e variância desconhecida. Então, seja uma amostra $x = (x_1, \dots, x_n)$ independente e identicamente distribuída segundo uma gaussiana de média zero e variância σ^2 desconhecida. Ou seja, $x_i \sim \mathcal{N}(0, \sigma^2)$, para todo $i = 1, \dots, n$. Suponha que estamos interessados em estimar $d = 1/(2\sigma^2)$. A função escore é dada por

$$L'(d) = \frac{n}{2d} - \sum_{i=1}^n x_i^2.$$

Portanto, a informação de Fisher é

$$I(d) = \frac{n}{2d^2}.$$

Logo, temos que

$$A_0 = -\frac{1}{d}, \quad A_1 = 0.$$

Assim, temos que o estimador d_1^* coincide com o estimador d_0^* proposto por Firth, que é obtido por encontrar o zero da função escore transladada $L' + A_0$ e é dado por

$$d_0^* = \frac{n-2}{2 \sum_{i=1}^n x_i^2}, \quad n > 2.$$

Observe que d_0^* é não viesado² para d , pois, $\sum_{i=1}^n x_i^2 / \sigma^2 \sim \chi_{(n)}^2$. Logo,

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{i=1}^n x_i^2 \right)^r \right] &= \frac{\sigma^r 2^{\frac{r}{2}} \Gamma(\frac{n+r}{2})}{\Gamma(\frac{n}{2})} \quad (\text{Desde que } n > -r), \\ \mathbb{E} \left[\frac{1}{2 \sum_{i=1}^n x_i^2} \right] &= \frac{\Gamma(\frac{n-2}{2})}{4\sigma^2 \Gamma(\frac{n}{2})} \quad (\text{Fazendo-se } r = -2), \\ &= \frac{1}{2(n-2)\sigma^2}. \end{aligned}$$

Ou seja,

$$\mathbb{E} \left[\frac{n-2}{2 \sum_{i=1}^n x_i^2} \right] = d.$$

Com o teorema seguinte generalizaremos esse raciocínio para encontrarmos um estimador θ_k^* cujo viés tenha ordem de magnitude $O(n^{-(k+2)})$, para qualquer $k = 0, 1, 2, \dots$.

Teorema 3.3.1 *Seja uma amostra $X = (X_1, \dots, X_n)$ independente e identicamente distribuída tal que a função escore satisfaz as suposições **S1** a **S9**. Considere a expansão do EMV dada na equação (3.3) até $k = 3$ e defina $EO(1)$ como os elementos dessa expansão cuja soma tem seu valor esperado $O(n^{-1})$. Então, θ_0^* que é zero da função $L' + A_0$, com $A_0 = -\mathbb{E}[EO(1)]$, é não viesado até ordem $O(n^{-2})$. Da mesma forma, considere a expansão de θ_0^* como na equação (3.4) até $k = 5$ e defina $EO(2)$ como os elementos dessa expansão cuja soma tem*

²Na verdade, d_0^* é o estimador uniformemente não viesado de mínima variância para d .

seu valor esperado $O(n^{-2})$. Então, θ_1^* que é zero da função $L' + A_0 + A_1$, com $A_1 = -IE[EO(2)]$, é não viésado até ordem $O(n^{-3})$. Sucessivamente, considere a expansão de θ_{h-1}^* como na equação (3.4) até $k = 2h + 3$ e defina $EO(h+1)$ como os elementos dessa expansão cuja soma tem seu valor esperado $O(n^{-(h+1)})$. Então, θ_h^* que é zero da função $L' + A_0 + A_1 + \dots + A_h$, com $A_h = -IE[EO(h+1)]$, é não viésado até ordem $O(n^{-(h+2)})$.

Dem. Vamos provar por indução matemática. Vimos que a equação (3.3) garante que podemos expandir o EMV da seguinte forma $\widehat{\theta} - \theta = b_1 + b_2 + O_p(n^{-\frac{3}{2}})$ e que θ_0^* , zero da função $L' + A_0$, com $A_0 = -IE[b_2]$, é não viésado até ordem $O(n^{-2})$. O termo b_3 foi desprezado porque seu valor esperado é $O(n^{-2})$. Observe que b_2 é o conjunto dos termos dessa expansão cuja soma tem valor esperado $O(n^{-1})$, pois, $b_i = O_p(n^{-\frac{i}{2}})$ e vale a suposição **S8**. Da mesma forma, a equação (3.4) garante que podemos expandir θ_0^* da seguinte forma $\theta_0^* - \theta = b_1^0 + b_2^0 + b_3^0 + b_4^0 + b_5 + O_p(n^{-3})$ e vimos que θ_1^* , zero da função $L' + A_0 + A_1$, com $A_1 = -I\{\mathcal{R}e_0 + E[b_3^0 + b_4^0 + b_5]\}$, é não viésado até a ordem $O(n^{-3})$. Observe que $\{\mathcal{R}e_0 + E[b_3^0 + b_4^0 + b_5]\}$ é a soma dos elementos cuja soma tem seu valor esperado $O(n^{-2})$. Assim, verificamos que o teorema é válido para $h = 0$ e $h = 1$. Suponha agora que este teorema é válido para $h - 1$, ou seja, por hipótese de indução,

$$\begin{aligned} E[\theta_{h-1}^* - \theta] &= E\left[\frac{L'_{h-1}}{I} + b_2^{h-1} + \dots + b_{2h+1}^{h-1} + O_p(n^{-(h+1)})\right] \\ &= O(n^{-(h+1)}). \end{aligned}$$

Vamos provar que essa teoria vale para h qualquer. Dada a hipótese de indução, fica fácil ver que, por definição,

$$EO(h+1) = b_1^{(h-1)} + b_2^{(h-1)} + \dots + b_{2h+1}^{(h-1)} + b_{h+2}^{(h-1)} + b_{2h+3}^{(2h-1)}.$$

A expansão da equação (3.4) até $k = 2h + 3$ é

$$\theta_h^* - \theta = b_1^h + \dots + b_{2h+3}^h + O_p(n^{-(h+2)}).$$

Observe que para cada parcela temos que

$$b_i^h = b_i + B_i^0 + \dots + B_i^{h-1} + B_i^h$$

$$\begin{aligned}
&= b_i + B_i^0 + \cdots + B_i^{h-1} + O_p(n^{-(h+2)}) \\
&= b_i^{h-1} + O_p(n^{-(h+2)}).
\end{aligned}$$

Isso quer dizer que para $i > 2$ podemos agrupar o termo B_i^h em $O_p(n^{-(h+2)})$, pois $B_i^h = O_p(n^{-\frac{i+1+2h}{2}})$. No caso em que $i = 2$, temos

$$B_2^h = \frac{L'_{h-1}A_h + A_h(L''_{h-1} + I + A'_h)}{I^2} + \frac{(2A_hL'_{h-1} + A_h^2)(L'''_{h-1} + A''_h)}{2I^3},$$

logo $E[B_2^h] = O(n^{-(h+2)})$. Assim, usando esses fatos e a hipótese de indução, temos que

$$\begin{aligned}
E[\theta_h^* - \theta] &= E[b_1^h + b_2^{h-1} + B_2^h + \cdots + b_{2h+3}^{h-1} + O_p(n^{-(h+2)})] \\
&= E\left[\frac{L'_{h-1} + A_h}{I} + b_2^{h-1} + B_2^h + \cdots + b_{2h+3}^{h-1} + O_p(n^{-(h+2)})\right] \\
&= \frac{A_h}{I} + E[EO(h-1)] + E[O_p(n^{-(h+2)})] \\
&= O(n^{-(h+2)}). \blacksquare
\end{aligned}$$

Este teorema garante que se a função escore satisfaz condições bastante gerais, que são as suposições **S1** a **S9**, então podemos fazer sucessivas translações na função escore de forma que o zero dessa função transladada gere um estimador cujo viés tem ordem de magnitude $O(n^{-(h+2)})$, para qualquer $h \geq 0$.

3.4 Correção de Viés Biparamétrica

Nesta seção iremos estender a correção de viés para o caso biparamétrico. Assim, como na Seção 3.3, a ideia básica é seguir os passos do Teorema 3.3.1 e propor um estimador cujo viés tem ordem de magnitude $O(n^{-(h+2)})$, para $h = 0, 1, \dots$

Dada uma função G real, definida no espaço paramétrico bidimensional, usaremos uma notação mais simples do que a dada na Seção 2.5 no caso multiparamétrico. Considere o parâmetro $\theta = (\alpha, \beta)^T$, então, o vetor gradiente $G'(\alpha, \beta)$ é definido como

$$\begin{aligned}
G' : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \times \mathbb{R} \\
(\alpha, \beta)^T &\longmapsto (G_\alpha, G_\beta)^T,
\end{aligned}$$

em que $G_\alpha = \frac{\partial G}{\partial \alpha}$ e $G_\beta = \frac{\partial G}{\partial \beta}$ são as derivadas parciais de G . Da mesma forma, as demais derivadas parciais são denotadas por $G_{\alpha\beta} = \frac{\partial^2 G}{\partial \alpha \partial \beta}$, $G_{\alpha \dots \alpha} = \frac{\partial^{(k)} G}{\partial \alpha^{(k)}}$. Por fim, as derivadas seguem a mesma notação matricial, $G'' = [\partial G'/\partial \alpha, \partial G'/\partial \beta]$, $\dots, G^{(k)} = [\partial G^{(k-1)}/\partial \alpha, \partial G^{(k-1)}/\partial \beta]$. Note que $G^{(k)}$ é uma matriz $2 \times 2(k-1)$, com $k > 1$.

Estabeleceremos algumas suposições de regularidade biparamétricas, são elas:

- **SB1** As derivadas parciais de G são contínuas.
- **SB2** O gradiente G' tem exatamente uma raiz θ^* e a matriz $G''(\theta^*)$ é negativa definida.
- **SB3** $E[G'G'^T]$ tem todos os elementos finitos.
- **SB4** G'' não depende dos dados.
- **SB5** $G' = O_p(n^{\frac{1}{2}})$.
- **SB6** Se uma função polinomial das derivadas parciais $f(\frac{\partial G}{\partial \alpha}, \frac{\partial^2 G}{\partial \alpha \partial \beta}, \frac{\partial^3 G}{\partial \alpha \partial \alpha \partial \beta}, \dots)$ tem ordem de magnitude $O_p(n^k)$, então, tanto $E[f(\frac{\partial G}{\partial \alpha}, \frac{\partial^2 G}{\partial \alpha \partial \beta}, \frac{\partial^3 G}{\partial \alpha \partial \alpha \partial \beta}, \dots)]$ quanto a derivada dessa esperança têm ordem de magnitude $O(n^k)$.
- **SB7** Os elementos da matriz $(G'')^{-1}$ são $O_p(n^{-1})$ em uma vizinhança de θ que inclui θ^* .
- **SB8** Os elementos das matrizes $G'', G''', \dots, G^{(k)}$ são $O_p(n)$ em uma vizinhança de θ que inclui θ^* .

Estas suposições são adaptações das condições de regularidade multiparamétricas para a função escore L' , dadas na Seção 2.5. Elas são importantes para que possamos inverter a função G' , como garantem as suposições **SB1** e **SB2**. Será necessário que o valor esperado da matriz $G'G'^T$ seja finito, pois sua matriz inversa será decomposta em parcelas com ordem de magnitude decrescente; isso é assegurado pela terceira suposição. Já **SB4** tem apenas a finalidade de simplificar os cálculos de valores esperados de elementos de uma matriz $G^{(k)}$. A relevância

das suposições **SB5** a **SB8** se dá na garantia da ordem de magnitude de seus elementos e de funções polinomiais de seus elementos.

Seja L'_k o k -ésimo elemento da sequência de translações da função escore $L'_k = L'_{k-1} + \mathcal{A}_k$, em que $\mathcal{A}_k = (A_k, B_k)^T$, com A_k e B_k sendo $O_p(n^{-k})$ e $L'_0 = L' + \mathcal{A}_0$. Denotaremos por $\theta_k^* = (\alpha_k^*, \beta_k^*)^T$ o zero da função L'_k . Em outras palavras, $L_{k\alpha}(\alpha_k^*, \beta_k^*) = L_{k\beta}(\alpha_k^*, \beta_k^*) = 0$.

Se G'_h satisfaz as suposições **SB1** e **SB2**, então, dada sua inversa H_h , a expansão de Taylor de θ_h^* , em torno de G'_h é dada por

$$\theta_h^* = \theta - H'_h G'_h + \frac{1}{2} H''_h (G'_h \otimes G'_h) + \cdots + \frac{(-1)^k}{k!} H_h^{(k)} (\underbrace{G'_h \otimes \cdots \otimes G'_h}_{k-\text{vezes}}) + r_k, \quad (3.6)$$

em que $|r_k| \leq |G'(\theta)|$.

Observe que as suposições **SB**'s se referem a uma função G e não necessariamente à função log-verossimilhança L ; isto se deve ao fato de que iremos fazer sucessivas translações na função escore L' de forma que, por hipótese, esta nova função transladada continue satisfazendo estas suposições. Dessa forma, acrescentamos a seguinte suposição:

- **SB9** Sejam G'_h uma função que satisfaz as suposições **SB1** a **SB8** e θ_h^* zero desta função, considere a expansão de θ_h^* como na equação (3.6) e defina $EO(h+2)$ como os elementos dessa expansão cuja soma tem valor esperado com ordem de magnitude $O(n^{-(h+2)})$. Então, a função $G'_{h+1} = G'_h + A_{h+1}$, com $A_{h+1} = -IE[EO(h+2)]$, satisfaz as suposições **SB1** a **SB8**. Além disso, tanto A_{h+1} quanto todas suas derivadas são de ordem $O_p(n^{-(h+1)})$, com $h \geq 0$.

Estamos impondo que pequenas translações deste tipo não impeçam que a função deixe de satisfazer as suposições biparamétricas **SB**'s.

Suponha que a função escore transladada L'_k satisfaz as condições de regularidade biparamétrica **SB1** a **SB9** e seja H_k sua função inversa. Ou seja,

$$\begin{aligned} H_k : \quad \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \times \mathbb{R} \\ (L_\alpha(\alpha, \beta), L_\beta(\alpha, \beta))^T &\longmapsto (H_k^1(L_\alpha, L_\beta), H_k^2(L_\alpha, L_\beta))^T, \end{aligned}$$

tal que $(H_k^1(L_\alpha, L_\beta), H_k^2(L_\alpha, L_\beta))^T = (\alpha, \beta)^T$. Em outras palavras, temos que $H_k(L'_k(\theta)) = \theta$.

Encontrar esta função inversa H_k pode ser uma tarefa árdua e até mesmo impossível. Então, uma solução para superar essa dificuldade é, assim como na Seção 3.2, em que tratamos do caso uniparamétrico, usar o teorema da função inversa [42], o qual garante que a matriz H'_k é igual à inversa da matriz L''_k . No caso biparamétrico sempre é fácil fazer essa inversão, pois

$$\begin{aligned} H'_k &= (L''_k)^{-1} \\ &= \begin{bmatrix} L_{k\beta\beta} & -L_{k\alpha\beta} \\ -L_{k\beta\alpha} & L_{k\alpha\alpha} \end{bmatrix} \frac{1}{|L''_k|} \\ &= \begin{bmatrix} \frac{\partial H_k^1}{\partial L_{k\alpha}} & \frac{\partial H_k^1}{\partial L_{k\beta}} \\ \frac{\partial H_k^2}{\partial L_{k\alpha}} & \frac{\partial H_k^2}{\partial L_{k\beta}} \end{bmatrix}. \end{aligned}$$

Para encontrar as derivadas de H'_k , utilizamos derivação implícita. Assim, dada

$$H''_k(L_k(\theta)) = \begin{bmatrix} \frac{\partial^2 H_k^1}{\partial L_{k\alpha}^2} & \frac{\partial^2 H_k^1}{\partial L_{k\alpha} \partial L_{k\beta}} & \frac{\partial^2 H_k^1}{\partial L_{k\beta} \partial L_{k\alpha}} & \frac{\partial^2 H_k^1}{\partial L_{k\beta}^2} \\ \frac{\partial^2 H_k^2}{\partial L_{k\alpha}^2} & \frac{\partial^2 H_k^2}{\partial L_{k\alpha} \partial L_{k\beta}} & \frac{\partial^2 H_k^2}{\partial L_{k\beta} \partial L_{k\alpha}} & \frac{\partial^2 H_k^2}{\partial L_{k\beta}^2} \end{bmatrix},$$

temos, por exemplo, que

$$\begin{aligned} \frac{\partial^2 H_k^1}{\partial L_{k\alpha} \partial L_{k\beta}} &= \frac{\partial}{\partial L_{k\alpha}} \left[\frac{\partial H_k^1}{\partial L_{k\beta}} \right] \\ &= \frac{\partial}{\partial H_k^1} \left[\frac{\partial H_k^1}{\partial L_{k\beta}} \right] \frac{\partial H_k^1}{\partial L_{k\alpha}} + \frac{\partial}{\partial H_k^2} \left[\frac{\partial H_k^1}{\partial L_{k\beta}} \right] \frac{\partial H_k^2}{\partial L_{k\alpha}} \\ &= \frac{\partial}{\partial \alpha} \left[\frac{-L_{k\alpha\beta}}{|L''_k|} \right] \frac{L_{k\beta\beta}}{|L''_k|} + \frac{\partial}{\partial \beta} \left[\frac{-L_{k\alpha\beta}}{|L''_k|} \right] \left(\frac{-L_{k\beta\alpha}}{|L''_k|} \right) \\ &= \frac{(L_{k\alpha\beta}|L''_k|_\alpha - L_{k\alpha\beta\alpha}|L''_k|)}{|L''_k|^3} L_{k\beta\beta} + \frac{(L_{k\alpha\beta\beta}|L''_k| - L_{k\alpha\beta}|L''_k|_\beta)}{|L''_k|^3} L_{k\beta\alpha}, \end{aligned}$$

em que $|L''_k|_\alpha$ e $|L''_k|_\beta$, são as derivadas do determinante de L''_k em relação a α e a β , respectivamente.

Agora, através do lema seguinte, veremos como se relaciona essa matriz H'_k com a matriz H'_{k-1} .

Lema 3.4.1 *Considere a sequência de translações da função escore, $L'_k = L'_{k-1} + \mathcal{A}_k$, com $L'_0 = L' + \mathcal{A}_0$ e $\mathcal{A}_k = (A_k, B_k)^T$, um vetor com ordem de magnitude*

$O_p(n^{-k})$. Suponha que cada L'_k satisfaz as suposições **SB1** a **SB9**. Então, a matriz H'_k pode ser decomposta em parcelas da seguinte forma:

$$H'_k = H'_{k-1} + \mathcal{R}e(k) + O_p(n^{-(2k+3)}),$$

em que

$$\mathcal{R}e(k) = \begin{bmatrix} B_{k\beta} & -A_{k\beta} \\ -B_{k\alpha} & A_{k\alpha} \end{bmatrix} \frac{1}{|L''_{k-1}|} - \frac{|L''_k| - |L''_{k-1}|}{|L''_k|} (L''_{k-1})^{-1}.$$

Dem. Vamos provar por indução matemática. Para $k = 0$, temos que

$$\begin{aligned} H'_0 &= \begin{bmatrix} L_{\beta\beta} + B_{0\beta} & -(L_{\alpha\beta} + A_{0\beta}) \\ -(L_{\beta\alpha} + B_{0\alpha}) & L_{\alpha\alpha} + A_{0\alpha} \end{bmatrix} \frac{1}{|L''_0|} \\ &= \left[\begin{bmatrix} L_{\beta\beta} & -L_{\alpha\beta} \\ -L_{\beta\alpha} & L_{\alpha\alpha} \end{bmatrix} + \begin{bmatrix} B_{0\beta} & -A_{0\beta} \\ -B_{0\alpha} & A_{0\alpha} \end{bmatrix} \right] \frac{1}{|L''| + R_0}, \end{aligned}$$

sendo $R_0 = |L''_0| - |L''|$. Ou seja, $R_0 = A_{0\alpha}L_{\beta\beta} + B_{0\beta}L_{\alpha\alpha} + A_{0\alpha}B_{0\beta} - A_{0\beta}L_{\beta\alpha} - B_{0\alpha}L_{\alpha\beta} - A_{0\beta}B_{0\alpha}$. Portanto, $R_0 = O_p(n)$.

Mas observe que $\frac{1}{|L''| + R_0} = \frac{1}{|L''|} - \frac{R_0}{(|L''| + R_0)|L''|}$, então,

$$\begin{aligned} H'_0 &= (L'')^{-1} + \begin{bmatrix} B_{0\beta} & -A_{0\beta} \\ -B_{0\alpha} & A_{0\alpha} \end{bmatrix} \frac{1}{|L''|} - \frac{(L'')^{-1}R_0}{|L''| + R_0} + O_p(n^{-3}) \\ &= H' + \mathcal{R}e(0) + O_p(n^{-3}). \end{aligned}$$

Suponha que a afirmação é válida para $k - 1$ e seja H'_k a função inversa de $L'_k = L'_{k-1} + \mathcal{A}_k$. De maneira análoga, define-se $R_k = |L''_k| - |L''_{k-1}|$. É fácil verificar que $R_k = O_p(n^{1-k})$. Logo,

$$\begin{aligned} H'_k &= \begin{bmatrix} L_{k\beta\beta} & -L_{k\alpha\beta} \\ -L_{k\beta\alpha} & L_{k\alpha\alpha} \end{bmatrix} \frac{1}{|L''_k|} \\ &= \left[\begin{bmatrix} L_{(k-1)\beta\beta} & -L_{(k-1)\alpha\beta} \\ -L_{(k-1)\beta\alpha} & L_{(k-1)\alpha\alpha} \end{bmatrix} + \begin{bmatrix} B_{k\beta} & -A_{k\beta} \\ -B_{k\alpha} & A_{k\alpha} \end{bmatrix} \right] \end{aligned}$$

$$\begin{aligned}
& \times \left[\frac{1}{|L''_{(k-1)}|} - \frac{R_k}{(|L''_{(k-1)}| + R_k)|L''_{k-1}|} \right] \\
& = (L''_{k-1})^{-1} + \begin{bmatrix} B_{k\beta} & -A_{k\beta} \\ -B_{k\alpha} & A_{k\alpha} \end{bmatrix} \frac{1}{|L''_{k-1}|} - \frac{(L''_{k-1})^{-1} R_k}{|L''_{k-1}| + R_k} + O_p(n^{-(2k+3)}) \\
& = H'_{k-1} + \mathcal{R}e(k) + O_p(n^{-(2k+3)}).
\end{aligned}$$

Logo, por indução, a afirmação é válida para qualquer k . ■

Como $|L''_k| - |L''_{k-1}|$ tem ordem de magnitude $O_p(n^{1-k})$, temos que $\mathcal{R}e(k) = O_p(n^{-(k+2)})$.

O Lema 3.4.1 é de relevância para nosso desenvolvimento, pois mostra como se comporta a matriz H'_k após uma translação da função escore já transladada. O próximo lema estabelece a ordem de magnitude das derivadas de H'_k .

Lema 3.4.2 *Seja L_k uma função que satisfaz as suposições **SB1** a **SB9**; então, a ordem de magnitude da l -ésima derivada da:*

a) função inversa H_k é $O_p(n^{-l})$.

b) função $\mathcal{R}e(k)$ é $O_p(n^{-(k+l+2)})$.

Dem.

a) Novamente vamos demonstrar esta afirmação usando o princípio da indução matemática. Para $l = 1$, $H'_k = (L'')^{-1}$, pelas suposições **SB3**, **SB6** a **SB9** claramente $H'_k = O_p(n^{-1})$. Suponha a hipótese de indução válida para $l - 1$. Por definição, $H_k^{(l)} = \left[\frac{\partial H_k^{(l-1)}}{\partial L_{k\alpha}}, \frac{\partial H_k^{(l-1)}}{\partial L_{k\beta}} \right]$. Sem perda de generalidade, considere $h_{k11}^{(l-1)}$ o elemento de posição $(1, 1)$ da matriz $H_k^{(l-1)}$, então

$$\begin{aligned}
h_{k11}^{(l)} &= \frac{\partial}{\partial L_{k\alpha}} \left[h_{k11}^{(l-1)} \right] \\
&= \frac{\partial}{\partial H_k^1} \left[h_{k11}^{(l-1)} \right] \frac{\partial H_k^1}{\partial L_{k\alpha}} + \frac{\partial}{\partial H_k^2} \left[h_{k11}^{(l-1)} \right] \frac{\partial H_k^2}{\partial L_{k\alpha}} \\
&= O_p(n^{-(l-1)}) O_p(n^{-1}) + O_p(n^{-(l-1)}) O_p(n^{-1}) \\
&= O_p(n^{-l}).
\end{aligned}$$

b) A demonstração deste item é análoga à do item (a). É fácil verificar que derivar $\mathcal{R}e(k)$ em relação a θ não muda sua ordem de magnitude e que $\frac{\partial H_k^1}{\partial L_{k\alpha}} = O(n^{-1})$. Assim, por exemplo, $\mathcal{R}e(k)' = \frac{\partial}{\partial H_k^1} [\mathcal{R}e(k)] \frac{\partial H_k^1}{\partial L_{k\alpha}} + \frac{\partial}{\partial H_k^2} [\mathcal{R}e(k)] \frac{\partial H_k^2}{\partial L_{k\alpha}}$ tem ordem $O(n^{-(k+3)})$. Então, suponha que a igualdade vale para $l-1$. Sem perda de generalidade, temos que

$$\begin{aligned} \mathcal{R}e(k)^{(l)} &= \frac{\partial}{\partial L_{k\alpha}} [\mathcal{R}e(k)^{(l-1)}] \\ &= \frac{\partial}{\partial H_k^1} [\mathcal{R}e(k)^{(l-1)}] \frac{\partial H_k^1}{\partial L_{k\alpha}} + \frac{\partial}{\partial H_k^2} [\mathcal{R}e(k)^{(l-1)}] \frac{\partial H_k^2}{\partial L_{k\alpha}} \\ &= O_p(n^{-(k+l+1)})O_p(n^{-1}) + O_p(n^{-(k+l+1)})O_p(n^{-1}) \\ &= O_p(n^{-(k+l+2)}). \blacksquare \end{aligned}$$

Dadas as suposições **SB5**, **SB6** e com a Afirmação 3.4.2 fica fácil observar que o l -ésimo termo da expressão (3.6) tem ordem de magnitude $O_p(n^{-\frac{l}{2}})$. Assim, considere os três primeiros termos dessa expressão para o EMV

$$\widehat{\theta} = \theta - H'L' + \frac{1}{2}H''(L' \otimes L') + O_p(n^{-\frac{3}{2}}).$$

Em notação matricial,

$$\begin{aligned} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\beta} - \beta \end{pmatrix} &= - \begin{pmatrix} h_{11}^{(1)}L_\alpha + h_{12}^{(1)}L_\beta \\ h_{21}^{(1)}L_\alpha + h_{22}^{(1)}L_\beta \end{pmatrix} + \\ &+ \frac{1}{2} \begin{pmatrix} h_{11}^{(2)}L_\alpha^2 + (h_{12}^{(2)} + h_{13}^{(2)})L_\alpha L_\beta + h_{14}^{(2)}L_\beta^2 \\ h_{21}^{(2)}L_\alpha^2 + (h_{22}^{(2)} + h_{23}^{(2)})L_\alpha L_\beta + h_{24}^{(2)}L_\beta^2 \end{pmatrix} + O_p(n^{-\frac{3}{2}}), \end{aligned}$$

sendo que $h_{ij}^{(k)}$ é o elemento de posição (i, j) da matriz $H^{(k)}$.

Defina $EO(1)$ como a soma dos elementos dessa expressão cujo valor esperado é $O(n^{-1})$ e $\mathcal{X}_0 = E[EO(1)]$, então, θ_0 , zero da função $L' + \mathcal{A}_0$, sendo $\mathcal{A}_0 = L''\mathcal{X}_0$, tem viés com ordem de magnitude $O(n^{-2})$. Observe que

$$\mathcal{X}_0 = E \left[\frac{1}{2}H''(L' \otimes L') \right]$$

$$= \frac{1}{2} \mathbb{E} \begin{bmatrix} \left(h_{11}^{(2)} L_\alpha^2 + (h_{12}^{(2)} + h_{13}^{(2)}) L_\alpha L_\beta + h_{14}^{(2)} L_\beta^2 \right) \\ \left(h_{21}^{(2)} L_\alpha^2 + (h_{22}^{(2)} + h_{23}^{(2)}) L_\alpha L_\beta + h_{24}^{(2)} L_\beta^2 \right) \end{bmatrix}.$$

Como a suposição **SB4** garante que $\mathbb{E}[H'] = H'$, temos que

$$\begin{aligned} \theta_0^* - \theta &= -H'_0 L'_0 + \frac{1}{2} H''_0 (L'_0 \otimes L'_0) + O_p(n^{-\frac{3}{2}}) \quad (\text{pelo Lema 3.4.1}) \\ &= -(H' + \mathcal{R}e(0)) (L' + L'' \mathcal{X}_0) + \\ &+ \frac{1}{2} (H'' + \mathcal{R}e'(0)) ((L' + \mathcal{A}_0) \otimes (L' + \mathcal{A}_0)) + O_p(n^{-\frac{3}{2}}). \end{aligned}$$

Assim, obtem-se

$$\begin{aligned} \mathbb{E} [\theta_0^* - \theta] &= -\mathcal{X}_0 + \frac{1}{2} \mathbb{E} [H''(L' \otimes L')] + O(n^{-2}) \\ &= O(n^{-2}). \end{aligned}$$

Dado que $H' L''$ é igual à matriz identidade, derivando este produto em relação a α , obtemos $\left(\frac{\partial H'}{\partial L_\alpha} \frac{\partial L_\alpha}{\partial \alpha} + \frac{\partial H'}{\partial L_\beta} \frac{\partial L_\beta}{\partial \alpha} \right) L'' + H' \frac{\partial L''}{\partial \alpha} = 0$. Podemos representar essa igualdade em forma matricial

$$H' \frac{\partial L''}{\partial \alpha} = -H'' \begin{pmatrix} \frac{\partial L_\alpha}{\partial \alpha} L'' \\ \frac{\partial L_\beta}{\partial \alpha} L'' \end{pmatrix}.$$

Se derivarmos o produto $H' L''$ em relação a β teremos uma igualdade similar. Então, usando a definição de L''' e concatenando as duas igualdades em uma única matriz teremos

$$\begin{aligned} H' L''' &= H' \begin{pmatrix} \frac{\partial L''}{\partial \alpha} & \frac{\partial L''}{\partial \beta} \end{pmatrix} \\ &= -H'' \begin{pmatrix} \frac{\partial L_\alpha}{\partial \alpha} L'' & \frac{\partial L_\alpha}{\partial \beta} L'' \\ \frac{\partial L_\beta}{\partial \alpha} L'' & \frac{\partial L_\beta}{\partial \beta} L'' \end{pmatrix} \\ &= -H'' (L'' \otimes L''). \end{aligned}$$

Ou seja, $L''' = -L'' H'' (L'' \otimes L'')$.

O produto de Kronecker satisfaz uma propriedade na qual garante que $\text{vec}[L'' H' L''] = (L'' \otimes L'') \text{vec}[H']$, sendo que o operador “vec” de uma matriz A é definido como o vetor formado pelas colunas de A sobrepostas uma a uma. Assim,

temos que a translação da função escore proposta por David Firth, examinada na Seção 2.6, é exatamente a translação proposta nesta seção, pois

$$\begin{aligned}
\mathcal{A}_0 &= \frac{1}{2} L''' \text{vec}[H'] \quad (\text{translação proposta por Firth}) \\
&= -\frac{1}{2} L'' H'' (L'' \otimes L'') \text{vec}[H'] \\
&= -\frac{1}{2} L'' H'' \text{vec}[L''] \\
&= \frac{1}{2} L'' H'' \mathbf{E}[L' \otimes L'] \quad (\text{translação proposta nesta seção}).
\end{aligned}$$

Na última igualdade, usamos o fato de que $\mathbf{E}[L' \otimes L'] = -\text{vec}[L'']$.

Podemos continuar a transladar a função escore; desta forma, considere a expansão de Taylor de θ_0^* em torno de $L'(\theta)$ até ordem $O_p(n^{-\frac{5}{2}})$. Assim,

$$\begin{aligned}
\theta_0^* &= \theta - H'_0 L'_0 + \frac{1}{2} H''_0 (L'_0 \otimes L'_0) - \frac{1}{6} H'''_0 (L'_0 \otimes L'_0 \otimes L'_0) + \\
&+ \frac{1}{24} H^{(iv)}_0 (L'_0 \otimes L'_0 \otimes L'_0 \otimes L'_0) + O_p(n^{-\frac{5}{2}}).
\end{aligned}$$

De maneira análoga, define-se $EO(2)$ como a soma dos elementos dessa expansão cuja soma tem valor esperado com ordem de magnitude $O(n^{-2})$ e $\mathcal{X}_1 = \mathbf{E}[EO(2)]$. Logo, θ_1^* que é zero da função $L'_0 + \mathcal{A}_1$, com $\mathcal{A}_1 = L''_0 \mathcal{X}_1$, é um estimador não viesado de θ até a ordem $O(n^{-3})$. Observe que

$$\begin{aligned}
\mathcal{X}_1 &= -\mathcal{R}e(0) \mathcal{A}_0 + \frac{1}{2} H'' \mathcal{A}_0 \otimes \mathcal{A}_0 + \frac{1}{2} \mathcal{R}e'(0) \mathbf{E}[L' \otimes L'] - \frac{1}{6} H''' \mathbf{E}[L' \otimes L' \otimes L' + \\
&+ 3L' \otimes L' \otimes \mathcal{A}_0] + \frac{1}{24} H^{(iv)} \mathbf{E}[L' \otimes L' \otimes L' \otimes L'].
\end{aligned}$$

Perceba que $H_1^{(l)} = H_0^{(l)} + \mathcal{R}e^{(l)}(1)$ e como a ordem de magnitude do termo $\mathcal{R}e^{(l)}(1)$ é no máximo $O_p(n^{-4})$, podemos dispensá-lo. Assim, $H_1^{(l)} = H^{(l)} + \mathcal{R}e^{(l)}(0)$; no caso em que $l \geq 3$, também, podemos dispensar o termo $\mathcal{R}e^{(l)}(0)$.

Antes de calcular o viés de θ_1^* observemos que

$$\begin{aligned}
\mathbf{E}[H'_1 L'_1] &= \mathbf{E}[(H'_0 + \mathcal{R}e(1))(L' + \mathcal{A}_0 + \mathcal{A}_1)] + O(n^{-3}) \\
&= (H' + \mathcal{R}e(0))(L'' \mathcal{X}_0 + L''_0 \mathcal{X}_1) + O(n^{-3}) \\
&= \mathcal{X}_0 + \mathcal{R}e(0) \mathcal{A}_0 + \mathcal{X}_1 + O(n^{-3}) \\
&= \mathcal{R}e(0) \mathcal{A}_0 + \mathcal{X}_0 + \mathcal{X}_1 + O(n^{-3}).
\end{aligned}$$

Da mesma forma

$$\begin{aligned}
 \mathbb{E}[H_1''(L'_1 \otimes L'_1)] &= \mathbb{E}[H_0''(L'_0 \otimes L'_0)] + O(n^{-3}) \\
 &= \mathbb{E}[(H'' + \mathcal{R}'e(0))((L' + \mathcal{A}_0) \otimes (L' + \mathcal{A}_0))] + O(n^{-3}) \\
 &= H''\mathbb{E}[L' \otimes L'] + H''(\mathcal{A}_0 \otimes \mathcal{A}_0) + \mathcal{R}'e(0)\mathbb{E}[L' \otimes L'] + O(n^{-3}) \\
 &= 2\mathcal{X}_0 + H''(\mathcal{A}_0 \otimes \mathcal{A}_0) + \mathcal{R}'e(0)\mathbb{E}[L' \otimes L'] + O(n^{-3}).
 \end{aligned}$$

Similarmente, temos

$$\mathbb{E}[H_1'''(L'_1 \otimes L'_1 \otimes L'_1)] = H'''\mathbb{E}[L' \otimes L' \otimes L' + 3L' \otimes L' \otimes \mathcal{A}_0] + O(n^{-3}).$$

Também é fácil verificar, que a menos dos termos de ordem $O(n^{-3})$, é válida a igualdade $\mathbb{E}[H^{(iv)}(L'_1 \otimes L'_1 \otimes L'_1 \otimes L'_1)] = H^{iv}\mathbb{E}[L' \otimes L' \otimes L' \otimes L']$.

Portanto,

$$\begin{aligned}
 \mathbb{E}[\theta_1^* - \theta] &= \mathbb{E}\left[-H'_1L'_1 + \frac{1}{2}H_1''(L'_1 \otimes L'_1) - \frac{1}{6}H_1'''(L'_1 \otimes L'_1 \otimes L'_1) + \right. \\
 &\quad \left. + \frac{1}{24}H_1^{(iv)}(L'_1 \otimes L'_1 \otimes L'_1 \otimes L'_1) + O_p(n^{-\frac{5}{2}})\right] \\
 &= -\mathcal{R}e_0\mathcal{A}_0 - \mathcal{X}_0 - \mathcal{X}_1 + \mathcal{X}_0 + H''(\mathcal{A}_0 \otimes \mathcal{A}_0) + \mathcal{R}e(0)\mathbb{E}[L' \otimes L'] + \\
 &\quad + \mathbb{E}\left[-\frac{1}{6}H''''(L' \otimes L' \otimes L' + 3L' \otimes L' \otimes \mathcal{A}_0)\right] + \\
 &\quad + \mathbb{E}\left[\frac{1}{24}H^{(iv)}(L' \otimes L' \otimes L' \otimes L')\right] + O_p(n^{-3}) \\
 &= O_p(n^{-3}).
 \end{aligned}$$

Observe que, por definição, a soma dos quatro últimos termos dessa soma mais o primeiro termo é exatamente \mathcal{X}_1 .

Assim, como na Seção 3.3, apresentamos o Teorema 3.4.1 no qual garantimos no espaço biparamétrico a construção de um estimador θ_h^* não viesado até a ordem $O(n^{-(h+2)})$, para $k = 0, 1, 2, \dots$. Este estimador também é definido como θ_h^* que é zero da função escore L'_h transladada k -vezes.

Teorema 3.4.1 *Seja uma amostra $X = (X_1, \dots, X_n)$ independente e identicamente distribuída tal que a função escore satisfaz as suposições **SB1** a **SB9**. Faça $G'_h = L'$ na equação (3.6); então, considere a expansão do EMV dada*

nesta equação até $k = 2$ e defina \mathcal{X}_0 como a esperança da soma dos elementos dessa expansão cuja soma tem valor esperado $O(n^{-1})$. Então, θ_0^* que é zero da função $L'_0 = L' + A_0$, com $A_0 = L''\mathcal{X}_0$, é um estimador não viesado até ordem $O(n^{-2})$. Da mesma forma, considere a expansão de θ_0^* como na equação (3.6), com $G'_0 = L'_0$, até $k = 4$ e defina \mathcal{X}_1 como a esperança da soma dos elementos dessa expansão cuja soma tem valor esperado $O(n^{-2})$. Então, θ_1^* que é zero da função $L'_1 = L' + A_0 + A_1$, com $A_1 = L''_0\mathcal{X}_1$, é não viesado até ordem $O(n^{-3})$. Sucessivamente, considere a expansão de θ_{h-1}^* como na equação (3.6), com $G'_{h-1} = L'_{h-1}$, até $k = 2h + 2$ e defina \mathcal{X}_h como a soma dos elementos dessa expansão cuja soma tem seu valor esperado $O(n^{-(h+1)})$. Então, θ_h^* que é zero da função $L'_h = L' + A_0 + A_1 + \dots + A_h$, com $A_h = L''_{h-1}\mathcal{X}_h$, é não viesado até ordem $O(n^{-(h+2)})$.

Dem. Novamente usaremos indução matemática nesta prova. Para $h = 0$ e $h = 1$, já provamos ser verdadeira esta afirmação. Suponha, por hipótese de indução, que este teorema é válido para $h - 1$, então, vamos provar que essa teoria vale para h qualquer. Observe que pelo Lema 3.4.1 temos $H'_h = H'_{h-1} + \mathcal{R}e(h) + O_p(n^{-(2h+3)})$ e pelo Lema 3.4.2 temos $H_h^{(l)} = H_{h-1}^{(l)} + O_p(n^{-(h+l+2)})$. A menos dos termos de ordem $O_p(n^{-\frac{2h+3}{2}})$, podemos escrever a expansão de Taylor de θ_h^* da seguinte forma:

$$\theta_h^* - \theta = -H'_h L'_h + \dots + \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} L'_h \otimes^{2h+2}, \quad (3.7)$$

em que, por abuso de notação, $L'_h \otimes^k = L'_h \underbrace{\otimes \dots \otimes}_{k-\text{vezes}} L'_h$. Vamos dividir essa equação em três partes, a primeira $P_0 = -H'_h A_h$, a segunda

$$P_1 = -H'_{h-1} L'_{h-1} + \frac{1}{2} H''_{h-1} L'_{h-1} \otimes^2 + \dots + \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} L'_{h-1} \otimes^{2h+2},$$

que por hipótese de indução tem valor esperado com ordem de magnitude $O(n^{-(h+1)})$.

Portanto, $\mathcal{X}_h = P_1$, logo $A_h = L''_{h-1} P_1$. A terceira parte é o que resta da equação (3.9) após a retirada das duas primeiras partes:

$$\begin{aligned} P_2 &= \frac{1}{2} H''_{h-1} (L'_h \otimes^2 - L'_{h-1} \otimes^2) + \dots + \\ &+ \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} (L'_h \otimes^{2h+2} - L'_{h-1} \otimes^{2h+2}). \end{aligned}$$

O valor esperado de P_2 é $O(n^{-(h+2)})$. Para verificarmos esse fato, observe que o produto de Kronecker satisfaz as propriedades associativas e distributivas, assim, temos que $(L'_{h-1} + A_h) \otimes (L'_{h-1} + A_h) = L'_{h-1} \otimes L'_{h-1} + 2A_h \otimes L'_{h-1} + A_h \otimes A_h$. Como o valor esperado de L'_{h-1} é $O(1)$, então,

$$\mathbb{E} [(L'_{h-1} + A_h) \otimes (L'_{h-1} + A_h)] = \mathbb{E} [L'_{h-1} \otimes L'_{h-1}] + O(n^{-h}).$$

Portanto, a esperança do primeiro termo de P_2 tem ordem $O(n^{-(h+2)})$. Suponha, por hipótese de indução que para algum l , vale a igualdade $\mathbb{E} [(L'_{h-1} + A_h)^{\otimes l}] = \mathbb{E} [L'_{h-1}^{\otimes l}] + O(n^{-\frac{2h-l+1}{2}})$. Então,

$$\begin{aligned} \mathbb{E} [(L'_{h-1} + A_h)^{\otimes l+1}] &= \mathbb{E} \left[\sum_{i=0}^{l+1} C_i^{l+1} (L'_{h-1}^{\otimes l+1-i}) \otimes (A_h^{\otimes i}) \right] \\ &= \mathbb{E} [L'_{h-1}^{\otimes l+1}] + \sum_{i=1}^{l+1} C_i^{l+1} \mathbb{E} [L'_{h-1}^{\otimes l+1-i}] \otimes (A_h^{\otimes i}) \\ &= \mathbb{E} [L'_{h-1}^{\otimes l+1}] + O(n^{-\frac{2h-l}{2}}). \end{aligned}$$

Ou seja, a igualdade vale para $l+1$. Assim, o valor esperado de um termo da expressão de P_2 tem magnitude $\mathbb{E} [H_{h-1}^{(l)} (L'_h \otimes^l - L'_{h-1} \otimes^l)] = O(n^{-l})O(n^{-\frac{2h-l+1}{2}})$. Como $l > 2$, temos que este valor esperado é no máximo $O(n^{-(h+2)})$. Portanto, como $\mathcal{R}e(h)A_h$ tem ordem maior que $O_p(n^{-\frac{2h+3}{2}})$, temos que

$$\begin{aligned} \mathbb{E} [\theta_h^* - \theta] &= \mathbb{E} [-(H'_{h-1} + \mathcal{R}e(h)) A_h] + \mathbb{E} [P_1] + \mathbb{E} [P_2] + \mathbb{E} [O_p(n^{-\frac{2h+3}{2}})] \\ &= -P_1 + P_1 + P_2 + \mathbb{E} [O_p(n^{-\frac{2h+3}{2}})] \\ &= O(n^{-(h+2)}). \blacksquare \end{aligned}$$

3.5 Estimação Multiparamétrica

Consideraremos agora o caso multiparamétrico $\theta = (\theta_1, \dots, \theta_p)^T$, como na Seção 2.5. Sejam Θ o conjunto de todos os possíveis valores de θ e G uma função com domínio em Θ , então estabeleceremos as seguintes suposições de regularidade multiparamétricas:

- **SM1** As derivadas parciais de G são contínuas.

- **SM2** O gradiente G' tem exatamente uma raiz θ^* e a matriz $G''(\theta^*)$ é negativa definida.
- **SM3** $E[G'G'^T]$ tem todos os elementos finitos.
- **SM4** $G' = O_p(n^{\frac{1}{2}})$.
- **SM5** Se uma função polinomial das derivadas parciais $f(\frac{\partial G}{\partial \theta_1}, \frac{\partial^2 G}{\partial \theta_1 \partial \theta_2}, \dots)$ tem ordem de magnitude $O_p(n^k)$, então, tanto $E[f(\frac{\partial G}{\partial \theta_1}, \frac{\partial^2 G}{\partial \theta_1 \partial \theta_2}, \dots)]$ como a derivada dessa esperança têm ordem de magnitude $O(n^k)$.
- **SM6** Os elementos da matriz $(G'')^{-1}$ são $O_p(n^{-1})$ em uma vizinhança de θ que inclui θ^* .
- **SM7** Os elementos das matrizes $G'', G''', \dots, G^{(k)}$ são $O_p(n)$ em uma vizinhança de θ que inclui θ^* .

Lembre que no caso biparamétrico utilizamos a suposição de que a função L'' não dependia dos dados. Agora estamos generalizando tanto a dimensão do espaço paramétrico como estamos abrindo mão da não-aleatoriedade da derivada da função escore.

Similarmente ao caso biparamétrico, considere que a função $G'_k = (G_{k\theta_1}, \dots, G_{k\theta_p})$ satisfaz as suposições acima e seja H_k sua inversa, ou seja,

$$H_k : \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{p-\text{vezes}} \longrightarrow \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{p-\text{vezes}} \\ (\dots, G_{k\theta_i}(\theta_1, \dots, \theta_p), \dots)^T \longmapsto (\dots, H_k^i(G_{k\theta_1}, \dots, G_{k\theta_p}), \dots)^T,$$

com $i = 1, \dots, p$, tal que, $(H_k^1(G_{k\theta_1}, \dots, G_{k\theta_p}), \dots, H_k^p(G_{k\theta_1}, \dots, G_{k\theta_p}))^T = (\theta_1, \dots, \theta_p)^T$. Em outras palavras, temos que $H_k(G'_k(\theta)) = \theta$.

Seja θ_h^* zero da função G'_h , então, a expansão de Taylor de θ_h^* em torno de $G'_h(\theta)$ é dada por

$$\theta_h^* = \theta - H'_h G'_h + \frac{1}{2} H''_h (G'_h \otimes G'_h) + \dots + \frac{(-1)^k}{k!} H_h^{(k)} (G'_h \underbrace{\otimes \dots \otimes}_{k-\text{vezes}} G'_h) + r_k, \quad (3.8)$$

sendo $|r_k| \leq |G'(\theta)|$.

Faremos sucessivas translações na função G'_k e desejamos que ela continue a satisfazer as suposições acima, dessa forma, acrescentamos a seguinte suposição:

- **SM8** Sejam G'_h uma função que satisfaz as suposições **SM1** a **SM7** e θ_h^* zero desta função, considere a expansão de θ_h^* como na equação (3.8) e defina $EO(h+2)$ como a esperança da soma dos elementos dessa expansão cuja soma tem valor esperado com ordem de magnitude $O(n^{-(h+2)})$. Então, a função $G'_{h+1} = G'_h + A_{h+1}$, sendo $A_{h+1} = E[H'_h]^{-1} E[EO(h+2)]$, satisfaz as suposições **SM1** a **SM7**. Além disso, tanto A_{h+1} como todas suas derivadas são de ordem $O_p(n^{-(h+1)})$, com $h \geq 0$.

O determinante de uma matriz tem ordem de magnitude relacionada com a dimensão do espaço paramétrico e com a magnitude de seus elementos; o próximo lema estabelece esta relação. Para demonstrar este lema precisamos da seguinte definição: dada uma matriz quadrada M , definimos a matriz $M_{(-i,-j)}$ como a matriz M sem a linha i e sem a coluna j .

Lema 3.5.1 *Seja A'_k uma matriz $p \times p$ tal que cada elemento $a_{k(i,j)}$ de A'_k tem ordem $O_p(n^{-k})$. Então, o determinante de A'_k tem ordem de magnitude $O_p(n^{-kp})$.*
Dem. Para $p = 1$ e $p = 2$, esta igualdade é trivialmente satisfeita. Suponha válida para $p - 1$. Fixada uma coluna j , temos que o determinante de A'_k é dado por

$$|A'_k| = \sum_{i=1}^p (-1)^{i+j} a_{k(i,j)} |A'_{k(-i,-j)}|.$$

Por hipótese de indução $|A'_{k(-i,-j)}|$ tem ordem $O_p(n^{-k(p-1)})$, logo cada parcela desta soma tem ordem $O_p(n^{-kp})$. Como p é constante, temos o resultado desejado.

■

A proposta desta tese é fazer sucessivas translações na função escore para eliminar vieses com magnitudes grandes. O Lema 3.5.2 estabelece como se relaciona o determinante de uma matriz G''_{k-1} com a matriz transladada $G''_k = G''_{k-1} + A'_k$. Já o Lema 3.5.3 estabelece qual a relação entre essas matrizes.

Lema 3.5.2 Sejam G''_{k-1} e A'_k duas matrizes $p \times p$ tais que cada elemento $g_{(k-1)(i,j)}$ de G''_{k-1} tem ordem $O_p(n)$ e cada elemento de $a_{k(i,j)}$ de A'_k tem ordem $O_p(n^{-k})$. Então, o determinante da matriz $G''_k = G''_{k-1} + A'_k$ é dado por

$$|G''_k| = |G''_{k-1}| + \mathcal{R}(k),$$

em que $\mathcal{R}(k) = \sum_{i=1}^p (-1)^{i+j} \left\{ g_{(k-1)(i,j)} \left| A'_{k(-i,-j)} \right| + a_{k(i,j)} \left| G''_{(k-1)(-i,-j)} + A'_{k(-i,-j)} \right| \right\}$ e tem ordem $O_p(n^{-(k-p+1)})$.

Dem. Para $p = 1$ trivialmente esta igualdade é válida. Para $p = 2$, temos que

$$\begin{aligned} |G''_k| &= g_{k(1,1)}g_{k(2,2)} - g_{k(1,2)}g_{k(2,1)} \\ &= (g_{(k-1)(1,1)} + a_{k(1,1)})(g_{(k-1)(2,2)} + a_{k(2,2)}) \\ &\quad - (g_{(k-1)(1,2)} + a_{k(1,2)})(g_{(k-1)(2,1)} + a_{k(2,1)}) \\ &= g_{(k-1)(1,1)}g_{(k-1)(2,2)} - g_{(k-1)(1,2)}g_{(k-1)(2,1)} + g_{(k-1)(1,1)}a_{k(2,2)} \\ &\quad + a_{k(1,1)}(g_{(k-1)(2,2)} + a_{k(2,2)}) - g_{(k-1)(1,2)}a_{k(2,1)} - a_{k(1,2)}(g_{(k-1)(2,1)} + a_{k(2,1)}) \\ &= |G''_{(k-1)}| + \mathcal{R}(k), \end{aligned}$$

em que $\mathcal{R}(k) = g_{(k-1)(1,1)}a_{k(2,2)} + a_{k(1,1)}(g_{(k-1)(2,2)} + a_{k(2,2)}) - g_{(k-1)(1,2)}a_{k(2,1)} - a_{k(1,2)}(g_{(k-1)(2,1)} + a_{k(2,1)})$ e tem ordem de magnitude $O_p(n^{-(k-1)})$. Suponha a igualdade válida para $p - 1$. Então, para um j qualquer fixo, o determinante de G''_k é dado por

$$\begin{aligned} |G''_k| &= \sum_{i=1}^p (-1)^{i+j} g_{k(i,j)} \left| G''_{k(-i,-j)} \right| \\ &= \sum_{i=1}^p (-1)^{i+j} (g_{(k-1)(i,j)} + a_{k(i,j)}) \left| G''_{(k-1)(-i,-j)} + A'_{k(-i,-j)} \right| \\ &= |G''_{k-1}| + \mathcal{R}(k), \end{aligned}$$

em que $\mathcal{R}(k) = \sum_{i=1}^p (-1)^{i+j} \left\{ g_{(k-1)(i,j)} \left| A'_{k(-i,-j)} \right| + a_{k(i,j)} \left| G''_{(k-1)(-i,-j)} + A'_{k(-i,-j)} \right| \right\}$ e tem ordem de magnitude $O_p(n^{-(k-p+1)})$, pois o termo com maior ordem é $a_{k(i,j)} \left| G''_{(k-1)(-i,-j)} \right| = O(n^{-k})O(n^{p-1})$. ■

Dada uma matriz quadrada M , definimos $\mathcal{M}(M)$ como a transposta da matriz formada pelos cofatores de M . Portanto, temos do teorema da função inversa que H'_k é igual a inversa de L''_k , ou seja,

$$H'_k = \frac{\mathcal{M}(L''_k)}{|L''_k|}.$$

Lema 3.5.3 *Considere a sequência de translação da função escore, $L'_k = L'_{k-1} + A_k$, com $L'_0 = L' + A_0$ e A_k um vetor p -dimensional com ordem de magnitude $O_p(n^{-k})$. Então, a matriz H'_k pode ser decomposta em parcelas da seguinte forma:*

$$H'_k = H'_{k-1} + \mathcal{R}e_1(k) + \mathcal{R}e_2(k) + O_p(n^{-kp-p-1}),$$

com

$$\begin{aligned} \mathcal{R}e_1(k) &= H'_{k-1} \frac{|L''_k| - |L''_{k-1}|}{|L''_k|}, \\ \mathcal{R}e_2(k) &= \frac{\mathcal{M}(A_k)}{|L''_{k-1}|}, \end{aligned}$$

em que $\mathcal{R}e_1(k)$ e $\mathcal{R}e_2(k)$ têm ordem de magnitude $O_p(n^{-(k+2)})$ e $O_p(n^{-p(k+1)})$, respectivamente.

Dem. Para $p = 1$, esta igualdade é válida pelo Lema 3.2.2. Suponha a igualdade válida para $p - 1$. O elemento de posição (i, j) da matriz de cofatores de L''_k é da forma

$$(-1)^{i+j} |L''_{k(-i,-j)}| = (-1)^{i+j} |L''_{(k-1)(-i,-j)} + A_{k(-i,-j)}|.$$

Sabemos que $\frac{1}{|L''_k|} = \frac{1}{|L''_{k-1}|} - \frac{|L''_k| - |L''_{k-1}|}{|L''_k||L''_{k-1}|}$, então, como a soma de uma matriz transposta é a matriz transposta da soma dessas matrizes, ao dividirmos a transposta da matriz de cofatores de L''_k por $|L''_k|$ temos que o elemento de posição (j, i) da matriz H'_k é

$$\begin{aligned} h_{k(j,i)} &= (-1)^{i+j} |L''_{(k-1)(-i,-j)} + A_{k(-i,-j)}| \left[\frac{1}{|L''_{k-1}|} - \frac{|L''_k| - |L''_{k-1}|}{|L''_k||L''_{k-1}|} \right] \\ &= h_{(k-1)(j,i)} - h_{(k-1)(j,i)} \frac{|L''_k| - |L''_{k-1}|}{|L''_k|} + \frac{(-1)^{i+j} |A_{k(-i,-j)}|}{|L''_{k-1}|} + O_p(n^{-kp-p-1}). \end{aligned}$$

Portanto,

$$H'_k = H'_{k-1} + H'_{k-1} \frac{|L''_k| - |L''_{k-1}|}{|L''_k|} + \frac{\mathcal{M}(A_k)}{|L''_{k-1}|} + O_p(n^{-kp-p-1}).$$

Observe que

$$H'_{k-1} \frac{|L''_k| - |L''_{k-1}|}{|L''_k|} = O_p(n^{-(k+2)}),$$

$$\frac{\mathcal{M}(A_k)}{|L''_{k-1}|} = O_p(n^{-k(p-1)-p})$$

e

$$|A_{k(-i,-j)}| \frac{|L''_k| - |L''_{k-1}|}{|L''_k||L''_{k-1}|} = O_p(n^{-kp-p-1}).$$

■

Na expansão de Taylor dada na equação (3.8) aparecem derivadas de ordem l da função H . Assim, o lema seguinte afirma qual a ordem de magnitude dessas derivadas e das parcelas de suas translações.

Lema 3.5.4 *Seja L'_k uma função satisfazendo as suposições **SM1** a **SM8**, então, a ordem de magnitude da l -ésima derivada das funções:*

- a) H_k é $O_p(n^{-l})$.
- b) $\mathcal{R}e_1(k)$ é $O_p(n^{-(k+2+l)})$.
- c) $\mathcal{R}e_2(k)$ é $O_p(n^{-k(p-1)-p-l})$.

Dem. Sem perda de generalidade, considere a derivada de uma dessas funções em relação a coordenada L_{θ_1} e o primeiro termo de H''_k . Ou seja,

$$h''_{k(1,1)} = \frac{\partial h'_{k(1,1)}}{\partial L_{\theta_1}} = \sum_{i=1}^p \frac{\partial h'_{k(1,1)}}{\partial H_k^i} \frac{\partial H_k^i}{\partial L_{\theta_1}} = O_p(n^{-2}).$$

Suponha por indução que a igualdade vale para $l-1$. Logo o termo de posição $(1,1)$ de $H_k^{(l)}$ é dado por

$$\begin{aligned} h_{k(1,1)}^{(l)} &= \frac{\partial h_{k(1,1)}^{l-1}}{\partial L_{\theta_1}} = \sum_{i=1}^p \frac{\partial h_{k(1,1)}^{(l-1)}}{\partial H_k^i} \frac{\partial H_k^i}{\partial L_{\theta_1}} \\ &= \sum_{i=1}^p O_p(n^{-(l-1)}) O_p(n^{-1}) = O_p(n^{-l}). \end{aligned}$$

Os itens b e c são idênticos na demonstração. ■

O principal resultado desta tese de doutorado é apresentado no Teorema 3.5.1, no qual propomos uma sequência de translações da função escore, de forma que o zero de cada função transladada é um estimador com viés corrigido.

Teorema 3.5.1 *Dada uma amostra $X = (X_1, \dots, X_n)$ independente e identicamente distribuída, tal que a função escore satisfaçõas as suposições **SM1** a **SM8**. Faça $G'_h = L'$ na equação (3.8), considere a expansão do EMV dada nesta equação até $k = 2$ e defina \mathcal{X}_0 como a esperança da soma dos elementos dessa expansão cuja soma tem valor esperado $O(n^{-1})$. Então, θ_0^* , que é zero da função $L'_0 = L' + A_0$, com $A_0 = \mathbb{E}[H']^{-1} \mathcal{X}_0$, é não viesado até ordem $O(n^{-2})$. Da mesma forma, considere a expansão de θ_0^* como na equação (3.8), com $G'_0 = L'_0$, até $k = 4$ e defina \mathcal{X}_1 como a esperança da soma dos elementos dessa expansão cuja soma tem valor esperado $O(n^{-2})$. Então, θ_1^* que é zero da função $L'_1 = L' + A_0 + A_1$, com $A_1 = \mathbb{E}[H'_0]^{-1} \mathcal{X}_1$, é não viesado até ordem $O(n^{-3})$. Sucessivamente, considere a expansão de θ_{h-1}^* como na equação (3.8), com $G'_{h-1} = L'_{h-1}$, até $k = 2h + 2$ e defina \mathcal{X}_h como a esperança da soma dos elementos dessa expansão cuja soma tem seu valor esperado $O(n^{-(h+1)})$. Então, θ_h^* que é zero da função $L'_h = L' + A_0 + A_1 + \dots + A_h$, com $A_h = \mathbb{E}[H'_{h-1}]^{-1} \mathcal{X}_h$, é não viesado até ordem $O(n^{-(h+2)})$.*

Dem. Novamente usaremos indução matemática nesta prova. Para $h = 0$, temos que $\mathcal{X}_0 = \mathbb{E}[-H'L' + \frac{1}{2}H''(L'_0 \otimes L'_0)]$. Então, o viés de θ_0^* é dado por

$$\begin{aligned} \mathbb{E}[\theta_0^* - \theta] &= \mathbb{E}\left[-H'_0 L'_0 + \frac{1}{2}H''(L'_0 \otimes L'_0) + O_p(n^{-\frac{3}{2}})\right] \\ &= \mathbb{E}\left[(-H' + \mathcal{R}e_1(0))(L' + A_0) + \frac{1}{2}H''(L' \otimes L') + O_p(n^{-\frac{3}{2}})\right] \\ &= \mathbb{E}\left[(-H')\left(L' + \mathbb{E}[H']^{-1} \mathcal{X}_0\right) + \frac{1}{2}H''(L' \otimes L') + O_p(n^{-\frac{3}{2}})\right] \\ &= O(n^{-2}). \end{aligned}$$

Usamos que $\mathbb{E}[O_p(n^{-\frac{3}{2}})] = O(n^{-2})$, devido à Afirmacão 3.3.1.

Suponha por hipótese de indução que este teorema é válido para $h - 1$, então, vamos provar que essa teoria vale para h qualquer. Observe que pelo Lema 3.5.3 temos $H'_h = H'_{h-1} + \mathcal{R}e_1(h) + \mathcal{R}e_2(h) + O_p(n^{-kp-p-1})$ e pelo Lema 3.5.4 temos $H_h^{(l)} = H_{h-1}^{(l)} + O_p(n^{-(h+l+2)})$. A menos dos termos de ordem $O_p(n^{-\frac{2h+3}{2}})$, podemos escrever a expansão de Taylor de θ_h^* da forma

$$\theta_h^* - \theta = -H'_h L'_h + \dots + \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} L'_h \otimes^{2h+2}. \quad (3.9)$$

Dividiremos essa equação em três partes: a primeira $P_0 = -H'_h A_h$, a segunda

$$P_1 = -H'_{h-1} L'_{h-1} + \frac{1}{2} H''_{h-1} L'_{h-1} \otimes^2 + \cdots + \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} L'_{h-1} \otimes^{2h+2},$$

que por hipótese de indução tem valor esperado com ordem de magnitude $O(n^{-(h+1)})$. Portanto, $\mathcal{X}_h = E[P_1]$, logo $A_h = E[H'_{h-1}]^{-1} E[P_1]$. A terceira parte é o que resta da equação (3.9), após a retirada das duas primeiras partes

$$\begin{aligned} P_2 &= \frac{1}{2} H''_{h-1} (L'_h \otimes^2 - L'_{h-1} \otimes^2) + \cdots + \\ &+ \frac{1}{(2h+2)!} H_{h-1}^{(2h+2)} (L'_h \otimes^{2h+2} - L'_{h-1} \otimes^{2h+2}). \end{aligned}$$

Assim como no caso bidimensional, temos que o valor esperado de P_2 é $O(n^{-(h+2)})$. Portanto, como $\mathcal{R}e_1(h) A_h$ tem ordem maior que $O_p(n^{-\frac{2h+3}{2}})$, temos que

$$\begin{aligned} E[\theta_h^* - \theta] &= E[-(H'_{h-1} + \mathcal{R}e(h)) A_h] + E[P_1] + E[P_2] + E\left[O_p(n^{-\frac{2h+3}{2}})\right] \\ &= E[-P_1 + P_1 + P_2] + E\left[O_p(n^{-\frac{2h+3}{2}})\right] \\ &= O(n^{-(h+2)}). \blacksquare \end{aligned}$$

Em outras palavras, dada uma amostra de variáveis aleatórias, na qual a função escore satisfaz as suposições **SM1** a **SM8**, é possível encontrar um estimador cujo viés tem ordem de magnitude $O(n^{-k})$, para $k = 2, 3, \dots$. Isto é, propomos uma nova alternativa para encontrar EMVs com viés de ordem de magnitude tão pequena quanto se deseja.

3.6 Considerações Finais

Foi proposta uma técnica de estimação nova, em que a partir de uma sequência de translações na função escore, obtemos um EMVC com ordem de magnitude tão pequena quanto se queira.

Uma grande vantagem dessa metodologia é que não há uma restrição muito forte nas classes de modelos em que podemos empregar tal técnica. Basta apenas que a função de verossimilhança satisfaça algumas suposições de regularidade, que são bastante gerais e satisfeitas por um grande número de modelos estatísticos.

Nesta metodologia precisamos apenas encontrar o zero da função escore modificada, desta forma torna-se uma técnica muito promissora para ser desenvolvida em outros métodos de estimação, não necessariamente em estimação por máxima verossimilhança.

Capítulo 4

AVALIAÇÃO NUMÉRICA

4.1 Introdução

No presente trabalho apresentamos uma técnica de estimação baseada em corrigir, de forma preventiva, o EMV. Esta técnica consiste em fazer sucessivas translações na função escore, de tal forma que o zero da função escore transladada é um estimador cujo viés tem ordem de magnitude reduzida. A primeira translação é a sugerida por David Firth; assim, obtemos uma sequência de estimadores sendo que o primeiro elemento da sequência é o EMV $\widehat{\theta}$, o segundo é o estimador θ_0^* e o terceiro elemento é θ_1^* , estes dois últimos têm vieses com ordem de magnitude $O(n^{-2})$ e $O(n^{-3})$, respectivamente. Outros dois estimadores com vieses corrigidos foram estudados no Capítulo 2, que são $\widehat{\theta}_1$, com viés de ordem de magnitude $O(n^{-2})$, e $\widetilde{\theta}_2$, com viés $O(n^{-3})$.

Realizamos neste capítulo um estudo aprofundado do desempenho destes quatro estimadores, mais o EMV. Foram realizadas simulações de Monte Carlo em vários modelos estatísticos, com diferentes tamanhos amostrais. A metodologia empregada para avaliar os estimadores foi baseada em comparar a média dos vieses estimados. Ou seja, em cada réplica de Monte Carlo, calculamos a estimativa do parâmetro, segundo cada estimador, no final da simulação calculamos a médias dessas estimativas e a subtraímos do valor verdadeiro do parâmetro. Essas simulações foram implementadas na linguagem de programação matricial Ox. Em muitos casos estudados, para encontrar o zero da função escore ou de

alguma de suas translações, existiu uma grande complexidade algébrica, então, fizemos uso do cálculo numérico através da função computacional “`SolveNLE`.” Esta função está disponível na linguagem de programação Ox através do pacote “`solvenle`”.

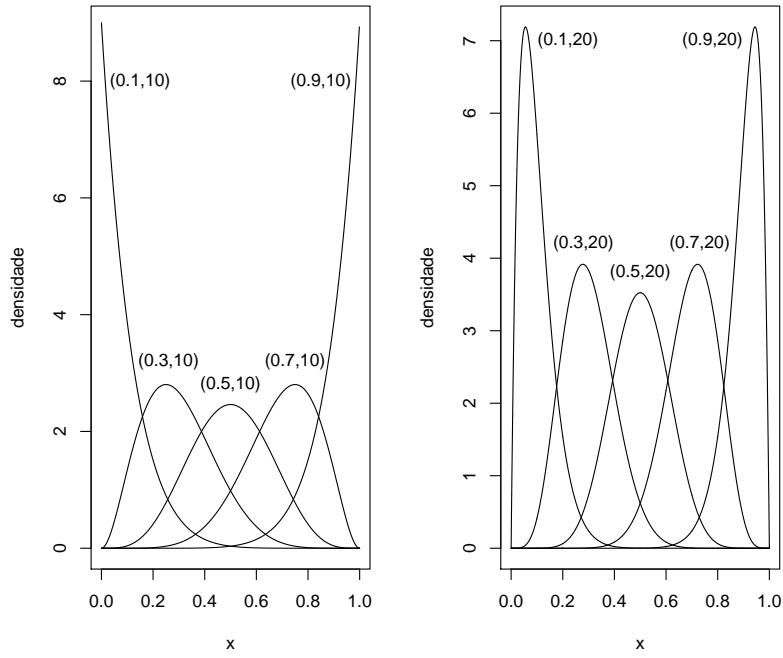
Na Seção 4.2 apresentamos os resultados das simulações feitas em modelos uniparamétricos, tais como, beta, gama, entre outros; a Seção 4.3 foi dedicada aos modelos biparamétricos; finalizamos o capítulo com a Seção 4.4, na qual colocamos as considerações finais sobre as avaliações numéricas realizadas.

4.2 Modelos Uniparamétricos

Com o intuito de avaliar o desempenho em pequenas amostras do EMVC até ordem $O(n^{-3})$, proposto na Seção 3.3, realizamos diversas simulações de Monte Carlo em uma grande variedade de modelos e de situações. Cada simulação foi composta de dez mil réplicas de Monte Carlo. Variamos as simulações com os modelos beta, gama, binomial, Poisson e geométrico. Nesta seção tratamos apenas do caso uniparamétrico; desta forma, caso um modelo seja caracterizado por mais de um parâmetro, consideramos sempre que apenas um dos parâmetros é desconhecido. Em cada modelo atribuímos diversos valores para o parâmetro e usando diversos estimadores foi calculado em cada réplica de Monte Carlo o valor estimado desses parâmetros; em seguida foi calculado o viés médio, ou seja, foi feita uma média da diferença entre o verdadeiro valor do parâmetro e seu valor estimado nas réplicas de Monte Carlo.

Consideramos em nossa avaliação o EMV $\widehat{\theta}$; o estimador proposto por David Firth θ_0^* , que é o zero da função escore transladada por $A_0 = \frac{L'''}{2T}$; mais os estimadores $\widehat{\theta}_1$ e $\widetilde{\theta}_2$, este último proposto por Ferrari et al., que são dados nas equações (2.18) e (2.19), respectivamente; por fim, o estimador θ_1^* proposto neste trabalho, o qual é obtido por encontrar o zero da função escore transladada por A_0 e por A_1 , definidos na Seção 3.3

Uma importante distribuição de probabilidade estudada neste trabalho é a distribuição beta. Uma variável aleatória X é dita ter distribuição beta com

Figura 4.1. Densidades da distribuição beta com parâmetros (μ, ϕ) .

parâmetros $p > 0$ e $q > 0$, se sua função de densidade é dada por

$$f(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}, \quad (4.1)$$

em que $0 < x < 1$ e $\Gamma(\cdot)$ é a função gama.

Uma reparametrização útil desta distribuição se dá ao definir $\mu = \frac{p}{p+q}$ e $\phi = p+q$. Desta forma o parâmetro μ é uma medida de locação, enquanto ϕ é uma medida de escala. Isto pode ser visto com mais facilidade na Figura 4.1, na qual está representado o gráfico desta densidade de probabilidade para vários valores de μ e ϕ .

Consideremos uma amostra de tamanho n , $x = x_1, \dots, x_n$, de variáveis aleatórias independentes e identicamente distribuídas, de acordo com a distribuição beta com parâmetros $p > 0$ e $q > 0$. A função de verossimilhança de p e q com base nesta amostra é dada por

$$f(p, q|x) = \prod_{i=1}^n f(x_i; p, q) = \left[\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \right]^n \prod_{i=1}^n x_i^{p-1} (1-x_i)^{q-1},$$

em que podemos escrever a função log-verossimilhança como

$$L(p, q) = n \log \Gamma(p+q) - n \log \Gamma(p) - n \log \Gamma(q)$$

$$+ \sum_{i=1}^n [(p-1) \log x_i + (q-1) \log(1-x_i)]. \quad (4.2)$$

Tabela 4.1. Vieses dos estimadores do parâmetro p da densidade beta, $n = 10$.

Parâmetro p	viés de $\hat{\theta}$ $O(n^{-1})$	viés de $\hat{\theta}_1$ $O(n^{-2})$	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$	viés de $\tilde{\theta}_2$ $O(n^{-3})$
0,50	0,037950	0,001252	0,000135	0,000696	0,001082
0,75	0,041520	-0,003327	-0,004582	-0,003746	-0,003421
1,00	0,047026	-0,003718	-0,005022	-0,003976	-0,003725
2,00	0,067623	-0,000583	-0,001481	-0,000583	-0,000445
3,00	0,085738	-0,000640	-0,001204	-0,000487	-0,000499
4,00	0,108000	-0,001852	-0,002307	-0,001715	-0,001725
5,00	0,148610	0,005796	0,005410	0,005917	0,005908
6,00	0,187100	-0,004991	-0,005328	-0,004883	-0,004891
7,00	0,287480	0,011278	0,010984	0,011374	0,011365
8,00	0,472380	0,023781	0,023546	0,023861	0,023850

Suponha que seja conhecido o parâmetro q e queiramos estimar p . A função log-verossimilhança e suas derivadas são dadas por

$$L(p) = n \log \Gamma(p+q) - n \log \Gamma(p) - n \log \Gamma(q) \\ + \sum_{i=1}^n [(p-1) \log x_i + (q-1) \log(1-x_i)],$$

$$L'(p) = n\psi(p+q) - n\psi(p) + \sum_{i=1}^n \log x_i,$$

$$L''(p) = n\psi'(p+q) - n\psi'(p),$$

.

.

.

$$L^{(k)}(p) = n\psi^{(k-1)}(p+q) - n\psi^{(k-1)}(p), \quad \text{se } k > 1.$$

Tabela 4.2. Vieses dos estimadores do parâmetro p da densidade beta, $n = 15$.

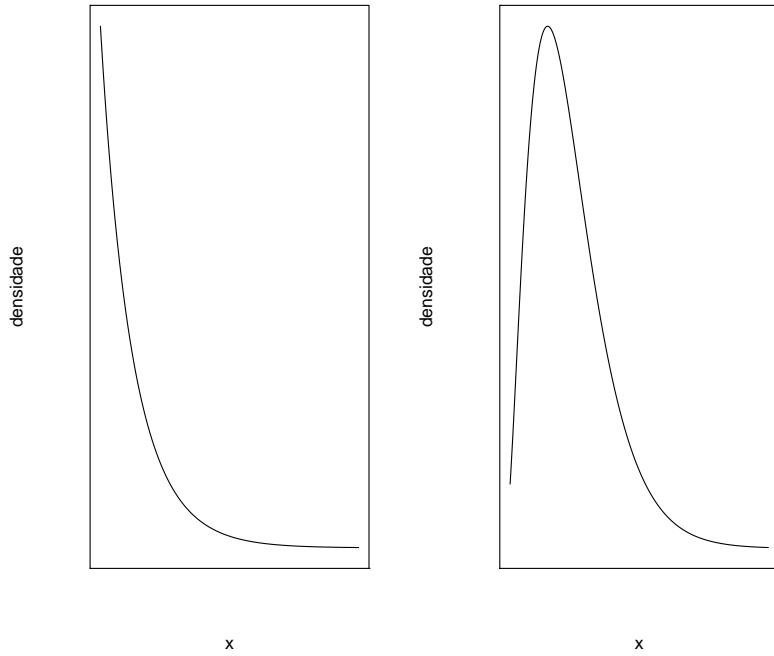
Parâmetro p	viés de $\hat{\theta}$ $O(n^{-1})$	viés de $\hat{\theta}_1$ $O(n^{-2})$	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$	viés de $\tilde{\theta}_2$ $O(n^{-3})$
0,50	0,022342	-0,001922	-0,002301	-0,002056	-0,002001
0,75	0,031352	0,001459	0,001034	0,001406	0,001415
1,00	0,030452	-0,003319	-0,003757	-0,003327	-0,003323
2,00	0,049168	0,003828	0,003499	0,003894	0,003890
3,00	0,056783	-0,000536	-0,000783	-0,000469	-0,000472
4,00	0,082390	0,009433	0,009234	0,009492	0,009490
5,00	0,090048	-0,004385	-0,004555	-0,004333	-0,004335
6,00	0,145980	0,018608	0,018461	0,018655	0,018653
7,00	0,191140	0,009150	0,009022	0,009192	0,009189
8,00	0,282350	-0,010375	-0,010477	-0,010342	-0,010345

Nas Tabelas 4.1 e 4.2 estão os resultados de várias simulações de Monte Carlo, nas quais fixamos o valor do parâmetro de locação de tal forma que $p + q = 10$, variamos o valor de p e avaliamos o comportamento dos estimadores supracitados em amostras com tamanhos $n = 10$ e $n = 15$. Podemos verificar que os quatro estimadores com vieses corrigidos são competitivos entre si e todos eles têm desempenho superior ao de máxima verossimilhança.

Outra distribuição de probabilidade de relevância é a densidade gama. Dizemos que uma variável aleatória X tem distribuição gama com parâmetros α e β , ambos positivos, se sua densidade é dada por

$$f(x|\alpha) = \frac{x^{\alpha-1} \beta^\alpha \exp(-\beta x)}{\Gamma(\alpha)},$$

com $x \geq 0$. Se o parâmetro α for menor ou igual um, o gráfico da densidade gama tem o formato como o do gráfico esquerdo da Figura 4.2, caso contrário, seu formato é como o do gráfico da direita.

Figura 4.2. Densidades da distribuição gama.

A função de verossimilhança é dada por

$$f(\alpha) = \Gamma(\alpha)^{-n} \beta^{n\alpha} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} \exp \left(-\beta \sum_{i=1}^n x_i \right).$$

Suponha que o parâmetro β seja conhecido e desejamos estimar α . Neste caso, a função de log-verossimilhança e suas derivadas são

$$\begin{aligned} L(\alpha) &= -n \log[\Gamma(\alpha)] + n\alpha \log(\beta) + (\alpha - 1) \sum_{i=1}^n \log(x_i) - \beta \sum_{i=1}^n x_i, \\ L'(\alpha) &= -n\psi(\alpha) + n \log(\beta) + \sum_{i=1}^n \log(x_i), \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ L^{(k)}(\alpha) &= -n\psi^{(k-1)}(\alpha), \quad \text{se } k > 1. \end{aligned}$$

Um resumo dos resultados das simulações de Monte Carlo para o modelo gama pode ser visto nas Tabelas 4.3 e 4.4. O parâmetro β foi considerado conhecido e de valor igual a um, o parâmetro α variou de 0.8 a 10. Utilizamos nessas simulações

Tabela 4.3. Vieses dos estimadores do parâmetro α da densidade gama, $n = 10$.

Parâmetro α	viés de $\hat{\theta}$ $O(n^{-1})$	viés de $\hat{\theta}_1$ $O(n^{-2})$	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$	viés de $\tilde{\theta}_2$ $O(n^{-3})$
0,80	0,037953	-0,003702	-0,005494	-0,004584	-0,003795
1,00	0,043309	-0,000880	-0,001900	-0,000907	-0,000897
1,50	0,045000	-0,002238	-0,003098	-0,002136	-0,002145
2,00	0,048607	0,000132	-0,000559	0,000271	0,000259
3,00	0,054748	0,005377	0,004908	0,005507	0,005499
4,00	0,062932	0,013266	0,012918	0,013372	0,013367
5,00	0,046020	-0,003773	-0,004049	-0,003686	-0,003689
6,00	0,050685	0,000824	0,000597	0,000898	0,000895
8,00	0,053646	0,003721	0,003554	0,003776	0,003774
10,00	0,047370	-0,002584	-0,002716	-0,002540	-0,002541

amostras com tamanhos 10 e 15. Todos os estimadores que têm a proposta de corrigir viés apresentaram desempenho superior ao EMV, além de serem muito competitivos entre si. Devemos observar que os estimadores θ_1^* e $\tilde{\theta}_2$ são muito parecidos, principalmente para valores grande de α .

No modelo de Poisson descrito nas Seções 2.7 e 3.2, sendo que o objetivo é estimar o parâmetro canônico, apesar de termos expressões fechadas para os cinco estimadores, só faz sentido estimar os vieses dos estimadores preventivo. Realizamos diversas simulações variando o parâmetro λ com valores entre um a sete. A Tabela 4.5 apresenta um resumo dessas simulações para o caso de amostras com tamanhos $n = 10$ e $n = 15$. Percebemos uma leve superioridade do estimador θ_1^* para o tamanho da amostra $n = 10$, por exemplo, quando o parâmetro é $\lambda = 1,5$ temos uma casa decimal de diferença entre os vieses.

Também foram realizados estudos no modelo binomial, sendo que $x \sim B(n, p)$. Neste caso, o objetivo é estimar o parâmetro canônico $\theta = \log\left(\frac{p}{1-p}\right)$; e no modelo geométrico, o qual o interesse está sobre o parâmetro $\theta = \log(1-p)$. As funções

Tabela 4.4. Vieses dos estimadores do parâmetro α da densidade gama, $n = 15$.

Parâmetro α	viés de $\hat{\theta}$ $O(n^{-1})$	viés de $\hat{\theta}_1$ $O(n^{-2})$	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$	viés de $\tilde{\theta}_2$ $O(n^{-3})$
0,80	0,029156	0,001302	0,000847	0,001252	0,001260
1,00	0,027680	-0,001815	-0,002266	-0,001826	-0,001823
1,50	0,030852	-0,000683	-0,001061	-0,000637	-0,000640
2,00	0,029288	-0,003050	-0,003354	-0,002989	-0,002992
3,00	0,032857	-0,000063	-0,000270	0,000006	0,000009
4,00	0,034346	0,001233	0,001079	0,001279	0,001278
5,00	0,032410	-0,000788	-0,000910	-0,000750	-0,000751
6,00	0,021970	-0,011271	-0,011372	-0,011239	-0,011240
8,00	0,032478	-0,000806	-0,000880	-0,000782	-0,000782
10,00	0,042934	0,009632	0,009573	0,009651	0,009651

escore dos modelos binomial e geométrico são dadas por

$$\begin{aligned} L'(\theta) &= -\frac{n \exp(\theta)}{1 + \exp(\theta)} + x \quad \text{e} \\ L'(\theta) &= -\frac{n \exp(\theta)}{1 - \exp(\theta)} + \sum_{i=1}^n x_i, \end{aligned}$$

respectivamente.

Nestes dois modelos, os vieses dos estimadores $\hat{\theta}$, $\hat{\theta}_1$ e $\tilde{\theta}_2$ não são definidos, similarmente ao caso do modelo de Poisson. Portanto, realizamos simulações de Monte Carlo para comparar os vieses dos estimadores preventivos θ_0^* e θ_1^* . Resumos dessas simulações estão apresentados nas Tabelas 4.6 e 4.7, sendo que no modelo binomial estão apresentados os resultados para amostras com tamanhos $n = 5$ e $n = 10$; enquanto no modelo geométrico as amostras têm tamanhos $n = 20$ e $n = 25$. No primeiro modelo com $n = 5$, o estimador θ_1^* é superior, porém quando $n = 10$, os dois estimadores ficam competitivos. Já no modelo geométrico, quando o tamanho da amostra é $n = 20$, os estimadores são competitivos e com

Tabela 4.5. Vieses dos estimadores do parâmetro canônico da densidade de Poisson.

Parâmetro λ	$n = 10$		Parâmetro λ	$n = 15$	
	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$		viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$
1	-0,000676	-0,000133	1	0,001889	0,002105
1,2	-0,003252	-0,002894	1,2	0,003675	0,003819
1,5	-0,000288	-0,000072	1,5	-0,002283	-0,002191
1,8	-0,002964	-0,002817	1,8	-0,001355	-0,001293
2	-0,003255	-0,003137	2	0,001144	0,001193
2,5	-0,001801	-0,001728	2,5	0,000290	0,000322
3	0,001131	0,001180	3	-0,001381	-0,001360
4	-0,002534	-0,002506	4	0,001916	0,001928
5	-0,001507	-0,001490	5	-0,000227	-0,000219
6	-0,001250	-0,001238	6	-0,001425	-0,001420
7	-0,000171	-0,000162	7	0,000441	0,000445

uma leve superioridade de θ_0^* . Porém, quando $n = 25$, o estimador θ_1^* tem um desempenho melhor.

4.3 Modelos Biparamétricos

Nesta seção realizamos simulações em modelos biparamétricos, em particular, em modelos com densidades de probabilidades beta. Foram feitas várias simulações de Monte Carlo, todas sempre com dez mil réplicas. Simulamos diversos valores para os parâmetros caracterizadores das distribuições, assim como utilizamos diferentes tamanhos de amostras.

Os estimadores avaliados neste trabalho são obtidos pela maximização da função log-verossimilhança $L(p, q)$, que na prática podem ser obtidos por encontrar

Tabela 4.6. Vieses dos estimadores do parâmetro canônico da densidade binomial.

Parâmetro p	$n = 5$		Parâmetro p	$n = 10$	
	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$		viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$
0,2	0,769614	0,716203	0,2	0,229317	0,210326
0,3	0,320291	0,282122	0,3	0,049678	0,034487
0,4	0,110073	0,105753	0,4	0,014187	0,010256
0,5	-0,031875	0,006015	0,5	0,001470	0,018874
0,6	-0,159366	-0,070086	0,6	-0,022954	0,030449
0,7	-0,386339	-0,253204	0,7	-0,071604	0,035575
0,8	-0,744097	-0,570776	0,8	-0,254515	-0,085245

Tabela 4.7. Vieses dos estimadores do parâmetro canônico da densidade geométrica.

Parâmetro p	$n = 20$		Parâmetro p	$n = 25$	
	viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$		viés de θ_0^* $O(n^{-2})$	viés de θ_1^* $O(n^{-3})$
0,1	0,000017	0,000086	0,1	-0,000453	-0,000410
0,2	0,001152	0,001439	0,2	-0,000136	0,000047
0,3	0,000283	0,000977	0,3	-0,000667	-0,000229
0,4	0,000331	0,001664	0,4	-0,000976	-0,000134
0,5	-0,001370	0,000928	0,5	0,000648	0,002088
0,6	-0,000681	0,003022	0,6	-0,003535	-0,001184
0,7	-0,002112	0,003616	0,7	-0,002363	0,001282

o zero das funções escore, através das equações

$$\frac{\partial L(p, q)}{\partial p} = 0 \quad \text{e} \quad \frac{\partial L(p, q)}{\partial q} = 0,$$

Tabela 4.8. Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 15$.

		Parâmetro	viés de $\hat{\theta}$ $O(n^{-1})$	viés de θ_1^* $O(n^{-3})$	viés de θ_0^* $O(n^{-2})$
p	0,10	0,018315	0,022217	0,034375	
	0,40	-0,254830	0,033870	0,018531	
p	0,15	-0,005010	0,008110	0,022005	
	0,35	-0,108090	-0,010738	0,012589	
p	0,20	-0,022442	0,006269	0,013969	
	0,30	-0,065074	-0,002502	0,008736	
p	0,25	-0,041650	0,001026	0,009295	
	0,25	-0,043501	0,000304	0,008634	
p	0,30	-0,064496	0,004097	0,009433	
	0,20	-0,024942	0,008464	0,013082	
p	0,35	-0,107980	-0,006167	0,009926	
	0,15	-0,009518	0,009282	0,017631	

sendo

$$\begin{aligned}
 \frac{\partial L(p, q)}{\partial p} &= n \frac{\partial}{\partial p} \log \Gamma(p + q) - n \frac{\partial}{\partial p} \log \Gamma(p) + \sum_{i=1}^n \log x_i \\
 &= n\psi(p + q) - n\psi(p) + \sum_{i=1}^n \log x_i,
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial L(p, q)}{\partial q} &= n \frac{\partial}{\partial q} \log \Gamma(p + q) - n \frac{\partial}{\partial q} \log \Gamma(q) + \sum_{i=1}^n \log(1 - x_i) \\
 &= n\psi(p + q) - n\psi(q) + \sum_{i=1}^n \log(1 - x_i),
 \end{aligned}$$

e $\psi(\cdot)$ denota a *função digama*.

Comparamos os desempenhos do EMV $\hat{\theta}$, do proposto por David Firth θ_0^* , visto na Seção 2.6, e o do estimador θ_1^* proposto neste trabalho. Todos estes estimadores são obtidos por encontrar o zero das funções L' , $L' + \mathcal{A}_0$ e $L' + \mathcal{A}_0 + \mathcal{A}_1$,

Tabela 4.9. Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 20$.

		Parâmetro	viés de $\hat{\theta}$ $O(n^{-1})$	viés de θ_1^* $O(n^{-3})$	viés de θ_0^* $O(n^{-2})$
p	0,10	0,023832	0,032750	0,033427	
	0,40	-0,138500	0,035910	0,014255	
p	0,15	-0,001134	0,016074	0,018669	
	0,35	-0,069844	0,009183	0,008610	
p	0,20	-0,017847	0,003447	0,007538	
	0,30	-0,048041	-0,003720	0,002021	
p	0,25	-0,027482	0,001130	0,007586	
	0,25	-0,029162	-0,000684	0,005852	
p	0,30	-0,044950	-0,006858	0,004492	
	0,20	-0,017420	0,001555	0,008690	
p	0,35	-0,074934	-0,012387	0,003440	
	0,15	-0,007660	0,006706	0,013597	

sendo que \mathcal{A}_0 e \mathcal{A}_1 são definidos como no Teorema 3.4.1. Os zeros dessas funções não podem ser calculados algebricamente, assim, utilizamos a função `SolveNLE` disponível na linguagem de programação *Ox*.

Os desempenhos desses estimadores foram avaliados a partir do viés estimado. Dado que sabemos os verdadeiros valores dos parâmetros de cada distribuição, calculamos a diferença entre estes valores verdadeiros e os obtidos pelos estimadores acima. Ao final do experimento de Monte Carlo calculamos a média das estimativas de cada parâmetro e em seguida a subtraímos do verdadeiro valor do parâmetro, obtendo, assim, o viés estimado.

No modelo beta, visto na seção anterior, usamos que o parâmetro q da equação (4.1) era conhecido; agora iremos supor que não temos informação sobre nenhum dos dois parâmetros. Assim, para empregarmos os métodos desenvol-

Tabela 4.10. Vieses dos estimadores dos parâmetros p e q da densidade beta, $n = 25$.

		Parâmetro	viés de $\hat{\theta}$ $O(n^{-1})$	viés de θ_1^* $O(n^{-3})$	viés de θ_0^* $O(n^{-2})$
p	0,10	0,024451	0,031187	0,033786	
	0,40	-0,097027	0,000492	0,009915	
p	0,15	-0,002121	0,009546	0,014289	
	0,35	-0,051727	-0,008071	0,005383	
p	0,20	-0,014100	0,002264	0,005589	
	0,30	-0,034147	-0,002006	0,003625	
p	0,25	-0,021878	-0,000537	0,003733	
	0,25	-0,022049	0,000176	0,004138	
p	0,30	-0,035263	-0,003879	0,001704	
	0,20	-0,014035	0,002237	0,005502	
p	0,35	-0,056865	-0,007910	0,000425	
	0,15	-0,005942	0,006513	0,010100	

vidos nesta tese necessitaremos fazer uso das derivadas parciais da função de log-verossimilhança. No caso da densidade beta, as derivadas de segunda ordem são dadas pelas expressões,

$$\begin{aligned} \frac{\partial^2 L(p, q)}{\partial p^2} &= n \frac{\partial^2}{\partial p^2} \log \Gamma(p+q) - n \frac{\partial^2}{\partial p^2} \log \Gamma(p) \\ &= n\psi'(p+q) - n\psi'(p), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 L(p, q)}{\partial p \partial q} &= n \frac{\partial^2}{\partial p \partial q} \log \Gamma(p+q) \\ &= n\psi'(p+q), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 L(p, q)}{\partial q^2} &= n \frac{\partial^2}{\partial q^2} \log \Gamma(p+q) - n \frac{\partial^2}{\partial q^2} \log \Gamma(q) \\ &= n\psi'(p+q) - n\psi'(q), \end{aligned}$$

senqo que $\psi'(\cdot)$ denota a função trigama. A partir dessas três derivadas fica fácil encontrar as expressões para as demais.

Nas Tabelas 4.8, 4.9 e 4.10 estão os vieses estimados dos três estimadores $\widehat{\theta}$, θ_0^* e θ_1^* . Percebemos que quando o tamanho da amostra é pequeno, $n = 15$, os desempenhos dos estimadores são próximos, porém com uma superioridade do estimador θ_1^* , pois este tem viés menor em muitos casos, por exemplo, quando $p = 0,2$ e $q = 0,3$. Esta superioridade torna-se mais nítida quando $n = 25$, veja como exemplo a Tabela 4.10, quando $p = q = 0,25$, em que o viés de θ_1^* é nulo até a terceira casa decimal, enquanto que o viés do EMV é nulo só na primeira casa decimal e o viés de θ_0^* é nulo até a segunda casa decimal.

4.4 Considerações Finais

Avaliamos o desempenho do EMV $\widehat{\theta}$ e de suas variações. Para isso, fizemos uso da metodologia de Monte Carlo em vários modelos estatísticos e com diferentes tamanhos amostrais. Dividimos o estudo em duas partes, uma dedicada aos modelos uniparamétricos e outra aos modelos biparamétricos.

Na estimação uniparamétrica, em particular nos modelos beta e gama, os quatro estimadores, $\widehat{\theta}_1$, θ_0^* , θ_1^* e $\widetilde{\theta}_2$, que têm como objetivo reduzir o viés dos estimador de máxima verossimilhança, apresentaram desempenhos similares. Já nos modelos de Poisson, binomial e geométrico não têm sentido estimar os vieses do EMV e dos estimadores corretivos $\widehat{\theta}_1$ e $\widetilde{\theta}_2$, pois os vieses ou são menos infinito ou não são definidos. Comparando apenas os estimadores preventivos, θ_0^* e θ_1^* , fica evidente na maioria das situações um desempenho melhor do estimador θ_1^* .

Comparamos os desempenhos apenas dos três estimadores $\widehat{\theta}$, θ_0^* , e θ_1^* , no caso biparamétrico. Quando o tamanho da amostra é pequeno, percebemos claramente que o estimador θ_1^* é superior, porém a medida que n aumenta os desempenhos dos três estimadores tornam-se parecidos, porém, com uma leve superioridade do estimador θ_1^* .

Capítulo 5

CONCLUSÕES

Em muitos modelos estatísticos, a técnica de estimação por máxima verossimilhança conduz a um estimador viesado. Em geral, estes estimadores são assintoticamente não viesados. Assim, um dos objetivos da inferência estatística é desenvolver metodologias de estimação que produzam estimadores com vieses cada vez menores. Ou seja, estimadores com vieses corrigidos.

Há vários anos muitos trabalhos relacionados à correção de vieses foram publicados. Destacamos o artigo apresentado por David Firth, nele é proposta uma técnica de correção de viés em EMVs, que consiste em fazer uma translação na função escore, de tal maneira que o zero desta função transladada é um estimador cujo viés tem ordem de magnitude $O(n^{-2})$. Podemos assim dizer que esta é uma técnica preventiva de correção de viés, já que antes de estimar o parâmetro se faz a correção do viés através de uma translação da função escore, diferentemente das metodologias nas quais estima-se o viés e em seguida subtrai-se do estimador, como é o caso dos estimadores $\hat{\theta}_1$ e $\tilde{\theta}_2$.

Foi apresentada por Bartlett uma expressão para uma aproximação do viés do EMV até ordem $O(n^{-\frac{3}{2}})$. Com base nessa expressão, generalizamos para o caso uniparamétrico a ideia de transladar a função escore, proposta por David Firth. Esta generalização consiste em fazer sucessivas translações na função escore, seno que a primeira translação é exatamente a que David Firth sugeriu. Assim, obtemos uma sequência de estimadores, sendo o primeiro elemento o EMV, cujo viés tem ordem de magnitude $O(n^{-1})$, o segundo é θ_0^* , com viés de ordem $O(n^{-2})$,

o terceiro é θ_1^* , o qual tem viés da ordem de $O(n^{-3})$, e o k -ésimo elemento tem viés com ordem $O(n^{-k})$. Ou seja, iniciamos a sequência com o EMV e a cada translação da função escore obtemos um novo estimador, cujo viés tem a ordem de magnitude diminuída de $O(n^{-k})$ para $O(n^{-(k+1)})$, com $k = 1, 2, \dots$.

Por meio de simulações de Monte Carlo, avaliamos no espaço uniparamétrico, os desempenhos de cinco estimadores. São eles: o EMV $\hat{\theta}$, o estimador de segunda ordem $\hat{\theta}_1$, o proposto por Firth θ_0^* , o estimador apresentado nesta tese θ_1^* e o elaborado por Ferrari et al. $\tilde{\theta}_2$. Desta forma, chegamos às seguintes conclusões:

1. Para amostras de tamanho pequeno, aproximadamente $n = 15$, exceto o EMV, estes estimadores muitas vezes são competitivos entre si, pois apresentam vieses muito próximos.
2. Os estimadores θ_1^* e $\tilde{\theta}_2$ são os que apresentam mais similaridade.
3. Nos modelos beta e gama os estimadores θ_1^* e $\tilde{\theta}_2$ apresentaram uma leve superioridade no seu desempenho em relação aos demais estimadores.
4. Nos modelos de Poisson, geométrico e binomial não faz sentido calcular o viés médio para os estimadores $\hat{\theta}$, $\hat{\theta}_1$ e $\tilde{\theta}_2$, pois o viés ou é menos infinito, como no caso do modelo de Poisson, ou não é definido. Assim, os estimadores θ_0^* e θ_1^* são boas alternativas para se produzir um estimador cujo viés tem ordem de magnitude $O(n^{-3})$.
5. Em geral, desde que estejam satisfeitas as suposições **S1** a **S9**, o estimador θ_1^* é uma boa metodologia para se obter estimadores com vieses com ordem de magnitude $O(n^{-3})$.

Expandimos esta metodologia para os casos biparamétrico e multiparamétrico. Para o caso em que o parâmetro da população pertence ao espaço biparamétrico, fizemos simulações de Monte Carlo com o intuito de avaliar apenas os estimadores $\hat{\theta}$, θ_0^* e θ_1^* . Concluímos que:

1. Os estimadores com vieses corrigidos, em geral, apresentam desempenho melhor que o EMV.

2. Os estimadores θ_0^* e θ_1^* são competitivos entre si, porém, há uma evidente superioridade do estimador θ_1^* , proposto neste trabalho, para tamanhos pequenos da amostra.
3. A metodologia empregada na construção do estimador θ_1^* é uma boa alternativa para produzir estimadores cujo viés seja da ordem $O(n^{-3})$.

Nesta tese realizamos correções de vieses de alta ordem em EMVs. A ideia principal deste trabalho é que em vez de maximizar a função verossimilhança, encontramos o zero da função escore transladada. Assim, deixamos como sugestão para um trabalho futuro tentar empregar esta metodologia em outras técnicas de estimação que se baseiam em encontrar zero de funções, como por exemplo, estimação por mínimos quadrados.

Referências Bibliográficas

- [1] Barreto-Souza, W. & Vasconcellos, K.L.P. (2011). “Bias and skewness in a general extreme-value regression model”. Computational Statistics and Data Analysis, 55, 1379-1393.
- [2] Bartlett, M.S. (1953). “Aproximate confidence intervals”, Biometrika, 40, 12-19.
- [3] Bolfarine, H. & Sandoval, M.C. (2001). “Introdução à Inferência Estatística”, SBM, São Paulo.
- [4] Box, M.J. (1971). “Bias in nonlinear estimation (with discussion)”. Journal of the Royal Statistical Society B, 33, 171-201.
- [5] Cordeiro, M.G. (1999). “Introdução à Teoria Assintótica”, 22º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro.
- [6] Cordeiro, G.M. & Botter, D.A., 2001. “Second-order biases of maximum likelihood estimates in overdispersed generalized linear models”. Statistics and Probability Letters, 55, 269-280.
- [7] Cordeiro, G.M., Vasconcellos, K.L.P. & Santos, M.L., 1998. “On the second-order bias of parameter estimates in nonlinear regression models with Student t errors”. Journal of Statistical Computation and Simulation, 60, 363-378.
- [8] Cordeiro, G.M., Ferrari, S.L.P., Uribe-Opazo, M.A. & Vasconcellos, K.L.P. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. Statistics and Probability Letters, 46, 317-328.

- [9] Cordeiro, G.M. & McCullagh, P. (1991). "Bias correction in generalized linear models". *Journal of the Royal Statistical Society B*, 53, 629-643.
- [10] Cordeiro, G.M. & Vasconcellos, K.L.P. (1997). "Bias correction for a class of multivariate nonlinear regression models". *Statistics and Probability Letters*, 35, 155-164.
- [11] Cordeiro, G.M. & Vasconcellos, K.L.P. (1999). Second-order biases of the maximum likelihood estimates in von Mises regression models. *Australian and New Zealand Journal of Statistics*, 41, 901-910.
- [12] Cox, D.R. & Hinkley, D.V. (1974). "Theoretical Statistics". London: Chapman and Hall.
- [13] Cox, D. R. & Snell, E. J. (1968). "A general definition of residuals (with discussion)", *Journal of the Royal Statistical Society B*, 30, 248-275.
- [14] Cribari-Neto, F. & Vasconcellos, K.L.P. (2002). "Nearly unbiased maximum likelihood estimation for the beta distribution". *Journal of Statistical Computation and Simulation*, 72, 107-118.
- [15] Cramér, H. (1946). "Mathematical Methods of Statistics". Princeton University Press.
- [16] Ferrari, S.L.P. & Cribari-Neto, F. (1998). "On bootstrap and analytical bias corrections". *Economics Letters*, 58, 7-15.
- [17] Ferrari, S.L.P., Botter, D.A., Cordeiro, G.M. & Cribari-Neto, F. (1996). "Second and third order bias reduction for one-parameter family models". *Statistics and Probability Letters*, 30, 339-345.
- [18] Firth, D. (1993). "Bias reduction of maximum likelihood estimates". *Biometrika*, 80, 27-38.
- [19] Fisher, R.A. (1921). "On Mathematical Foundations of Theoretical Statistics", *Philosophical Transactions of the Royal Society of London A*, 222, 309-368.

- [20] Greene, W. H. (2000), “Econometric Analysis”, 4^a ed., Prentice-Hall, Upper Saddle River, New Jersey.
- [21] Gourieroux, C. & Monfort, A. (1995), “Statistics and Econometric Models”, Cambridge University Press, Cambridge.
- [22] Haldane, J. B. S. (1953). “The estimation of two parameters from a sample”, *Sankhya*, 12, 313-320.
- [23] Haldane, J. B. S. & Smith, S. M. (1953). “The sampling distribution of a maximum likelihood estimate”, *Biometrika*, 43, 96-103.
- [24] Hogg, R. V. & Craig, A. T. (1995) “Introduction to Mathematical Statistics”, Prentice Hall, Uper Sadlle River, New Jersey.
- [25] Horn, R. A. & Johnson, C. R. (1991). “Topics in Matrix Analysis”, Cambridge University Press.
- [26] James, B.R. (1996). “Probabilidade: Um Curso em Nível Intermediário”, IMPA, Rio de Janeiro.
- [27] Jeffreys, H. (1939/1961). “Theory of Probability”, Oxford University Press, Oxford.
- [28] Jeffreys, H. (1946). “An invariant form for the prior probability in estimation problems”, *Proc. R. Soc. A*, 186, 453-461.
- [29] Lamport, L. (1994). LATEX, A Document Preparation System. Addison-Wesley, Reading, 2nd. edition.
- [30] Lawley, D. (1956). “A general method for approximating to the distribution of likelihood ratio criteria”, *Biometrika*, 43, 295-303.
- [31] Lehmann, E. L. & Casella, G. (1998). “Theory of Point Estimation”, 2nd Ed., Springer, New York.

- [32] Lemonte, A.J., Cribari-Neto, F. & Vasconcellos, K.L.P. (2007). “Improved statistical inference for the two-parameter Birnbaum-Saunders distribution”. Computational Statistics and Data Analysis, 51, 4656-4681.
- [33] McCullagh, P. (1987). “Tensor Methods in Statistics”, Chapman and Hall, London.
- [34] Magalhães, M.N. (2004). “Probabilidade e Variáveis Aleatórias”, USP, São Paulo.
- [35] Magnus, J. R. & Neudecker, H. (1988). “Matrix Differential Calculus with Applications in Statistics and Econometrics”, John Wiley & Sons, New York.
- [36] Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). “Introduction to the Theory of Statistics”, McGraw-Hill.
- [37] Ospina, R., Cribari-Neto, F. & Vasconcellos, K.L.P. (2006). “Improved point and interval estimation for a beta regression model”. Computational Statistics and Data Analysis, 51, 960-981.
- [38] Paulino, C. D., Turkman, M. A. A. & Murteira, B. (2003). “Estatística Bayesiana”, Fundação Calouste Gulbenkian, Lisboa.
- [39] Pike, M., Hill, A. & Smith, P. (1980). “Bias and efficiency in logistic analysis of stratified case-control studies”. International Journal of Epidemiology, 9, 89-95.
- [40] Quenouille, M. H. (1949). “Approximate tests of correlation in time-series”. J. R. Statist. Soc. B 11, 68-84.
- [41] Quenouille, M. H. (1956). “Notes on bias in estimation”. Biometrika 43, 353-60.
- [42] Rudin, W. (1989). “Principles of Mathematical Analysis”. McGraw-Hill, New York.
- [43] Shenton, L. R. & Bowman, K. (1977). “Maximum Likelihood Estimation in Small Samples”, MacMillan, New York.

- [44] Stuart, A. & Ord, J. K. (1987). "Kendall's Advanced Theory of Statistics", Vol. I, 1nd Ed., Oxford University Press, New York.
- [45] Vasconcellos, K.L.P. (2010). "Correção de Viés de Estimadores". Artigo não publicado.
- [46] Vasconcellos, K.L.P. & Cordeiro, G.M. (1997). "Approximate bias for multivariate nonlinear heteroscedastic regressions". *Brazilian Journal of Probability and Statistics*, 11, 141-159.
- [47] Vasconcellos, K.L.P. & Cordeiro, G.M. (1997). "Estimadores corrigidos para modelos SUR não-lineares". *Brazilian Review of Econometrics*, 17, 45-65.
- [48] Vasconcellos, K.L.P. & Cordeiro, G.M. (2000). "Bias corrected estimates in multivariate Student t regression models". *Communications in Statistics - Theory and Methods*, 29, 797-822.
- [49] Vasconcellos, K.L.P. & Cribari-Neto, F. (2005). "Improved maximum likelihood estimation in a new class of beta regression models". *Brazilian Journal of Probability and Statistics*, 19, 13-31.
- [50] Vasconcellos, K.L.P., Cordeiro, G.M. & Barroso, L.P. (2000). "Improved estimation for robust econometric regression models". *Brazilian Journal of Probability and Statistics*, 14, 141-157.
- [51] Vasconcellos, K.L.P., Frery, A.C. & Silva, L.B. (2005). "Improving estimation in speckled imagery". *Computational Statistics*, 20, 503-519.
- [52] Vasconcellos, K.L.P. & Dourado, G.B. (2008). "Nearly unbiased estimation in a biparametric exponential family". *Journal of Statistical Computation and Simulation*, 78, 387-404.
- [53] Vasconcellos, K.L.P. & Silva, S.G. (2005). "Corrected estimates for Student t regression models with unknown degrees of freedom". *Journal of Statistical Computation and Simulation*, 75, 409-423.

- [54] Young, D.H. & Bakir, S.T. (1987). "Bias correction for a generalized log-gamma regression model. *Technometrics*", 29, 183-191.

Apêndice A

Expansão da Equação (3.3)

Com o auxílio da Tabela 3.1 podemos expandir a equação (3.3) até $k = 5$, assim,

$b_1^h = O_p(n^{-\frac{1}{2}})$:

$$\frac{\mathbf{L}'_{\mathbf{h}}}{\mathbf{I}}.$$

$b_2^h = O_p(n^{-\frac{2}{2}})$:

$$\begin{aligned} &= \frac{11L'_h(L''_h + I)}{I^2} + \frac{-10L'_h(L''_h + I)}{I^2} + \frac{-(L'_h)^2(-L'''_h)}{2I^3} \\ &= \frac{\mathbf{L}'_{\mathbf{h}}(\mathbf{L}''_{\mathbf{h}} + \mathbf{I})}{\mathbf{I}^2} + \frac{(\mathbf{L}'_{\mathbf{h}})^2\mathbf{L}'''_{\mathbf{h}}}{2\mathbf{I}^3}. \end{aligned}$$

$b_3^h = O_p(n^{-\frac{3}{2}})$:

$$\begin{aligned} &= \frac{55L'_h(L''_h + I)^2}{I^3} + \frac{-99L'_h(L''_h + I)^2}{I^3} + \frac{45L'_h(L''_h + I)^2}{I^3} \\ &+ \frac{-11(L'_h)^2(-L'''_h)(L''_h + I)}{2I^4} + \frac{8(L'_h)^2(-L'''_h)(L''_h + I)}{2I^4} + \frac{(L'_h)^3\mathbf{H}_3}{6I^5} \\ &= \frac{\mathbf{L}'_{\mathbf{h}}(\mathbf{L}''_{\mathbf{h}} + \mathbf{I})^2}{\mathbf{I}^3} + \frac{3(\mathbf{L}'_{\mathbf{h}})^2\mathbf{L}'''_{\mathbf{h}}(\mathbf{L}''_{\mathbf{h}} + \mathbf{I})}{2\mathbf{I}^4} + \frac{(\mathbf{L}'_{\mathbf{h}})^3\mathbf{H}_3}{6\mathbf{I}^5}. \end{aligned}$$

Por exemplo, para o caso em que $h = 0$, temos $L'_0 = L' + A_0$.

$$\begin{aligned} b_3^0 &= \frac{(L' + A_0)(L'' + A'_0 + I)^2}{I^3} + \frac{3(L' + A_0)^2(L''' + A''_0)(L'' + A'_0 + I)}{2I^4} \\ &+ \frac{(L' + A_0)^3[3(L''' + A''_0)^2 - (L'' + A'_0)(L^{iv} + A'''_0)]}{6I^5}. \end{aligned}$$

Como $b_3^0 = b_3 + B_3^0$, retirando-se da equação anterior os termos pertencentes a b_3 , temos que

$$\begin{aligned}
 B_3^0 = & \frac{L' [2A'_0(L'' + I) + (A'_0)^2]}{I^3} + \frac{A_0(L'' + A'_0 + I)^2}{I^3} \\
 + & \frac{3(L')^2 L''' A'_0}{2I^4} + \frac{3(L')^2 A''_0(L'' + A'_0 + I)}{2I^4} \\
 + & \frac{3(2L'A_0 + A_0^2)(L''' + A''_0)(L'' + A'_0 + I)}{2I^4} \\
 + & \frac{(L')^3 [3(2L''' A''_0 + (A''_0)^2) - L'' A'''_0 - A'_0(L^{iv} + A'''_0)]}{6I^5} \\
 + & \frac{[3(L')^2 A_0 + 3L'A_0^2 + A_0^3] [3(L''' + A''_0)^2 - (L'' + A'_0)(L^{iv} + A'''_0)]}{6I^5}.
 \end{aligned}$$

$b_4^h = O_p(n^{-\frac{4}{2}})$:

$$\begin{aligned}
 & = \frac{165L'_h(L''_h + I)^3}{I^4} + \frac{-440L'_h(L''_h + I)^3}{I^4} \frac{396L'_h(L''_h + I)^3}{I^4} \\
 + & \frac{-120L'_h(L''_h + I)^3}{I^4} + \frac{-55(L'_h)^2(-L'''_h)(L''_h + I)^2}{2I^5} \\
 + & \frac{77(L'_h)^2(-L'''_h)(L''_h + I)^2}{2I^5} + \frac{-28(L'_h)^2(-L'''_h)(L''_h + I)^2}{2I^5} \\
 + & \frac{11(L'_h)^3 \mathbf{H}_3(L''_h + I)}{6I^6} + \frac{-6(L'_h)^3 \mathbf{H}_3(L''_h + I)}{6I^6} + \frac{-(L'_h)^4 \mathbf{H}_4}{24I^7} \\
 = & \frac{\mathbf{L}'_h(\mathbf{L}''_h + \mathbf{I})^3}{\mathbf{I}^4} + \frac{6(\mathbf{L}'_h)^2 \mathbf{L}'''_h(\mathbf{L}''_h + \mathbf{I})^2}{2\mathbf{I}^5} + \frac{5(\mathbf{L}'_h)^3 \mathbf{H}_3(\mathbf{L}''_h + \mathbf{I})}{6\mathbf{I}^6} \\
 - & \frac{(\mathbf{L}'_h)^4 \mathbf{H}_4}{24\mathbf{I}^7}.
 \end{aligned}$$

$b_5^h = O_p(n^{-\frac{5}{2}})$:

$$\begin{aligned}
 & = \frac{330L'_h(L''_h + I)^4}{I^5} + \frac{-1155L'_h(L''_h + I)^4}{I^5} + \frac{1540L'_h(L''_h + I)^4}{I^5} \\
 + & \frac{-924L'_h(L''_h + I)^4}{I^5} + \frac{210L'_h(L''_h + I)^4}{I^5} + \frac{-165(L'_h)^2(-L'''_h)(L''_h + I)^3}{2I^6} \\
 + & \frac{330(L'_h)^2(-L'''_h)(L''_h + I)^3}{2I^6} + \frac{-231(L'_h)^2(-L'''_h)(L''_h + I)^3}{2I^6} \\
 + & \frac{56(L'_h)^2(-L'''_h)(L''_h + I)^3}{2I^6} + \frac{55(L'_h)^3 \mathbf{H}_3(L''_h + I)^2}{6I^7} \\
 + & \frac{-55(L'_h)^3 \mathbf{H}_3(L''_h + I)^2}{6I^7} + \frac{15(L'_h)^3 \mathbf{H}_3(L''_h + I)^2}{6I^7} \\
 + & \frac{-11(L'_h)^4 \mathbf{H}_4(L''_h + I)}{24I^8} + \frac{4(L'_h)^4 \mathbf{H}_4(L''_h + I)}{24I^8} + \frac{(L'_h)^5 \mathbf{H}_5}{120I^9} \\
 = & \frac{\mathbf{L}'_h(\mathbf{L}''_h + \mathbf{I})^4}{\mathbf{I}^5} + \frac{10(\mathbf{L}'_h)^2(-\mathbf{L}'''_h)(\mathbf{L}''_h + \mathbf{I})^3}{2\mathbf{I}^6} + \frac{15(\mathbf{L}'_h)^3 \mathbf{H}_3(\mathbf{L}''_h + \mathbf{I})^2}{6\mathbf{I}^7}
 \end{aligned}$$

$$+ \frac{7(\mathbf{L}'_h)^4 \mathbf{H}_4 (\mathbf{L}''_h + \mathbf{I})}{24\mathbf{I}^8} + \frac{(\mathbf{L}'_h)^5 \mathbf{H}_5}{120\mathbf{I}^9}.$$

Para completar essa expansão precisamos encontrar os \mathbf{H}_i 's, sabemos que $\mathbf{H}_1 = 1$ e que $\mathbf{H}_i = \mathbf{H}'_{i-1} L''_h - (2i-3)L'''_h \mathbf{H}_{i-1}$, assim teremos,

- $\mathbf{H}_1 = 1.$
- $\mathbf{H}_2 = -\mathbf{L}'''_h.$
- $\mathbf{H}_3 = 3(\mathbf{L}'''_h)^2 - \mathbf{L}''\mathbf{L}^{iv}_h.$ Logo temos que $H'_3 = 5L'''_h L^{iv}_h - L''_h L^v_h.$
- $\mathbf{H}_4 = 10\mathbf{L}''_h \mathbf{L}'''_h \mathbf{L}^{iv}_h - (\mathbf{L}''_h)^2 \mathbf{L}^v - 15(\mathbf{L}'''_h)^3.$ Pois,

$$\begin{aligned} \mathbf{H}_4 &= (5L'''_h L^{iv}_h - L''_h L^v_h) L''_h - (8-3)L'''_h (3(L'''_h)^2 - L''_h L^{iv}_h) \\ &= 5L''_h L'''_h L^{iv}_h - (L''_h)^2 L^v - 15(L'''_h)^3 + 5L''_h L'''_h L^{iv}_h \\ &= 10L''_h L'''_h L^{iv}_h - (L''_h)^2 L^v_h - 15(L'''_h)^3. \end{aligned}$$

Da mesma forma, sua derivada é dada por

$$\begin{aligned} \mathbf{H}'_4 &= 10 \left\{ (L'''_h)^2 L^{iv}_h + L''_h [(L^{iv}_h)^2 + L'''_h L^v] \right\} - 2L''_h L'''_h L^v_h \\ &\quad - (L''_h)^2 L^{vi}_h - 45(L'''_h)^2 L^{iv}_h \\ &= 10(L'''_h)^2 L^{iv}_h + 10L''_h (L^{iv}_h)^2 + 10L''_h L'''_h L^v_h - 2L''_h L'''_h L^v_h \\ &\quad - (L''_h)^2 L^{vi}_h - 45(L'''_h)^2 L^{iv}_h \\ &= -35(L'''_h)^2 L^{iv}_h + 10L''_h (L^{iv}_h)^2 + 8L''_h L'''_h L^v_h - (L''_h)^2 L^{vi}_h. \end{aligned}$$

- $\mathbf{H}_5 = -105\mathbf{L}''_h (\mathbf{L}'''_h)^2 \mathbf{L}^{iv}_h + 10(\mathbf{L}''_h)^2 (\mathbf{L}^{iv}_h)^2 + 78(\mathbf{L}''_h)^2 \mathbf{L}'''_h \mathbf{L}^v_h - (\mathbf{L}''_h)^3 \mathbf{L}^{vi}_h$
 $- 105(\mathbf{L}'''_h)^4.$

Seguindo o algoritmo, temos

$$\begin{aligned} \mathbf{H}_5 &= \mathbf{H}'_4 L''_h - (10-3)L'''_h \mathbf{H}_4 \\ &= -35L''_h (L'''_h)^2 L^{iv}_h + 10(L''_h)^2 (L^{iv}_h)^2 + 8(L''_h)^2 L'''_h L^v_h - (L''_h)^3 L^{vi}_h \\ &\quad - 70L''_h (L'''_h)^2 L^{iv}_h + 70(L''_h)^2 L'''_h L^v_h + 105(L'''_h)^4 \\ &= -105L''_h (L'''_h)^2 L^{iv}_h + 10(L''_h)^2 (L^{iv}_h)^2 + 78(L''_h)^2 L'''_h L^v_h - (L''_h)^3 L^{vi}_h \\ &\quad - 105(L'''_h)^4. \end{aligned}$$

Apêndice B

Valores Esperados de Potências da Função Escore

Observe que, de forma geral, temos que

$$\begin{aligned} E\left[\frac{f_{\alpha\beta}}{f}\right] &= \int \frac{f_{\alpha\beta}}{f} f dx = \int \frac{\partial^2}{\partial\alpha\partial\beta} f dx \\ &= \frac{\partial^2}{\partial\alpha\partial\beta} \left[\int f dx \right] = \frac{\partial^2}{\partial\alpha\partial\beta} [1] = 0. \end{aligned}$$

Assim, temos que

$$\begin{aligned} E[L_{\alpha\beta}] &= E\left[\frac{\partial}{\partial\beta} [L_\alpha]\right] = E\left[\frac{\partial}{\partial\beta} \left[\frac{f_\alpha}{f}\right]\right] \\ &= E\left[\frac{f_{\alpha\beta}}{f} - \frac{f_\alpha f_\beta}{f^2}\right] = -E[L_\alpha L_\beta]. \end{aligned}$$

Da mesma forma temos que

$$\begin{aligned} E[L_{\alpha\alpha}] &= E\left[\frac{\partial}{\partial\alpha} [L_\alpha]\right] = E\left[\frac{\partial}{\partial\alpha} \left[\frac{f_\alpha}{f}\right]\right] \\ &= E\left[\frac{f_{\alpha\alpha}}{f} - \frac{f_\alpha f_\alpha}{f^2}\right] = -E[L_\alpha L_\alpha]. \end{aligned}$$

$$\begin{aligned} E[L_{\alpha\alpha\beta}] &= E\left[\frac{\partial}{\partial\alpha} [L_{\alpha\beta}]\right] = E\left[\frac{\partial}{\partial\alpha} \left[\frac{f_{\alpha\beta}}{f} - \frac{f_\alpha f_\beta}{f^2}\right]\right] \\ &= E\left[\frac{f_{\alpha\beta\alpha}}{f} - \frac{f_{\alpha\beta} f_\alpha}{f^2} - \left[\frac{f_{\alpha\alpha} f_\beta + f_\alpha f_{\beta\alpha}}{f^2} - 2\frac{f_\alpha^2 f_\beta}{f^3}\right]\right] \\ &= E\left[-2\frac{f_{\alpha\beta} f_\alpha}{f^2} - \frac{f_{\alpha\alpha} f_\beta}{f^2} + 2\frac{f_\alpha^2 f_\beta}{f^3}\right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[-2L_\alpha \frac{f_{\alpha\beta}}{f} - L_\beta \frac{f_{\alpha\alpha}}{f} + 2L_\alpha^2 L_\beta \right] \\
&= \mathbb{E} \left[-2L_\alpha [L_{\alpha\beta} + L_\alpha L_\beta] - L_\beta [L_{\alpha\alpha} + L_\alpha^2] + 2L_\alpha^2 L_\beta \right] \\
&= \mathbb{E} \left[-2L_\alpha L_{\alpha\beta} - L_\beta L_{\alpha\alpha} - L_\alpha^2 L_\beta \right].
\end{aligned}$$

Ou seja,

$$\mathbb{E} [L_\alpha^2 L_\beta] = -2\mathbb{E} [L_\alpha L_{\alpha\beta}] - \mathbb{E} [L_\beta L_{\alpha\alpha}] - \mathbb{E} [L_{\alpha\alpha\beta}].$$

Com o intuito de simplificar os cálculos, considere o caso onde a matriz L'' não é aleatória. Observe que

$$\frac{f_{\alpha\beta}}{f} = L_{\alpha\beta} + L_\alpha L_\beta.$$

Se derivarmos $\frac{f_{\alpha\beta}}{f}$ em relação a α temos

$$\begin{aligned}
\frac{f_{\alpha\alpha\alpha}}{f} &= \frac{f_{\alpha\alpha} f_\alpha}{f^2} + L_{\alpha\alpha\alpha} + 2L_\alpha L_{\alpha\alpha} \\
&= (L_{\alpha\alpha} + L_\alpha^2) L_\alpha + L_{\alpha\alpha\alpha} + 2L_\alpha L_{\alpha\alpha} \\
&= L_\alpha^3 + L_{\alpha\alpha\alpha} + 3L_\alpha L_{\alpha\alpha}.
\end{aligned}$$

Derivando em relação a β a expressão $\frac{f_{\alpha\alpha\alpha}}{f}$, temos que

$$\begin{aligned}
\frac{f_{\alpha\alpha\alpha\beta}}{f} &= \frac{f_{\alpha\alpha\alpha} f_\beta}{f^2} + 3L_\alpha^2 L_{\alpha\beta} + L_{\alpha\alpha\alpha\beta} + 3L_{\alpha\beta} L_{\alpha\alpha} + 3L_\alpha L_{\alpha\alpha\beta} \\
&= (L_\alpha^3 + L_{\alpha\alpha\alpha} + 3L_\alpha L_{\alpha\alpha}) L_\beta + 3L_\alpha^2 L_{\alpha\beta} + L_{\alpha\alpha\alpha\beta} + 3L_{\alpha\beta} L_{\alpha\alpha} + 3L_\alpha L_{\alpha\alpha\beta}.
\end{aligned}$$

Calculando o valor esperado dessa expressão obtemos

$$\begin{aligned}
\mathbb{E} [L_\alpha^3 L_\beta] &= -3\mathbb{E} [L_\alpha L_\beta L_{\alpha\alpha}] - 3\mathbb{E} [L_\alpha^2 L_{\alpha\beta}] - L_{\alpha\alpha\alpha\beta} - 3L_{\alpha\beta} L_{\alpha\alpha} \\
&= 3L_{\alpha\beta} L_{\alpha\alpha} - L_{\alpha\alpha\alpha\beta}.
\end{aligned}$$

Derivando em relação a β a expressão $\frac{f_{\alpha\alpha\alpha}}{f}$ temos que

$$\frac{f_{\alpha\alpha\beta}}{f} = L_\beta L_{\alpha\alpha} + L_\alpha^2 L_\beta + L_{\alpha\alpha\beta} + 2L_\alpha L_{\alpha\beta}.$$

Assim,

$$\begin{aligned}\frac{f_{\alpha\alpha\beta\beta}}{f} &= (L_\beta L_{\alpha\alpha} + L_\alpha^2 L_\beta + L_{\alpha\alpha\beta} + 2L_\alpha L_{\alpha\beta}) L_\beta + L_{\beta\beta} L_{\alpha\alpha} + L_\beta L_{\alpha\alpha\beta} \\ &+ 2L_\alpha L_\beta L_{\alpha\beta} + L_\alpha^2 L_{\beta\beta} + L_{\alpha\alpha\beta\beta} + 2L_{\alpha\beta} L_{\alpha\beta} + 2L_\alpha L_{\alpha\beta\beta}.\end{aligned}$$

Logo,

$$\begin{aligned}\mathbb{E}[L_\alpha^2 L_\beta^2] &= L_{\alpha\alpha} L_{\beta\beta} + 2L_{\alpha\beta}^2 - L_{\alpha\alpha} L_{\beta\beta} + 2L_{\alpha\beta}^2 + L_{\alpha\alpha} L_{\beta\beta} - L_{\alpha\alpha\beta\beta} - 2L_{\alpha\beta}^2 \\ &= L_{\alpha\alpha} L_{\beta\beta} + 2L_{\alpha\beta}^2 - L_{\alpha\alpha\beta\beta}.\end{aligned}$$

Apêndice C

Derivadas da Função Inversa da Função Escore

Seja a função $H(H_1, H_2)$ a inversa da função escore $L'(\alpha, \beta) = (L_\alpha, L_\beta)$, então,

$$\begin{aligned} H' &= \text{inv}(L'')^{-1} \\ &= \begin{bmatrix} L_{\beta\beta} & -L_{\alpha\beta} \\ -L_{\beta\alpha} & L_{\alpha\alpha} \end{bmatrix} \frac{1}{|L''|}. \end{aligned}$$

A derivada da matriz H' em relação a L' é dada por

$$H'(L(\theta)) = \begin{bmatrix} \frac{\partial H_1}{\partial L_\alpha} & \frac{\partial H_1}{\partial L_\beta} \\ \frac{\partial H_2}{\partial L_\alpha} & \frac{\partial H_2}{\partial L_\beta} \end{bmatrix}.$$

Adicionalmente,

$$\begin{aligned} H''(L(\theta)) &= \begin{bmatrix} \frac{\partial H'}{\partial L_\alpha} & \frac{\partial H'}{\partial L_\beta} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial^2 H_1}{\partial L_\alpha^2} & \frac{\partial^2 H_1}{\partial L_\alpha \partial L_\beta} & \frac{\partial^2 H_1}{\partial L_\beta \partial L_\alpha} & \frac{\partial^2 H_1}{\partial L_\beta^2} \\ \frac{\partial^2 H_2}{\partial L_\alpha^2} & \frac{\partial^2 H_2}{\partial L_\alpha \partial L_\beta} & \frac{\partial^2 H_2}{\partial L_\beta \partial L_\alpha} & \frac{\partial^2 H_2}{\partial L_\beta^2} \end{bmatrix}. \end{aligned}$$

Podemos simplificar essa derivada da seguinte forma:

$$H' = \begin{bmatrix} x & y \\ z & w \end{bmatrix},$$

$$H'' = \begin{bmatrix} H'_\alpha x + H'_\beta z & \sim & H'_\alpha y + H'_\beta w \end{bmatrix},$$

$$H''' = \begin{bmatrix} H''_\alpha x + H''_\beta z & \sim & H''_\alpha y + H''_\beta w \end{bmatrix},$$

$$H^{(iv)} = \begin{bmatrix} H'''_\alpha x + H'''_\beta z & \sim & H'''_\alpha y + H'''_\beta w \end{bmatrix},$$

$$H''_\alpha = \begin{bmatrix} H'_{\alpha\alpha} x + H'_{\alpha} x_\alpha + H'_{\beta\alpha} z + H'_{\beta} z_\alpha & \sim & H'_{\alpha\alpha} y + H'_{\alpha} y_\alpha + H'_{\beta\alpha} w + H'_{\beta} w_\alpha \end{bmatrix},$$

$$H''_{\alpha\alpha} = \begin{bmatrix} H'_{\alpha\alpha\alpha} x + H'_{\alpha\alpha} x_\alpha + H'_{\alpha\alpha} x_\alpha + H'_{\alpha} x_{\alpha\alpha} + H'_{\beta\alpha\alpha} z + H'_{\beta\alpha} z_\alpha + H'_{\beta\alpha} z_\alpha + H'_{\beta} z_{\alpha\alpha} & \sim \\ H'_{\alpha\alpha\alpha} y + H'_{\alpha\alpha} y_\alpha + H'_{\alpha\alpha} y_\alpha + H'_{\alpha} y_{\alpha\alpha} + H'_{\beta\alpha\alpha} w + H'_{\beta\alpha} w_\alpha + H'_{\beta\alpha} w_\alpha + H'_{\beta} w_{\alpha\alpha} & \end{bmatrix},$$

$$H''_{\alpha\beta} = \begin{bmatrix} H'_{\alpha\alpha\beta} x + H'_{\alpha\alpha} x_\beta + H'_{\alpha\beta} x_\alpha + H'_{\alpha} x_{\alpha\beta} + H'_{\beta\alpha\beta} z + H'_{\beta\alpha} z_\beta + H'_{\beta\alpha} z_\beta + H'_{\beta} z_{\alpha\beta} & \sim \\ H'_{\alpha\alpha\beta} y + H'_{\alpha\alpha} y_\beta + H'_{\alpha\beta} y_\alpha + H'_{\alpha} y_{\alpha\beta} + H'_{\beta\alpha\beta} w + H'_{\beta\alpha} w_\beta + H'_{\beta\alpha} w_\beta + H'_{\beta} w_{\alpha\beta} & \end{bmatrix},$$

$$H''_\beta = \begin{bmatrix} H'_{\alpha\beta} x + H'_{\alpha} x_\beta + H'_{\beta\beta} z + H'_{\beta} z_\beta & \sim & H'_{\alpha\beta} y + H'_{\alpha} y_\beta + H'_{\beta\beta} w + H'_{\beta} w_\beta \end{bmatrix},$$

$$H''_{\beta\beta} = \begin{bmatrix} H'_{\alpha\beta\beta} x + H'_{\alpha\beta} x_\beta + H'_{\alpha\beta} x_\beta + H'_{\alpha} x_{\beta\beta} + H'_{\beta\beta\beta} z + H'_{\beta\beta} z_\beta + H'_{\beta\beta} z_\beta + H'_{\beta} z_{\beta\beta} & \sim \\ H'_{\alpha\beta\beta} y + H'_{\alpha\beta} y_\beta + H'_{\alpha\beta} y_\beta + H'_{\alpha} y_{\beta\beta} + H'_{\beta\beta\beta} w + H'_{\beta\beta} w_\beta + H'_{\beta\beta} w_\beta + H'_{\beta} w_{\beta\beta} & \sim \end{bmatrix},$$

$$H'''_\alpha = \begin{bmatrix} H''_{\alpha\alpha} x + H''_{\alpha} x_\alpha + H''_{\beta\alpha} z + H''_{\beta} z_\alpha & \sim & H''_{\alpha\alpha} y + H''_{\alpha} y_\alpha + H''_{\beta\alpha} w + H''_{\beta} w_\alpha \end{bmatrix},$$

$$H_{\beta}''' = \begin{bmatrix} H_{\alpha\beta}''x + H_{\alpha}''x_{\beta} + H_{\beta\beta}''z + H_{\beta}''z_{\beta} & \sim & H_{\alpha\beta}''y + H_{\alpha}''y_{\beta} + H_{\beta\beta}''w + H_{\beta}''w_{\beta} \end{bmatrix}.$$

Outras derivadas utilizadas ao longo do texto são:

$$\begin{aligned} |L''| &= L_{\alpha\alpha}L_{\beta\beta} - 2L_{\alpha\beta}^2, \\ |L''|_{\alpha} &= L_{\alpha\alpha\alpha}L_{\beta\beta} + L_{\alpha\alpha}L_{\beta\beta\alpha} - 2L_{\alpha\beta\alpha}L_{\alpha\beta}, \\ |L''|_{\beta} &= L_{\alpha\alpha\beta}L_{\beta\beta} + L_{\alpha\alpha}L_{\beta\beta\beta} - 2L_{\alpha\beta\beta}L_{\alpha\beta}, \end{aligned}$$

$$|L''|_{\alpha\beta} = L_{\alpha\alpha\alpha\beta}L_{\beta\beta} + L_{\alpha\alpha\alpha}L_{\beta\beta\beta} + L_{\alpha\alpha\beta}L_{\beta\beta\alpha} + L_{\alpha\alpha}L_{\beta\beta\alpha\beta} - 2L_{\alpha\beta\alpha\beta}L_{\alpha\beta} - 2L_{\alpha\beta\alpha}L_{\alpha\beta\beta},$$

$$\begin{aligned} |L''|_{\alpha\alpha} &= L_{\alpha\alpha\alpha\alpha}L_{\beta\beta} + L_{\alpha\alpha\alpha}L_{\beta\beta\alpha} + L_{\alpha\alpha\alpha}L_{\beta\beta\alpha} + L_{\alpha\alpha}L_{\beta\beta\alpha\alpha} - 2L_{\alpha\beta\alpha\alpha}L_{\alpha\beta} - 2L_{\alpha\beta\alpha}^2 \\ &= L_{\alpha\alpha\alpha\alpha}L_{\beta\beta} + 2L_{\alpha\alpha\alpha}L_{\beta\beta\alpha} + L_{\alpha\alpha}L_{\beta\beta\alpha\alpha} - 2L_{\alpha\beta\alpha\alpha}L_{\alpha\beta} - 2L_{\alpha\beta\alpha}^2, \end{aligned}$$

$$\begin{aligned} |L''|_{\alpha\alpha\beta} &= L_{\alpha\alpha\alpha\alpha\beta}L_{\beta\beta} + L_{\alpha\alpha\alpha\alpha}L_{\beta\beta\beta} + 2L_{\alpha\alpha\alpha\beta}L_{\beta\beta\alpha} + 2L_{\alpha\alpha\alpha}L_{\beta\beta\alpha\beta} + L_{\alpha\alpha\beta}L_{\beta\beta\alpha\alpha} \\ &+ L_{\alpha\alpha}L_{\beta\beta\alpha\alpha\beta} - 2L_{\alpha\beta\alpha\alpha\beta}L_{\alpha\beta} - 2L_{\alpha\beta\alpha\alpha}L_{\alpha\beta\beta} - 4L_{\alpha\beta\alpha}L_{\alpha\beta\alpha\beta}, \end{aligned}$$

$$\begin{aligned} |L''|_{\alpha\alpha\alpha} &= L_{\alpha\alpha\alpha\alpha\alpha}L_{\beta\beta} + L_{\alpha\alpha\alpha\alpha}L_{\beta\beta\alpha} + 2L_{\alpha\alpha\alpha\alpha}L_{\beta\beta\alpha} + 2L_{\alpha\alpha\alpha}L_{\beta\beta\alpha\alpha} \\ &+ L_{\alpha\alpha\alpha}L_{\beta\beta\alpha\alpha} + L_{\alpha\alpha}L_{\beta\beta\alpha\alpha\alpha} - 2L_{\alpha\beta\alpha\alpha\alpha}L_{\alpha\beta} \\ &+ -2L_{\alpha\beta\alpha\alpha}L_{\alpha\beta\alpha} - 4L_{\alpha\beta\alpha}L_{\alpha\beta\alpha\alpha}, \end{aligned}$$

$$\begin{aligned} |L''|_{\alpha\alpha\alpha} &= L_{\alpha\alpha\alpha\alpha\alpha}L_{\beta\beta} + 3L_{\alpha\alpha\alpha\alpha}L_{\beta\beta\alpha} + 3L_{\alpha\alpha\alpha}L_{\beta\beta\alpha\alpha} \\ &+ L_{\alpha\alpha}L_{\beta\beta\alpha\alpha\alpha} - 2L_{\alpha\beta\alpha\alpha\alpha}L_{\alpha\beta} - 2L_{\alpha\beta\alpha\alpha}L_{\alpha\beta\alpha} - 4L_{\alpha\beta\alpha}L_{\alpha\beta\alpha\alpha}, \end{aligned}$$

$$\begin{aligned}
|L''|_{\beta\beta} &= L_{\alpha\alpha\beta\beta}L_{\beta\beta} + L_{\alpha\alpha\beta}L_{\beta\beta\beta} + L_{\alpha\alpha\beta}L_{\beta\beta\beta} + L_{\alpha\alpha}L_{\beta\beta\beta\beta} - 2L_{\alpha\beta\beta\beta}L_{\alpha\beta} - 2L_{\alpha\beta\beta}L_{\alpha\beta\beta} \\
&= L_{\alpha\alpha\beta\beta}L_{\beta\beta} + 2L_{\alpha\alpha\beta}L_{\beta\beta\beta} + L_{\alpha\alpha}L_{\beta\beta\beta\beta} - 2L_{\alpha\beta\beta\beta}L_{\alpha\beta} - 2L_{\alpha\beta\beta}^2,
\end{aligned}$$

$$\begin{aligned}
|L''|_{\beta\beta\alpha} &= L_{\alpha\alpha\beta\beta\alpha}L_{\beta\beta} + L_{\alpha\alpha\beta\beta}L_{\beta\beta\alpha} + 2L_{\alpha\alpha\beta\alpha}L_{\beta\beta\beta} + 2L_{\alpha\alpha\beta}L_{\beta\beta\beta\alpha} + L_{\alpha\alpha\alpha}L_{\beta\beta\beta\beta} \\
&+ L_{\alpha\alpha}L_{\beta\beta\beta\beta\alpha} - 2L_{\alpha\beta\beta\beta\alpha}L_{\alpha\beta} - 2L_{\alpha\beta\beta\beta}L_{\alpha\beta\alpha} - 4L_{\alpha\beta\beta}L_{\alpha\beta\beta\alpha},
\end{aligned}$$

$$\begin{aligned}
|L''|_{\beta\beta\beta} &= L_{\alpha\alpha\beta\beta\beta}L_{\beta\beta} + L_{\alpha\alpha\beta\beta}L_{\beta\beta\beta} + 2L_{\alpha\alpha\beta\beta}L_{\beta\beta\beta} + 2L_{\alpha\alpha\beta}L_{\beta\beta\beta\beta} \\
&+ L_{\alpha\alpha\beta}L_{\beta\beta\beta\beta} + L_{\alpha\alpha}L_{\beta\beta\beta\beta\beta} - 2L_{\alpha\beta\beta\beta\beta}L_{\alpha\beta} \\
&+ -2L_{\alpha\beta\beta\beta}L_{\alpha\beta\beta} - 4L_{\alpha\beta\beta}L_{\alpha\beta\beta\beta},
\end{aligned}$$

$$\begin{aligned}
|L''|_{\beta\beta\beta} &= L_{\alpha\alpha\beta\beta\beta}L_{\beta\beta} + 3L_{\alpha\alpha\beta\beta}L_{\beta\beta\beta} + 3L_{\alpha\alpha\beta}L_{\beta\beta\beta\beta} \\
&+ L_{\alpha\alpha}L_{\beta\beta\beta\beta\beta} - 2L_{\alpha\beta\beta\beta\beta}L_{\alpha\beta} - 2L_{\alpha\beta\beta\beta}L_{\alpha\beta\beta} - 4L_{\alpha\beta\beta}L_{\alpha\beta\beta\beta}.
\end{aligned}$$

As derivadas parciais de H' são

$$H'_\alpha = \begin{bmatrix} L_{\beta\beta\alpha} & -L_{\alpha\beta\alpha} \\ -L_{\alpha\beta\alpha} & L_{\alpha\alpha\alpha} \end{bmatrix} \frac{1}{|L''|} - H' \frac{|L''|_\alpha}{|L''|}.$$

Uma notação alternativa é definir a matriz

$$M = \begin{bmatrix} L_{\beta\beta} & -L_{\alpha\beta} \\ -L_{\alpha\beta} & L_{\alpha\alpha} \end{bmatrix}.$$

Assim,

$$H' = \frac{M}{|L''|},$$

$$H'_\alpha = \frac{M_\alpha - H' |L''|_\alpha}{|L''|},$$

$$\begin{aligned} H'_{\alpha\alpha} &= \frac{M_{\alpha\alpha} - H'_\alpha |L''|_\alpha - H' |L''|_{\alpha\alpha} - H'_\alpha |L''|_\alpha}{|L''|} \\ &= \frac{M_{\alpha\alpha} - 2H'_\alpha |L''|_\alpha - H' |L''|_{\alpha\alpha}}{|L''|}, \end{aligned}$$

$$H'_{\alpha\beta} = \frac{M_{\alpha\beta} - H'_\beta |L''|_\alpha - H' |L''|_{\alpha\beta} - H'_\alpha |L''|_\beta}{|L''|},$$

$$\begin{aligned} H'_{\alpha\beta\alpha} &= \frac{M_{\alpha\beta\alpha} - H'_{\beta\alpha} |L''|_\alpha - H'_\beta |L''|_{\alpha\alpha} - H'_\alpha |L''|_{\alpha\beta} - H' |L''|_{\alpha\beta\alpha}}{|L''|} \\ &+ \frac{-H'_{\alpha\alpha} |L''|_\beta - H'_\alpha |L''|_{\beta\alpha} - H'_{\alpha\beta} |L''|_\alpha}{|L''|}, \end{aligned}$$

$$\begin{aligned} H'_{\alpha\beta\beta} &= \frac{M_{\alpha\beta\beta} - H'_{\beta\beta} |L''|_\alpha - H'_\beta |L''|_{\alpha\beta} - H'_\beta |L''|_{\alpha\beta} - H' |L''|_{\alpha\beta\beta}}{|L''|} \\ &+ \frac{-H'_{\alpha\beta} |L''|_\beta - H'_\alpha |L''|_{\beta\beta} - H'_{\alpha\beta} |L''|_\beta}{|L''|}, \end{aligned}$$

$$H'_{\alpha\beta\beta} = \frac{M_{\alpha\beta\beta} - H'_{\beta\beta} |L''|_\alpha - 2H'_\beta |L''|_{\alpha\beta} - H' |L''|_{\alpha\beta\beta} - 2H'_{\alpha\beta} |L''|_\beta - H'_\alpha |L''|_{\beta\beta}}{|L''|},$$

$$\begin{aligned} H'_{\alpha\alpha\alpha} &= \frac{M_{\alpha\alpha\alpha} - 2H'_{\alpha\alpha} |L''|_\alpha - 2H'_\alpha |L''|_{\alpha\alpha} - H'_\alpha |L''|_{\alpha\alpha} - H' |L''|_{\alpha\alpha\alpha} - H'_{\alpha\alpha} |L''|_\alpha}{|L''|} \\ &= \frac{M''_{\alpha\alpha\alpha} - 3H'_{\alpha\alpha} |L''|_\alpha - 3H'_\alpha |L''|_{\alpha\alpha} - H' |L''|_{\alpha\alpha\alpha}}{|L''|}, \end{aligned}$$

$$H'_{\alpha\alpha\beta} = \frac{M_{\alpha\alpha\beta} - 2H'_{\alpha\beta} |L''|_\alpha - 2H'_\alpha |L''|_{\alpha\beta} - H'_\beta |L''|_{\alpha\alpha} - H' |L''|_{\alpha\alpha\beta} - H'_{\alpha\alpha} |L''|_\beta}{|L''|},$$

$$H'_\beta = \frac{M_\beta - H' |L''|_\beta}{|L''|},$$

$$H'_{\beta\beta} = \frac{M_{\beta\beta} - 2H'_\beta |L''|_\beta - H' |L''|_{\beta\beta}}{|L''|},$$

$$H'_{\beta\beta\beta} = \frac{M_{\beta\beta\beta} - 3H'_{\beta\beta}|L''|_{\beta} - 3H'_{\beta}|L''|_{\beta\beta} - H'|L''|_{\beta\beta\beta}}{|L''|}.$$

Para calcular R_0 , observe que

$$\mathcal{A}_0 = \frac{1}{2|L''|} \begin{bmatrix} |L''|_{\alpha} \\ |L''|_{\beta} \end{bmatrix}$$

Portanto,

$$\begin{aligned} A_{0\alpha} &= \frac{|L''|_{\alpha\alpha}}{2|L''|} - 2A_0^2, \\ B_{0\beta} &= \frac{|L''|_{\beta\beta}}{2|L''|} - 2B_0^2, \\ A_{0\beta} &= \frac{|L''|_{\alpha\beta}}{2|L''|} - 2A_0B_0, \\ B_{0\alpha} &= A_{0\beta}, \end{aligned}$$

$$\mathcal{R}_e(0) = \begin{bmatrix} B_{0\beta} & -A_{0\beta} \\ -B_{0\alpha} & A_{0\alpha} \end{bmatrix} \frac{1}{|L''|} - \frac{R_0}{|L''| + R_0} H'.$$

Para calcularmos $\mathcal{R}'_e(0)$ precisamos encontrar sua derivada em relação a α e a β . Assim, temos que

$$\begin{aligned} A_{0\alpha\alpha} &= \frac{|L''|_{\alpha\alpha\alpha}}{2|L''|} - \frac{|L''|_{\alpha\alpha}|L|_{\alpha}}{2|L''|^2} - 4A_0A_{0\alpha} \\ &= \frac{|L''|_{\alpha\alpha\alpha}}{2|L''|} - (A_{0\alpha} + 2A_0^2)2A_0 - 4A_0A_{0\alpha}, \end{aligned}$$

$$\begin{aligned} A_{0\alpha\alpha} &= \frac{|L''|_{\alpha\alpha\alpha}}{2|L''|} - 4A_0^3 - 6A_0A_{0\alpha}, \\ B_{0\beta\beta} &= \frac{|L''|_{\beta\beta\beta}}{2|L''|} - 4B_0^3 - 6B_0B_{0\beta}, \end{aligned}$$

$$\begin{aligned} A_{0\alpha\beta} &= \frac{|L''|_{\alpha\alpha\beta}}{2|L''|} - 2A_{0\alpha}B_0 - 4A_0^2B_0 - 4A_0A_{0\beta}, \\ B_{0\beta\alpha} &= \frac{|L''|_{\beta\beta\alpha}}{2|L''|} - 2A_0B_{0\beta} - 4A_0B_0^2 - 4B_0B_{0\alpha}, \end{aligned}$$

$$\begin{aligned} A_{0\beta\beta} &= \frac{|L''|_{\alpha\beta\beta}}{2L''} - 4A_{0\beta}B_0 - 4A_0B_0^2 - 2A_0B_{0\beta}, \\ A_{0\beta\alpha} &= \frac{|L''|_{\alpha\beta\alpha}}{2L''} - 4A_0A_{0\beta} - 4A_0^2B_0 - 2A_{0\alpha}B_0, \end{aligned}$$

$$B_{0\alpha\alpha} = A_{0\beta\alpha},$$

$$B_{0\alpha\beta} = A_{0\beta\beta}.$$

Como $A_{0\beta} = B_{0\alpha}$ e $L_{\alpha\beta} = L_{\beta\alpha}$, temos que R_0 é dado por

$$\begin{aligned} R_0 &= A_{0\alpha}L_{\beta\beta} + B_{0\beta}L_{\alpha\alpha} + A_{0\alpha}B_{0\beta} - A_{0\beta}L_{\beta\alpha} - B_{0\alpha}L_{\alpha\beta} - A_{0\beta}B_{0\alpha} \\ &= A_{0\alpha}L_{\beta\beta} + B_{0\beta}L_{\alpha\alpha} + A_{0\alpha}B_{0\beta} - 2A_{0\beta}L_{\beta\alpha} - A_{0\beta}B_{0\alpha}. \end{aligned}$$

Logo sua derivada em relação a α é

$$\begin{aligned} R_{0\alpha} &= A_{0\alpha\alpha}L_{\beta\beta} + A_{0\alpha}L_{\beta\beta\alpha} + B_{0\beta\alpha}L_{\alpha\alpha} + B_{0\beta}L_{\alpha\alpha\alpha} + A_{0\alpha\alpha}B_{0\beta} \\ &\quad + A_{0\alpha}B_{0\beta\alpha} - 2A_{0\beta\alpha}L_{\beta\alpha} - 2A_{0\beta}L_{\beta\alpha\alpha} - A_{0\beta\alpha}B_{0\alpha} - A_{0\beta}B_{0\alpha\alpha}. \end{aligned}$$

A derivada em relação a β é

$$\begin{aligned} R_{0\beta} &= A_{0\alpha\beta}L_{\beta\beta} + A_{0\alpha}L_{\beta\beta\beta} + B_{0\beta\beta}L_{\alpha\alpha} + B_{0\beta}L_{\alpha\alpha\beta} + A_{0\alpha\beta}B_{0\beta} \\ &\quad + A_{0\alpha}B_{0\beta\beta} - 2A_{0\beta\beta}L_{\beta\alpha} - 2A_{0\beta}L_{\beta\alpha\beta} - A_{0\beta\beta}B_{0\alpha} - A_{0\beta}B_{0\alpha\beta}. \end{aligned}$$

Para calcular a derivada de $\mathcal{R}_e(0)$ defina a seguinte matriz:

$$\mathcal{MA} = \begin{bmatrix} B_{0\beta} & -A_{0\beta} \\ -B_{0\alpha} & A_{0\alpha} \end{bmatrix}.$$

Logo,

$$\mathcal{R}_e(0) = \frac{\mathcal{MA}}{|L''|} - \frac{R_0}{|L''| + R_0} H'.$$

Portanto

$$\begin{aligned}\mathcal{R}_e(0)_\alpha &= \frac{\mathcal{MA}_\alpha}{|L''|} - \frac{\mathcal{MA}|L''|_\alpha}{|L''|^2} - \left[\frac{R_{0\alpha}}{|L''| + R_0} - \frac{|L''|_\alpha + R_{0\alpha}}{(|L''| + R_0)^2} \right] H' - \frac{R_0}{|L''| + R_0} H'_\alpha, \\ \mathcal{R}_e(0)_\beta &= \frac{\mathcal{MA}_\beta}{|L''|} - \frac{\mathcal{MA}|L''|_\beta}{|L''|^2} - \left[\frac{R_{0\beta}}{|L''| + R_0} - \frac{|L''|_\beta + R_{0\beta}}{(|L''| + R_0)^2} \right] H' - \frac{R_0}{|L''| + R_0} H'_\beta.\end{aligned}$$

Logo a derivada de $\mathcal{R}_e(0)$ em relação a L' é dada por,

$$\mathcal{R}'_e(0) = \frac{1}{|L''|} [\mathcal{R}_e(0)_\alpha L_{\beta\beta} - \mathcal{R}_e(0)_\beta L_{\alpha\beta}, -\mathcal{R}_e(0)_\alpha L_{\beta\alpha} + \mathcal{R}_e(0)_\beta L_{\alpha\alpha}].$$