
ESTIMAÇÃO PONTUAL E INTERVALAR EM UM MODELO DE REGRESSÃO BETA

RAYDONAL OSPINA MARTÍNEZ

Orientador: Prof. Dr. Klaus Leite Pinto Vasconcellos

Co-orientador: Prof. Dr. Francisco Cribari-Neto

Área de concentração: Estatística Matemática

Dissertação submetida como requerimento parcial para obtenção do
grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, 16 de fevereiro de 2004

*Dedico este trabalho a minha mãe, Martha
Martínez, e a meu pai, Guillermo Ospina.*

A Yuri.

Agradecimentos

A Deus, que sempre me acompanha, me guia e me consola em todos os momentos de minha vida. Ainda, por suas provas que me ensinam e me tornam mais forte.

Ao professor Klaus Leite Pinto Vasconcellos, pela confiança, paciência e orientação segura; assim como as nossas longas discussões e horas de leitura, que tornaram possível este trabalho.

Ao professor Francisco Cribari-Neto, pelo estímulo, dedicação, orientação incondicional e leitura cuidadosa deste trabalho. Mais ainda, pelo apoio em todos os momentos decisivos de minha estadia no Brasil.

A minha família, pelo amor incondicional e por mesmo estando tão longe de mim sempre ter-me apoiado durante toda a duração do curso.

A Patrícia, por todos os momentos de alegria e amor que me tem dedicado e por me ensinar que com fé tudo é possível.

Às professoras Viviana Giampaoli e Silvia Ferrari, pelas pertinentes sugestões e recomendações na elaboração deste trabalho

Pelas dicas e inspiração, que chamo de ajuda divina, pois sem elas não conseguiria entender e enxergar além de meus olhos as soluções de alguns problemas.

Ao professor Alejandro César Frery, pela confiança e incentivo para estudar na UFPE.

Aos professores do Programa de Mestrado em Estatística da UFPE.

A todos meus professores da Universidad Nacional de Colombia, em particular a Liliana Blanco, Jimmy Corzo, Myriam Muñoz e Alberto Lopez pelo apoio e confiança.

A meus grandes amigos na Colômbia, Hanner, Yofre, Emilse, Edilneyi, Adriana, Karen, Angela, Rosa, Claudia, Alexis, Rainer, Arjuna, Diego e Oscar, pelos momentos de amizade e carinho.

A Felipe, Batolomeu, Ricardo e Gilson, pela convivência.

A meus irmãos no Brasil, Felipe, Bartolomeu e Silvia, pela amizade incondicional e pelas alegrias compartilhadas durante estes dois anos de luta.

A Karl-Heinrich e Mozart, pelo trabalho intenso e amizade sem fronteiras.

A todos os colegas e grandes amigos de minha turma, Felipe, Bartolomeu, Gilson, Silvia, Keila, Tatiene, Tarciana, Moises, João Marcelo, Patrícia e Rubens, pelo companheirismo, pelas horas de diversão e pelos intermináveis momentos de trabalho acadêmico compartilhado.

A Valéria, pela presteza, competência e carinho com os quais sempre resolveu os problemas da pós-graduação.

A Michelli, Diana, Carla e Pedro, pela ajuda dada nos primeiros meses de minha estadia no Recife e pela amizade.

Ao Brasil que me possibilitou a continuidade de meus estudos de pós-graduação e a todas as pessoas que me acolheram com tanto carinho.

A todos que participaram direta ou indiretamente da concretização deste ideal: Minha família, todos os meus amigos na Colômbia, meus ex-professores da Universidad Nacional de Colombia, meus companheiros de graduação, meus colegas de trabalho no DANE, a família Leone Espinheira, meus amigos da colônia Colombiana no Brasil. Agradeço também às cidades de Recife, Salvador e São Paulo.

À trindade $\text{\TeX}$, $\text{\Ox}$ e $\text{\R}$, por serem as melhores ferramentas computacionais com que já trabalhei.

Aos participantes da banca examinadora, pelas sugestões.

Ao CNPq, pelo apoio financeiro.

À vida e à esperança de um melhor amanhã.

An appropriate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. —John Tukey

Resumo

O modelo de regressão beta possui potencialmente aplicabilidade prática, em particular, na modelagem de taxas e proporções. Assim, o cálculo dos vieses dos estimadores dos parâmetros deste modelo torna-se importante, visto que, em geral, para modelos regulares, quanto menores são os tamanhos de amostra, mais viesados são os estimadores de máxima verossimilhança. A obtenção de expressões que permitam calcular os vieses desses estimadores possibilita a obtenção de estimadores corrigidos, que em princípio são mais precisos que os não corrigidos.

O objetivo deste trabalho é fornecer expressões para os vieses de segunda ordem dos estimadores de máxima verossimilhança no modelo de regressão beta proposto por Ferrari & Cribari-Neto (2003). Com a finalidade de reduzir os vieses destes estimadores em amostras finitas, utilizam-se correções de viés obtidas a partir de esquemas analíticos (Cox & Snell, 1968; Firth, 1993) e de bootstrap. Deduzimos uma fórmula para o cálculo dos vieses de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta. Em seguida, fornecemos estimativas corrigidas do tipo corretivo, preventivo e de bootstrap, mostrando numericamente que as estimativas corrigidas de tipo corretivo e de bootstrap apresentam desempenhos superiores em termos de viés e erro médio quadrático às suas respectivas estimativas de máxima verossimilhança.

Apresentamos intervalos de confiança do tipo assintótico, bootstrap percentil e bootstrap BCa para os parâmetros do modelo de regressão beta. A avaliação numérica revelou que os intervalos de tipo percentil para os parâmetros baseados nas estimativas corrigidas apresentam os melhores desempenhos em termos de cobertura, balanceamento e comprimento. Adicionalmente, mostramos que os intervalos de confiança para o parâmetro de precisão são bastante assimétricos, sendo que os intervalos do tipo assintótico baseados nas estimativas de máxima verossimilhança e corrigida corretivamente possuem as melhores coberturas e menores comprimentos.

Abstract

In this thesis we consider the beta regression model recently proposed by Ferrari & Cribari-Neto (2003), which is tailored to situations where the response is restricted to the standard unit interval and has a regression structure involving regressors and unknown parameters. We derive the second order biases of the maximum likelihood estimators, and use them to define bias-adjusted estimators. As an alternative to the two analytically corrected estimators, we consider a bias correction mechanism via parametric bootstrap. The numerical evidence favors the bootstrap-biased estimator and the one obtained from the analytical corrective adjustment. Several different strategies for interval estimation are also considered. The thesis closes with an empirical application.

Índice

1. Introdução	1
1.1. Introdução	1
1.1.1. Organização da dissertação	2
1.1.2. Suporte computacional	2
1.2. Preliminares	3
1.3. Estimadores de máxima verossimilhança	5
1.4. Condições de regularidade	6
1.5. Identidades de Bartlett	8
1.6. Intervalos de confiança assintóticos	9
1.7. Otimização não-linear em inferência estatística	10
1.7.1. Algoritmo BFGS	11
1.7.2. Algoritmo de Newton-Raphson	11
2. O modelo de regressão beta	13
2.1. Modelo beta	13
2.2. Definição e estimação do modelo beta	13
2.3. O modelo de regressão beta	21
3. Correção de viés	31
3.1. Revisão sobre correção de viés	31
3.2. Correção de Cox & Snell	34
3.2.1. Correção de viés dos estimadores de máxima verossimilhança do modelo de regressão beta	35
3.2.2. O viés do vetor de médias	40
3.3. Correção preventiva	42
3.4. Correção por bootstrap	44
3.5. Intervalos de confiança bootstrap	47
4. Avaliação numérica	51
4.1. Introdução	51
4.2. Detalhes metodológicos	52
4.3. Resultados e discussão	54
4.3.1. Resultados da correção de viés	54
4.3.2. Resultados dos intervalos de confiança	57
5. Aplicação e conclusões	77

5.1. Aplicação do modelo de regressão beta	77
5.2. Conclusões	78
★ Apêndice	80
A. Cálculo dos momentos	80
B. Programa de simulação	93
B.1. Programa principal	93
B.2. Biblioteca de funções	103
B.3. Programa de ligação	111
B.4. Programa de geração do histograma das 5000 estimativas de máxima verossimilhança, e as diferentes versões de intervalos de confiança, para níveis de confiança de 90%, 95% e 99%	112
★ Referências	114

Capítulo 1

Introdução

1.1. Introdução

A distribuição beta fornece uma ferramenta útil para modelagem de dados estatísticos que assumem valores no intervalo $(0, 1)$. Uma estratégia analítica baseada em máxima verossimilhança para modelagem de dados usando a distribuição beta é aqui estudada no contexto de um novo modelo de regressão, o qual sob uma adequada parametrização pode ser utilizado para modelar, por exemplo, taxas e proporções (Ferrari & Cribari-Neto, 2003). Nesta linha, uma importante área de pesquisa é o estudo do comportamento de estimadores de máxima verossimilhança (MV) em pequenas amostras, em particular, análise de viés; este em geral, mede quão distante em média está um estimador do verdadeiro parâmetro estudado. Muitas vezes, esta medida é utilizada como um critério para avaliar a “qualidade” de um estimador. O viés não é um problema muito sério quando os tamanhos das amostras são grandes, pois tipicamente é de ordem $O(n^{-1})$ ao passo que o erro padrão assintótico é de ordem $O(n^{-1/2})$, onde n é o tamanho de amostra. No entanto, para valores moderados de tamanho de amostra n o viés pode ser problemático. Assim, a correção de viés torna-se importante quando a informação disponível é pequena. Desta forma, a formulação de expressões que permitam corrigir o viés possibilita a obtenção de estimadores mais precisos que os não-corrigidos.

O mecanismo de ajuste do viés foi amplamente pesquisado por Box (1971), o qual forneceu uma expressão geral para o viés de ordem $O(n^{-1})$ em modelos não-lineares multivariados com matriz de covariâncias conhecida. Cook, Tsai & Wei (1986) relacionam o viés com a posição das variáveis explicativas no espaço amostral. Young & Bakir (1987) mostram que a correção de viés pode melhorar a estimação em modelos de regressão generalizados log-gama. Cordeiro & McCullagh (1991) fornecem fórmulas matriciais para a correção de viés em modelos lineares generalizados. Trabalhos mais recentes nesta área incluem Cordeiro & Vasconcellos (1997), que obtiveram fórmulas para a correção de viés em modelos de regressão não-lineares com erros normais, e Cordeiro & Vasconcellos (1999), que calculam o viés de segunda ordem de estimadores de máxima verossimilhança em modelos de regressão Von Mises, entre alguns outros.

Trabalhos de correção de viés de estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição beta foram desenvolvidos por Cordeiro, Rocha, Rocha & Cribari-Neto (1997) e Cribari-Neto & Vasconcellos (2002). Entretanto, tais resultados não envolvem modelos com estrutura de regressão. Vasconcellos & Cribari-Neto (2003) estudam o comportamento dos estimadores de máxima verossimilhança em uma nova classe de modelos de regressão com distribuição beta, i.e., eles consideram um modelo de regressão que relaciona o comportamento de variáveis aleatórias que seguem uma distribuição beta a um conjunto de variáveis explicativas. O modelo estabelece uma relação entre os parâmetros que indexam a distribuição $\text{Beta}(p, q)$ e um preditor linear. Nesta mesma linha, o objetivo principal deste trabalho é fornecer expressões que permitam corrigir os vieses até ordem $O(n^{-1})$ dos esti-

madores de máxima verossimilhança dos parâmetros no modelo de regressão beta proposto por Ferrari & Cribari–Neto (2003), visto que para pequenas amostras os estimadores de máxima verossimilhança podem ser bastante viesados. No modelo de Ferrari & Cribari–Neto (2003), ao invés do modelo de Vasconcellos & Cribari–Neto (2003), a variável resposta segue um comportamento beta sob uma nova parametrização, i.e., a resposta se distribui como $\text{Beta}(\mu, \phi)$, onde a resposta média μ é modelada através de uma adequada estrutura de regressão sujeita a um parâmetro de precisão ϕ . Assim, o parâmetro μ da distribuição beta pode ser interpretado como uma taxa ou proporção, e o parâmetro ϕ como uma medida de sua precisão. Este trabalho apresenta expressões gerais para o cálculo dos vieses no modelo de regressão beta de Ferrari & Cribari–Neto (2003), para o caso de diferentes funções de ligação, que permitem modelar a média de uma quantidade estudada através de uma estrutura de regressão. Na obtenção da fórmula geral para os vieses, até segunda ordem, dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta foi utilizada a expressão dada por Cox & Snell (1968).

1.1.1. Organização da dissertação

A presente dissertação de mestrado está dividida em cinco capítulos. Neste primeiro capítulo abordamos em forma geral aspectos de estimação e inferência, concentrando-nos principalmente nas características e propriedades do estimador de máxima verossimilhança sob condições de regularidade. No segundo capítulo estudamos a distribuição beta e algumas de suas propriedades amostrais. Além disso, apresentamos o modelo de regressão beta proposto por Ferrari & Cribari–Neto (2003) com a estimação pontual de seus parâmetros por máxima verossimilhança e algumas de suas propriedades assintóticas, entre as quais, a normalidade assintótica das estimativas de máxima verossimilhança, que nos permite fornecer intervalos de confiança de tipo assintótico. No terceiro capítulo mostramos uma revisão geral sobre algumas técnicas de correção de viés, em particular o método de Cox & Snell (1968). Naquele capítulo fornecemos a fórmula geral para os vieses de segunda ordem dos estimadores de máxima verossimilhança do modelo de regressão beta. Através da análise do viés obtemos estimadores corrigidos por viés utilizando as metodologias correctiva, preventiva e de bootstrap para os estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta. Adicionalmente, apresentamos os respectivos intervalos de confiança assintóticos e de bootstrap para ditas estimativas corrigidas. No quarto capítulo, com base em simulações de Monte Carlo para o modelo de regressão beta investigamos o comportamento dos estimadores de máxima verossimilhança, e suas versões corrigidas corretivamente, preventivamente e de bootstrap. Também são analisadas as diferentes estimativas intervalares dos parâmetros do modelo de regressão beta em suas diferentes versões. No quinto capítulo aplicamos o modelo de regressão beta a um conjunto de dados reais e comparamos as estimativas dos parâmetros do modelo por máxima verossimilhança com suas respectivas estimativas corrigidas. Além disso, apresentamos as conclusões deste trabalho.

1.1.2. Suporte computacional

As avaliações numéricas realizadas ao longo desta dissertação foram feitas através da linguagem matricial de programação `0x` em sua versão 3.30 para sistemas operacionais

Linux. `0x` está distribuído gratuitamente para uso acadêmico e se encontra disponível no site <http://www.nuff.ox.ac.uk/Users/Doornik>. Maiores detalhes sobre esta linguagem de programação podem ser encontrados em Cribari-Neto & Zarkos (2003) e em Doornik (2001). O apêndice B desta dissertação contém os programas de simulação que permitem avaliar os estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta e suas versões corrigidas. Os gráficos apresentados no trabalho foram produzidos utilizando o ambiente de programação, análise de dados e gráficos `R` em sua versão 1.8.0 para o sistema operacional Linux e que se encontra disponível gratuitamente no endereço <http://www.r-project.org>. Por fim, a dissertação foi digitada usando o sistema de tipografia (Plain) `TEX` desenvolvido por Donald Knuth em 1986. `TEX` também é distribuído de forma gratuita.

1.2. Preliminares

O objetivo desta seção é fornecer alguns conceitos básicos para o desenvolvimento de nosso trabalho, assim como introduzir a notação e terminologia pertinentes. Consideramos um experimento aleatório, o qual é representado por um espaço de probabilidade $(\Omega, \mathfrak{S}, P)$, sendo Ω o conjunto de todos os possíveis resultados do experimento, $\mathfrak{S}$ uma σ -álgebra de subconjuntos de Ω e P uma medida de probabilidade definida nos elementos de $\mathfrak{S}$. Associada ao experimento está uma variável aleatória (v.a.), a qual é uma transformação $Y : \Omega \rightarrow R$ tal que $Y^{-1}(B) \in \mathfrak{S}$, para todo B , conjunto de Borel, e definimos

$$\begin{aligned} P_Y(B) &= P(Y^{-1}(B)) \\ &= P(Y \in B), \quad \forall B \text{ boreliano.} \end{aligned}$$

Se $B = (-\infty, y]$, então a função $F_Y(y) = P(Y \leq y) = P_Y((-\infty, y]) = P(\{\omega : Y(\omega) \leq y\})$ é chamada função de distribuição de Y . Do ponto de vista estatístico, uma sequência de experimentos pode ser simbolizada por um vetor de observações $\underline{Y} = (Y_1, \dots, Y_n)$, denominado vetor aleatório, que assume valores em $\mathbb{R}^n$. A distribuição de $\underline{Y}$ é determinada por uma função de distribuição conjunta definida em R^n , dada por

$$\begin{aligned} F_{\underline{Y}}(\underline{y}) &= F_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = P(\{\omega \in \Omega : Y_1(\omega) \leq y_1, \dots, Y_n(\omega) \leq y_n\}), \\ &\quad -\infty < y_1, \dots, y_n < \infty. \end{aligned}$$

Se assumimos que os Y_i 's são independentes, então

$$F_{\underline{Y}}(\underline{y}) = F_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \prod_{i=1}^n P(\{\omega \in \Omega : Y_i(\omega) \leq y_i\}) = \prod_{i=1}^n F_{Y_i}(y_i).$$

O valor esperado de uma variável aleatória Y com relação à medida de probabilidade P_Y é definido como a integral de Lebesgue-Stieltjes:

$$EY = \int y dF_Y(y).$$

Supondo que Y é absolutamente contínua com densidade $f(y)$, a distribuição de probabilidade é tal que $P_Y(B) = \int_B f(y) dy$, para todo B boreliano e o valor esperado pode ser

expresso como $EY = \int yf(y)dy$. Se $E|Y| < \infty$, então dizemos que Y é integrável. Algumas características de v.a.'s podem ser representadas convenientemente em termos de valores esperados. Definimos o k -ésimo momento de Y , se existir, como $m_k = EY^k$. No caso de Y ser integrável, então $\mu_k = E(Y - EY)^k$, se existir, é chamado k -ésimo momento central de Y . Por exemplo, a variância de Y é definida por $\text{Var}(Y) = E\{(Y - EY)^2\}$. É possível generalizar a noção de momentos a vetores aleatórios. Por exemplo, se Y_1 e Y_2 são duas variáveis aleatórias e i, j são números naturais, o momento conjunto de ordem (i, j) de Y_1 e Y_2 é definido por $m_{ij} = E(Y_1^i Y_2^j)$ e de forma análoga se define o momento central conjunto de ordem (i, j) como $\mu_{ij} = E\{(Y_1 - EY_1)^i (Y_2 - EY_2)^j\}$. Desta forma, por exemplo, a covariância de duas variáveis aleatórias X e Y é definida por $\text{Cov}(X, Y) = \mu_{11} = E\{(X - EX)(Y - EY)\}$. No caso do vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_n)$, a matriz de covariância é dada por $\Sigma = (\sigma_{ij})_{n \times n}$ com $\sigma_{ij} = \text{Cov}(Y_i, Y_j)$, $i, j = 1, \dots, n$. Para maiores detalhes sobre esses fundamentos, ver Billingsley (1995).

Suponhamos que $E(\exp(t_0|Y|)) < \infty$ para algum $t_0 > 0$. Neste caso, a aplicação M_Y definida por

$$M_Y(t) = E(\exp(tY)), \quad |t| \leq t_0,$$

é chamada função *geradora de momentos* de Y . Se $M_Y(t)$ está definida numa vizinhança $\{t : |t| < t_0\}$ de zero, então todos os momentos de Y são finitos e

$$M_Y(t) = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k = \sum_{k=0}^{\infty} \frac{m_k}{k!} t^k, \quad |t| < t_0.$$

Além disso, M_Y tem derivadas de toda ordem em $t = 0$ e $\frac{d^k}{dt^k} M_Y(t)|_{t=0} = E(X^k) = m_k$. Definimos a função *geradora de cumulantes* de Y como $K_Y(t) = \log M_Y(t)$ e pode-se representá-la por uma expansão de Taylor da forma $K_Y(t) = \sum_{j=0}^{\infty} \frac{\kappa_j}{j!} t^j$. As constantes $\kappa_1, \kappa_2, \dots$ são chamadas cumulantes de Y e são obtidas pela diferenciação de $K(t)$ da forma

$$\frac{d^j}{dt^j} K_Y(t)|_{t=0} = \kappa_j,$$

com $\kappa_0 = K_Y(0) = 0$. Os dois primeiros cumulantes são $\kappa_1 = EY$, $\kappa_2 = \text{Var}(Y)$. Pode-se mostrar que existe uma relação biunívoca entre momentos e cumulantes, já que o j -ésimo cumulante de Y é uma função dos primeiros j -ésimos momentos de Y e vice-versa. Isto pode ser verificado expandindo $M_Y(t)$ na forma $M_Y(t) = 1 + \sum_{j=1}^{\infty} \frac{m_j}{j!} t^j$ e

$$K_Y(t) = \log(M_Y(t)) = \sum_{j=0}^{\infty} \frac{\kappa_j}{j!} t^j = \log\left(1 + \sum_{j=1}^{\infty} \frac{m_j}{j!} t^j\right).$$

Então, usando uma expansão em série de Taylor de $\log(1+x)$ ao redor do ponto $x = 0$, temos que

$$\log\left(1 + \sum_{j=1}^{\infty} \frac{m_j}{j!} t^j\right) = m_1 t + \frac{1}{2!}(m_2 - m_1^2)t^2 + \frac{1}{3!}(m_3 - 3m_1 m_2 + 2m_1^3)t^3 + \dots,$$

de onde se tem que $\kappa_1 = m_1$, $\kappa_2 = (m_2 - m_1^2)$, $\kappa_3 = (m_3 - 3m_1 m_2 + 2m_1^3)$, e assim por diante.

O terceiro cumulante de Y é uma medida de assimetria, no sentido de que se $\kappa_3 = E(Y - \kappa_1)^3$ for nulo, então Y é uma v.a. simétrica. Os coeficientes de assimetria $\gamma_1 = \mu_3/\sigma^3$ e de excesso de curtose $\gamma_2 = (\mu_4/\sigma^4) - 3$ podem ser escritos como $\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}}$ e $\gamma_2 = \kappa_4/\kappa_2^2$.

A análise dos cumulantes é importante no estudo assintótico de estatísticas, já que estes se anulam para a distribuição normal, para $j > 2$. Uma discussão detalhada sobre a teoria de momentos e cumulantes pode ser encontrada em McCullagh (1987).

1.3. Estimadores de máxima verossimilhança

Neste trabalho, a função f é tipicamente desconhecida; portanto, o objetivo é tentar obter conclusões acerca de f com base no conhecimento da observação de $\mathcal{Y}=\underline{y}$. Consideramos um modelo paramétrico como um conjunto $\mathcal{D}$ de possíveis funções de densidade, onde se encontra nossa verdadeira função de densidade f . O conjunto $\mathcal{D}$ é caracterizado por um parâmetro θ que assume valores num espaço paramétrico $\Theta \subseteq \mathbb{R}^k$, tal que

$$\mathcal{D} = \{f(\cdot; \theta) : \theta \in \Theta\},$$

onde $f(\cdot; \theta)$ é a função modelo. Neste contexto, a inferência paramétrica propõe e avalia mecanismos de seleção de estatísticas $T_n = T_n(\mathcal{Y})$ como funções amostrais apropriadas para estimar θ , i.e., “adivinhar” seu verdadeiro valor, especificando, conseqüentemente, o modelo. Desta forma, a análise estatística consiste na especificação de um modelo para os dados (Severini, 2000, p. 1). Tais estatísticas, utilizadas para estimar θ , são chamadas estimadores de θ . No estudo de tais estatísticas esperamos que elas possuam boas propriedades.

Uma propriedade desejável é a suficiência. Suponhamos que $T_n(\mathcal{Y}) = t$ é conhecido, então, T_n é suficiente para θ se a distribuição condicional de $\mathcal{Y}$ dado que $T_n(\mathcal{Y}) = t$ não depende do valor de θ , i.e., T_n esgota toda a informação existente na amostra a respeito de θ . Além disso, se $\hat{\theta}_1$ e $\hat{\theta}_2$ são dois estimadores de θ , dizemos que $\hat{\theta}_1$ é mais eficiente que $\hat{\theta}_2$ se $\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$ com $\hat{\theta}_1 \neq \hat{\theta}_2$; esta propriedade é fundamental no momento de escolher o “melhor” modelo. Adicionalmente, um estimador $\hat{\theta}$ de θ é não-viesado, se $E(\hat{\theta} - \theta) = 0$, $\forall \theta \in \Theta$; se $E(\hat{\theta} - \theta) \neq 0$, definimos o viés de $\hat{\theta}$ como

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta,$$

sendo desejável, em princípio, que o viés seja zero ou tenda a zero quando o tamanho de amostra n tende para o infinito. Neste último caso, dizemos que $\hat{\theta}$ é assintoticamente não-viesado. Consistência é outra propriedade desejável. Uma sequência de estimadores $\{\hat{\theta}_n\}_{n \geq 1}$ de θ é dita ser consistente, se $\hat{\theta}_n$ converge em probabilidade para θ . Nesta classe de estimadores estão aqueles assintoticamente não-viesados cuja precisão aumenta com o aumento do tamanho de amostra n .

Talvez a mais importante ferramenta de estimação paramétrica seja o método de máxima verossimilhança. Suponhamos que a distribuição do vetor $\underline{y}$ de dados observados possa ser modelada através da densidade $f(\underline{y}; \underline{\theta})$ de um vetor aleatório $\mathcal{Y}$, onde o parâmetro $\underline{\theta}$ da distribuição é desconhecido, $\underline{\theta} \in \Theta \subseteq \mathbb{R}^k$. A função de verossimilhança para $\underline{\theta}$ baseada na observação $\mathcal{Y}=\underline{y}$ é dada por

$$L(\underline{\theta}) \equiv L(\underline{\theta}; \underline{y}) = f(\underline{y}; \underline{\theta}), \quad \underline{\theta} \in \Theta, \quad \Theta \subseteq \mathbb{R}^k,$$

sendo vista como uma função de $\underline{\theta}$ para $\underline{y}$ fixo. Dada a função $L(\underline{\theta})$, o estimador de máxima verossimilhança $\hat{\underline{\theta}} = \hat{\underline{\theta}}(\underline{y})$ é o valor que maximiza $L(\underline{\theta})$. Formalmente, o estimador de máxima verossimilhança é definido como a variável aleatória

$$\hat{\underline{\theta}} = \operatorname{argmax}_{\underline{\theta} \in \Theta} L(\underline{\theta}).$$

É conveniente trabalhar com o logaritmo da função de verossimilhança, que é denotado por

$$\ell(\underline{\theta}) = \log(L(\underline{\theta})) = \ell(\underline{\theta}; \underline{y}).$$

Seja $\underline{\theta} = (\theta_1, \dots, \theta_k)$. Define-se a função escore $U(\underline{\theta}) = (U_1(\underline{\theta}), \dots, U_k(\underline{\theta}))$ por

$$\begin{aligned} U_r(\underline{\theta}) &= \partial \ell(\underline{\theta}) / \partial \theta_r, \quad r = 1, \dots, k, \\ U(\underline{\theta}) &= \nabla_{\underline{\theta}} \ell(\underline{\theta}), \end{aligned}$$

onde $\nabla_{\underline{\theta}} = (\partial / \partial \theta_1, \dots, \partial / \partial \theta_k)^\top$ é o operador gradiente. A função escore descreve como a função de log-verossimilhança varia com $\underline{\theta}$.

1.4. Condições de regularidade

As seguintes condições de regularidade, quase sempre satisfeitas na prática, são de grande auxílio no desenvolvimento da teoria assintótica de modelos paramétricos, em particular, no estudo de estimadores de máxima verossimilhança. Tais condições permitem justificar e delimitar os erros de aproximação em expansões de séries de Taylor e provar propriedades como consistência, unicidade, eficiência e suficiência dos estimadores de máxima verossimilhança. Assim, consideramos as seguintes suposições (Cox & Hinkley, 1974):

- (i) Se $\underline{\theta}_1 \neq \underline{\theta}_2$, $\underline{\theta}_1, \underline{\theta}_2 \in \Theta$, então as medidas de probabilidade $P_{\underline{\theta}_1}$ e $P_{\underline{\theta}_2}$ associadas a $f(\cdot; \underline{\theta}_1)$, $f(\cdot; \underline{\theta}_2)$ são tais que $P_{\underline{\theta}_1} \neq P_{\underline{\theta}_2}$, i.e., as distribuições são identificáveis.
- (ii) As distribuições $P_{\underline{\theta}}$ têm o mesmo suporte para todo $\underline{\theta} \in \Theta$, ou seja, o conjunto $\mathcal{C} = \{y : f(y; \underline{\theta}) > 0\}$ é o mesmo para todo $\underline{\theta}$.

As suposições (iii)-(v) abaixo garantem a regularidade de $f(y; \underline{\theta})$ como função de $\underline{\theta}$ e a existência de um conjunto aberto $\Theta_1 \subseteq \Theta$ tal que o parâmetro verdadeiro $\underline{\theta}_0$ pertença a Θ_1 . Aqui, $E_{\underline{\theta}_0}$ é o valor esperado supondo que $\underline{\theta}_0$ é o valor verdadeiro do parâmetro.

- (iii) Existe $\Theta_1 \subseteq \Theta$ aberto, com $\underline{\theta}_0 \in \Theta_1$, tal que a função de densidade $f(y; \underline{\theta})$ admite derivadas de terceira ordem em relação a $\underline{\theta}$ para todo $\underline{\theta} \in \Theta_1$.
- (iv) $E_{\underline{\theta}_0}\{U(\underline{\theta})\} = \underline{0}$ e a matriz $K(\underline{\theta}) = E_{\underline{\theta}_0}\{U(\underline{\theta})U(\underline{\theta})^\top\}$ existe e é positiva-definida para todo $\underline{\theta} \in \Theta_1$.
- (v) Existem funções $H_{rsu}(y)$ não envolvendo $\underline{\theta}$ tais que, para $r, s, u = 1, \dots, k$,

$$\left| \frac{\partial^3 \log f(y; \underline{\theta})}{\partial \theta_r \partial \theta_s \partial \theta_u} \right| < H_{rsu}(y),$$

para todo $\underline{\theta} \in \Theta_1$, onde $E_{\underline{\theta}_0}\{H_{rsu}(Y)\} < \infty$.

A condição (iii) garante a existência das derivadas até terceira ordem de $f(y; \underline{\theta})$ em Θ_1 , (iv) indica que $K(\underline{\theta})$ é finita e positiva-definida numa vizinhança aberta de $\underline{\theta}_0$, o que é uma condição suficiente para garantir que $\hat{\underline{\theta}}$ é um máximo local nesta vizinhança, e (v) afirma que as terceiras derivadas de $\ell(\underline{\theta})$ são limitadas por uma função integrável de y . Um modelo que apresenta estas condições de regularidade é chamado de *modelo regular*. Em modelos regulares, a diferenciação com respeito a θ e integração sobre o espaço amostral de algumas funções mais importantes podem ser trocadas. Seja $J(\underline{\theta}) = -\partial^2 \ell(\underline{\theta}; Y) / \partial \underline{\theta} \partial \underline{\theta}^\top$ o negativo da matriz hessiana da função de log-verossimilhança; J é chamada matriz de informação observada (Fisher, 1925). Como consequência das condições de regularidade, temos que

$$\text{Cov}_{\underline{\theta}}\{U(\underline{\theta})\} = E_{\underline{\theta}}\{U(\underline{\theta})U(\underline{\theta})^\top\} = E_{\underline{\theta}}\{J(\underline{\theta})\},$$

ou seja,

$$K(\underline{\theta}) = E_{\underline{\theta}}\{-\nabla \nabla^\top \ell(\underline{\theta})\}.$$

A matriz $K(\underline{\theta})$ é chamada de matriz de informação de Fisher de $\underline{\theta}$ e corresponde à matriz da condição de regularidade (iv). Uma justificativa qualitativa para o uso da palavra “informação” é que se a matriz J “cresce”, i.e., se é maior a informação sobre $\underline{\theta}$, menor será a dispersão de $\hat{\underline{\theta}}$ ao redor de $\underline{\theta}$ (Barndorff-Nielsen & Cox, 1994).

Sob as condições de regularidade apresentadas acima, $\ell(\underline{\theta})$ é diferenciável em $\underline{\theta}$, $\hat{\underline{\theta}}$ existe, $\hat{\underline{\theta}}$ satisfaz a equação de verossimilhança

$$U(\hat{\underline{\theta}}) = \nabla_{\underline{\theta}} \ell(\hat{\underline{\theta}}) = \nabla_{\underline{\theta}} \ell(\hat{\underline{\theta}}, \underline{y}) = \underline{0}, \quad (1.4.1)$$

e

$$\ell(\hat{\underline{\theta}}) \geq \ell(\underline{\theta}), \quad \forall \underline{\theta} \in \Theta.$$

Além disso, se $\underline{y} = (y_1, \dots, y_n)$, onde os Y_i 's são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com densidades $f(y_i; \underline{\theta})$, então, a função de verossimilhança será

$$L(\underline{\theta}) = \prod_{i=1}^n f(y_i; \underline{\theta}) \quad (1.4.2)$$

e a função de log-verossimilhança

$$\ell(\underline{\theta}) = \sum_{i=1}^n \log\{f(y_i; \underline{\theta})\}, \quad (1.4.3)$$

sendo a equação de verossimilhança (1.4.1) simplificada na forma

$$\sum_{i=1}^n \nabla_{\underline{\theta}} \log f(y_i; \hat{\underline{\theta}}) = \underline{0}. \quad (1.4.4)$$

Lehmann & Casella (1998, cap. 6) mostram que, sob as condições de regularidade citadas, o estimador de máxima verossimilhança $\hat{\underline{\theta}}$, obtido como raiz única da equação (1.4.4), apresenta as seguintes propriedades:

- (1) É consistente.
- (2) É suficiente.
- (3) É intrinsecamente assintoticamente eficiente, i.e., entre todos os estimadores consistentes de $\underline{\theta}$, o estimador de máxima verossimilhança possui variância assintótica mínima.
- (4) É invariante sob transformações biunívocas, i.e., se $\hat{\underline{\theta}}$ é o estimador de máxima verossimilhança de $\underline{\theta}$ e g é uma função bijetora de $\underline{\theta}$, então $g(\hat{\underline{\theta}})$ é o estimador de máxima verossimilhança de $g(\underline{\theta})$.
- (5) Assintoticamente se distribui como uma v.a. normal k -variada da forma

$$\hat{\underline{\theta}} \stackrel{A}{\sim} \mathcal{N}_k(\underline{\theta}, K^{-1}(\underline{\theta})). \quad (1.4.5)$$

- (6) Em geral, é viesado, embora seja assintoticamente não-viesado.

1.5. Identidades de Bartlett

Sob as condições de regularidade listadas e supondo (1.4.2), introduzimos a seguinte notação, dada por Lawley (1956), para as derivadas da log-verossimilhança (1.4.3), na qual todos os índices variam de 1 a k : $U_r = \partial \ell / \partial \theta_r$, $U_{rs} = \partial^2 \ell / \partial \theta_r \partial \theta_s$, etc. Os momentos conjuntos de derivadas de $\ell(\underline{\theta})$ são $\mu_r = E(U_r)$, $\mu_{rs} = E(U_{rs})$, $\mu_{r,s} = E(U_r U_s)$, $\mu_{r,st} = E(U_r U_{st})$, e assim por diante. Como $\mu_r = 0$, os cumulantes conjuntos expressos em termos de momentos são $\kappa_{r,s} = \mu_{r,s}$, $\kappa_{rs} = \mu_{rs}$, $\kappa_{rs,t} = \mu_{rs,t}$, $\kappa_{r,s,t} = \mu_{r,s,t}$, e $\kappa_{r,s,t,u} = \mu_{r,s,t,u} - \sum_{(3)} \mu_{r,s} \mu_{t,u}$, onde $\sum_{(j)}$ representa o somatório sobre todas as j -ésimas combinações de índices, etc. Como caso particular, $-\kappa^{rs} = \kappa^{r,s}$ representa o elemento (r, s) da inversa da matriz de informação $K^{-1}(\underline{\theta})$. As derivadas dos cumulantes são representadas com sobrescritos: $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \theta_t$, $\kappa_{rs}^{(tu)} = \partial \kappa_{rs} / \partial \theta_t \partial \theta_u$, $\kappa_{rst}^{(u)} = \partial \kappa_{rst} / \partial \theta_u$, etc. Os momentos e cumulantes são intra-relacionados, mas satisfazem algumas equações que facilitam seus cálculos. Estas equações, que decorrem de condições de regularidade, são denominadas *identidades de Bartlett*, as mais importantes sendo $\kappa_r = 0$ e $\kappa_{rs} + \kappa_{r,s} = 0$. Além disso, pode-se mostrar que, sobre a amostra total, os cumulantes são de ordem $O(n)^1$. A idéia de Bartlett é bastante engenhosa; supondo problemas regulares tem-se

$$\frac{\partial}{\partial \underline{\theta}} E\{T(\underline{Y})\} = \int T(\underline{y}) \frac{\partial}{\partial \underline{\theta}} f(\underline{y}; \underline{\theta}) d\underline{y}, \quad (1.5.1)$$

para diversas estatísticas $T(\underline{Y})$, i.e., pode-se inverter a ordem das operações de diferenciação em relação a $\underline{\theta}$ e integração com respeito a $\underline{y}$. Expressando a identidade (1.5.1) em termos dos momentos e diferenciando sucessivamente com relação às componentes de $\underline{\theta}$ se encontram as ditas identidades. Assim, por exemplo, da definição da função de escore temos que $U_r = L_r / L$, com $L_r = \partial L / \partial \theta_r$. Diferenciando $\int L d\underline{y} = 1$ em relação a θ_r vem $\int L_r d\underline{y} = 0$, i.e., $\kappa_r = E(U_r) = 0$, onde diferenciando em relação a θ_s encontramos $\int (U_{rs} L + U_r U_s L) d\underline{y} = 0$, e

¹ Neste contexto, se $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são duas seqüências de números reais dizemos que a_n é de ordem menor que b_n e escrevemos como $a_n = o(b_n)$ se $\lim_{n \rightarrow \infty} a_n / b_n = 0$. Dizemos que a_n é de ordem pelo menos b_n , denotado por $a_n = O(b_n)$, se existe um número real $M > 0$ tal que $|a_n / b_n| \leq M$. Observamos que se $a_n = o(b_n)$ então $a_n = O(b_n)$ e que quando $b_n \rightarrow 0$ a ordem dá noção da taxa de convergência para zero de a_n .

assim obtemos $\kappa_{rs} + \kappa_{r,s} = 0$. Diferenciando novamente a integral em relação a θ_t , obtém-se $\kappa_{r,s,t} + \kappa_{rst} + \kappa_{s,rt} + \kappa_{r,st} = 0$, e de forma análoga outras identidades são deduzidas. A grande vantagem das identidades de Bartlett é facilitar a obtenção dos cumulantes κ 's, já que determinada parametrização pode conduzir a um cálculo simples de alguns cumulantes, sendo os demais calculados indiretamente através destas identidades. Esses cumulantes têm uma grande aplicabilidade no cálculo do viés de segunda ordem dos estimadores de máxima verossimilhança, como será visto mais adiante. Uma discussão profunda sobre as identidades de Bartlett e suas aplicações pode ser encontrada em Barndorff-Nielsen & Cox (1989, § 5.5).

1.6. Intervalos de confiança assintóticos

Suponha que queremos gerar uma região de confiança para o parâmetro $\underline{\theta} \in \mathbb{R}^k$, baseados num estimador $\hat{\underline{\theta}} \in \mathbb{R}^k$ construído a partir da observação $\underline{y}$ de um vetor $\underline{Y}$ com $F_{\underline{Y}}(\underline{y})$. Assuma que $n^{1/2}(\hat{\underline{\theta}} - \underline{\theta})$ tende em distribuição a uma variável aleatória normal k -variada $\mathcal{N}_k(0, V)$, quando o tamanho de amostra n cresce para infinito, e seja Σ um estimador consistente da matriz de covariância assintótica V . Daí, utilizando a informação contida em $\underline{y}$, obtemos um intervalo de confiança de tamanho aproximado $(1 - \alpha)$ para a r -ésima componente θ_r de $\underline{\theta}$, dado por:

$$(\hat{\theta}_r - z_{1-\frac{\alpha}{2}}(\Sigma_{rr}/n)^{1/2}, \hat{\theta}_r + z_{1-\frac{\alpha}{2}}(\Sigma_{rr}/n)^{1/2}), \quad (1.6.1)$$

onde z_δ é tal que $P(Z \leq z_\delta) = \delta$, Z tem distribuição normal padrão $\mathcal{N}(0, 1)$ e $(\Sigma_{rr}/n)^{1/2}$ é o r -ésimo erro-padrão assintótico de $\hat{\theta}_r$. A probabilidade de cobertura $1 - \alpha$ é chamada *nível nominal*. Assim, (1.6.1) é um resultado assintótico e o *erro de cobertura* é igual à *cobertura nominal* menos a verdadeira probabilidade de cobertura, que pode ser significativo, especialmente quando o tamanho de amostra no experimento estudado não é suficientemente grande.

No caso de modelos paramétricos e ao utilizar o estimador de máxima verossimilhança sob as condições de regularidade dadas na Seção (1.4), $\hat{\underline{\theta}}$ é um estimador consistente de $\underline{\theta}$ e, como consequência de (1.4.5), a quantidade

$$T = (\hat{\underline{\theta}} - \underline{\theta})^\top K(\underline{\theta})^{1/2}$$

é assintoticamente pivotal² e $K(\hat{\underline{\theta}})$ é consistente para a matriz de informação $K(\underline{\theta})$. Assim, para a r -ésima componente de $\underline{\theta}$

$$(\hat{\theta}_r - \theta_r)(K(\underline{\theta})^{rr})^{1/2} \stackrel{A}{\sim} \mathcal{N}(0, 1),$$

onde $K(\underline{\theta})^{rr}$ é o (r, r) -ésimo elemento da matriz $K(\underline{\theta})^{-1}$. Desta forma, ao utilizar (1.6.1) e usando o quantil $\alpha/2$ da distribuição $\mathcal{N}(0, 1)$, podemos construir um intervalo de confiança de tamanho $1 - \alpha$ com erro de cobertura tendendo para zero. Portanto, um intervalo de confiança

² Dizemos que uma variável aleatória é uma quantidade pivotal se sua distribuição não depende de parâmetros desconhecidos

assintótico de $100(1 - \alpha)\%$ para a r -ésima componente θ_r de $\underline{\theta}$ baseado no estimador de máxima verossimilhança é da forma

$$(\hat{\theta}_r - z_{1-\frac{\alpha}{2}}(K(\hat{\underline{\theta}})^{rr})^{1/2}, \hat{\theta}_r + z_{1-\frac{\alpha}{2}}(K(\hat{\underline{\theta}})^{rr})^{1/2}), \quad (1.6.2)$$

com erro padrão assintótico $(K(\hat{\underline{\theta}})^{rr})^{1/2}$. Esta construção pode ser imprecisa na prática, já que a distribuição pode ser bem diferente da assintótica, e a quantidade T pode ter uma distribuição assimétrica ou ser viesada para o vetor nulo, e como estes intervalos são simétricos em torno de θ_r o erro de cobertura pode ser bem diferente de zero. Outra desvantagem é que esta construção não impede a possibilidade de que o intervalo abranja valores que não pertençam ao espaço paramétrico. Para maiores detalhes, ver Davidson & Hinkley (1997, cap. 5) e Efron & Tibshirani (1993, cap. 12).

1.7. Otimização não-linear em inferência estatística

Um dos problemas mais freqüentes na inferência estatística é encontrar o estimador de máxima verossimilhança $\hat{\underline{\theta}}$ de $\underline{\theta}$, que em geral não apresenta uma forma analítica fechada para seu cálculo. Existem algoritmos de busca numérica que no contexto deste trabalho serão de muita utilidade. Entre tais técnicas se encontram os métodos quasi-Newton (com derivadas analíticas), os quais se baseiam na idéia descrita a seguir.

Suponha que ∇g (escalar ou vetorial) pode ser aproximado por uma série de Taylor até primeira ordem, numa vizinhança do ponto $\theta(k)$, na forma

$$\nabla g(\theta) \approx \nabla g(\theta(k)) + H(\theta(k))(\theta - \theta(k)).$$

No método de Newton fazemos $\nabla g(\theta) = 0$ para determinar o próximo ponto da iteração, i.e.,

$$0 = \nabla g(\theta(k)) + H(\theta(k))(\theta - \theta(k)) \Rightarrow \theta = \theta(k) - H(\theta(k))^{-1} \nabla g(\theta(k)),$$

onde θ é uma aproximação para o ponto que maximiza a função. Desta forma, obtemos um aproximação do ponto de máximo de g através da recorrência

$$\theta(k+1) = \theta(k) - H(\theta(k))^{-1} \nabla g(\theta(k))$$

e que podemos generalizar por

$$\theta(k+1) = \theta(k) - s(k)H(\theta(k))^{-1} \nabla g(\theta(k)),$$

onde $s(k)$ é um escalar determinado por algum procedimento de busca linear a partir de $\theta(k)$ na direção $-H(\theta(k))^{-1} \nabla g(\theta(k))$ de forma que $g(\theta(k))$ cresça nesta direção. A velocidade de crescimento de g é caracterizada por

$$\nabla^\top(\theta(k))(\theta - \theta(k)) > 0 \Rightarrow -(\theta - \theta(k))^\top H(\theta(k))(\theta - \theta(k)),$$

Assim, H deve ser negativa-definida para todo k . Se θ está longe do máximo H pode não ser negativa-definida e o incremento pode mover-se na direção errada, provocando a não-convergência do algoritmo. Neste caso, uma opção é utilizar o algoritmo BFGS que não apresenta tal problema.

1.7.1. Algoritmo BFGS

Este algoritmo foi desenvolvido por Broyden, Fletcher, Goldfarf e Shanno. Ele utiliza o princípio do método de Newton, diferenciando-se na forma como é calculado o passo, já que neste caso o passo não depende da k -ésima iteração e da matriz hessiana, pois ao invés de trabalhar com $-H^{-1}$, é utilizada uma sequência de matrizes simétricas e positivas-definidas $Q(k)$, tal que

$$\lim_{k \rightarrow \infty} Q(k) = -H^{-1}.$$

Comumente, toma-se como matriz inicial a identidade da mesma ordem, pois ela é positiva-definida e simétrica, assim conduzindo a aproximações $Q(k)$ positivas-definidas e simétricas. A forma recursiva para obter tais matrizes é dada pela expressão

$$\begin{aligned} Q(k+1) = Q(k) + & \frac{(\theta(k+1) - \theta(k)) \otimes (\theta(k+1) - \theta(k))}{(\theta(k+1) - \theta(k))^\top (\nabla g(\theta(k+1)) - \nabla g(\theta(k)))} \\ & - \frac{[Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))) \otimes [Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))]}{(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))^\top Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))} \\ & + (\nabla g(\theta(k+1)) - \nabla g(\theta(k)))^\top Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k))) U \otimes U, \end{aligned}$$

onde U é o vetor coluna definido por

$$\begin{aligned} U = & \frac{\theta(k+1) - \theta(k)}{(\theta(k+1) - \theta(k))^\top (\nabla g(\theta(k+1)) - \nabla g(\theta(k)))} \\ & - \frac{Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))}{(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))^\top Q(k)(\nabla g(\theta(k+1)) - \nabla g(\theta(k)))}. \end{aligned}$$

O operador $\otimes$ denota o produto direto de dois vetores, cujo resultado para τ, v é uma matriz de elemento $(i, j) = \tau_i v_j$. Assim, mesmo que θ esteja longe do ponto de máximo, as matrizes $Q(k)$ garantem que as iterações se darão na direção crescente. De forma análoga ao método anterior, o máximo é obtido pela recorrência

$$\theta(k+1) = \theta(k) - s(k)Q(k)\nabla g(\theta(k)).$$

Este método está implementado na linguagem de programação `0x` através da função `MaxBFGS`; maiores detalhes, ver Doornik (2001, §. 10.2).

1.7.2. Algoritmo de Newton-Raphson

Um problema muito comum em inferência estatística é resolver sistemas de equações não-lineares, por exemplo, ao tentar procurar os zeros da função score. Entre os métodos mais poderosos para solucionar este tipo de problemas está o método de Newton-Raphson, o qual apresenta algumas desvantagens, como a necessidade de cálculo da matriz jacobiana, que em algumas situações é de difícil obtenção. Similarmente do método apresentado anteriormente, suponhamos que g pode ser aproximada em série de Taylor até primeira ordem ao redor de um ponto θ pertencente a uma vizinhança de $\theta(k)$; logo

$$g(\theta(k)) = g(\theta) + J(\theta)(\theta - \theta(k)),$$

onde J é a matriz jacobiana de derivadas parciais. Como $g(\theta) = 0$, então

$$\theta(k) - \theta = J(\theta)^{-1}g(\theta),$$

logo, usando iterativamente a equação anterior, podemos obter

$$\theta(k+1) = \theta(k) + J(\theta(k))^{-1}g(\theta(k)).$$

O processo anterior é repetido até que a distância entre $\theta(k+1)$ e $\theta(k)$ seja menor que uma tolerância especificada. Este método está implementado na linguagem de programação **Ox** através da função **SolveNLE**. Para maiores detalhes, ver <http://www.nuff.ox.ac.uk/Users/Doornik>. Uma discussão mais detalhada sobre métodos de otimização não-linear, obtenção de zeros de funções e suas implementações em **C**, pode ser encontrada em Press et al. (1992, capítulos 9 e 10).

Capítulo 2

O modelo de regressão beta

2.1 Modelo beta

A distribuição beta é uma das mais usadas para modelar experimentos aleatórios que produzem resultados no intervalo $(0, 1)$, dada a grande flexibilidade de ajuste de seus parâmetros. Bury (1999) lista um conjunto de aplicações da distribuição beta em engenharia. Janardan & Padmanabhan (1986) modelam variáveis hidrológicas usando a distribuição beta. McNally (1990) utiliza a distribuição beta no estudo de algumas variáveis que afetam a reprodutibilidade de vacas. Graham & Hollands (1990) e Milyutin & Yaromenko (1991) usam a distribuição beta para estudar índices relacionados à transmissão de radiação solar. A potência de sinais de radar é modelada por Maffet & Wackerman (1991) através da distribuição beta. Wiley, Herschokoru & Padiou (1989) desenvolvem um modelo beta para estimar a probabilidade de transmissão de HIV durante o contato sexual entre um indivíduo infectado e um indivíduo sadio. Johnson, Kotz & Balakrishnan (1995, p. 235) observam: “the beta distributions are among the most frequently employed to model theoretical distributions”; além disso, mostram como a distribuição beta é usada no ajuste de certos testes estatísticos de verossimilhança e como é possível ser utilizada no ajuste de dados empíricos. A seguir, consideramos algumas propriedades da distribuição beta, como sua função de densidade, função de distribuição acumulada e seus primeiros momentos.

2.2. Definição e estimação do modelo beta

Seja Y uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathfrak{F}, P)$. Dizemos que Y tem distribuição beta com parâmetros $p, q > 0$, denotado por $Y \sim \mathcal{B}(p, q)$, se sua função de densidade é dada por

$$f(y; p, q) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, & y \in (0, 1), \\ 0, & \text{caso contrário.} \end{cases} \quad (2.2.1)$$

Temos que $f(y) = f(y; p, q)$ é efetivamente uma densidade. Inicialmente, observamos que $f(y)$ não assume valores negativos, já que para qualquer valor $y \in (0, 1)$ a função $f(y)$ é produto de funções positivas, portanto $f(y) \geq 0$. Adicionalmente, como veremos a seguir, a densidade em (2.2.1) integra um. Seja $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$, $x > 0$, a função gama.

Temos

$$\begin{aligned}\Gamma(p)\Gamma(q) &= \left(\int_0^\infty x^{p-1} e^{-x} dx \right) \left(\int_0^\infty y^{q-1} e^{-y} dy \right) \\ &= \int_0^\infty \int_0^\infty x^{p-1} y^{q-1} e^{-(x+y)} dx dy.\end{aligned}$$

Fazendo a troca de variáveis $u = x/(x+y)$, com $u \in (0,1)$, temos que $x = uy/(1-u)$ e $dx = ydu/(1-u)^2$ e, assim,

$$\begin{aligned}\Gamma(p)\Gamma(q) &= \int_0^\infty \int_0^1 \frac{1}{(1-u)^2} u^{p-1} (1-u)^{1-p} y^{p-1} y^{q-1} e^{-y/(1-u)} y du dy \\ &= \int_0^\infty \int_0^1 \frac{u^{p-1}}{(1-u)^{p+1}} y^{p+q-1} e^{-y/(1-u)} du dy.\end{aligned}$$

Se $v = y/(1-u)$, então $y = v(1-u)$ e $dy = (1-u)dv$, onde $v \in (0, \infty)$. Desta forma,

$$\begin{aligned}\Gamma(p)\Gamma(q) &= \int_0^\infty \int_0^1 u^{p-1} (1-u)^{q-1} v^{p+q-1} e^{-v} du dv \\ &= \int_0^\infty v^{p+q-1} e^{-v} dv \int_0^1 u^{p-1} (1-u)^{q-1} du \\ &= \Gamma(p+q) \int_0^1 u^{p-1} (1-u)^{q-1} du.\end{aligned}$$

Portanto,

$$\Gamma(p)\Gamma(q) = \Gamma(p+q) \int_0^1 x^{p-1} (1-x)^{q-1} dx.$$

Logo,

$$\int_{-\infty}^\infty f(y) dy = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 y^{p-1} (1-y)^{q-1} dy = 1.$$

Desta forma $f(y)$ é de fato uma densidade.

Os parâmetros p e q são parâmetros de ajuste, pois através da escolha de p e q podem ser obtidas diferentes distribuições em $(0,1)$. Quando $p = q = 1/2$, a distribuição beta se reduz à chamada “distribuição arco seno,” a qual é útil no estudo de passeios aleatórios. Distribuições beta para as quais $p+q = 1$ com $p \neq 1/2$ são conhecidas como distribuições “arco seno generalizadas”. Se $p = q$ obtemos densidades simétricas e, quando $p = q = 1$, temos a distribuição uniforme. Assim, a distribuição beta na realidade é uma família de distribuições, como pode ser visto na Figura 2.2.1.

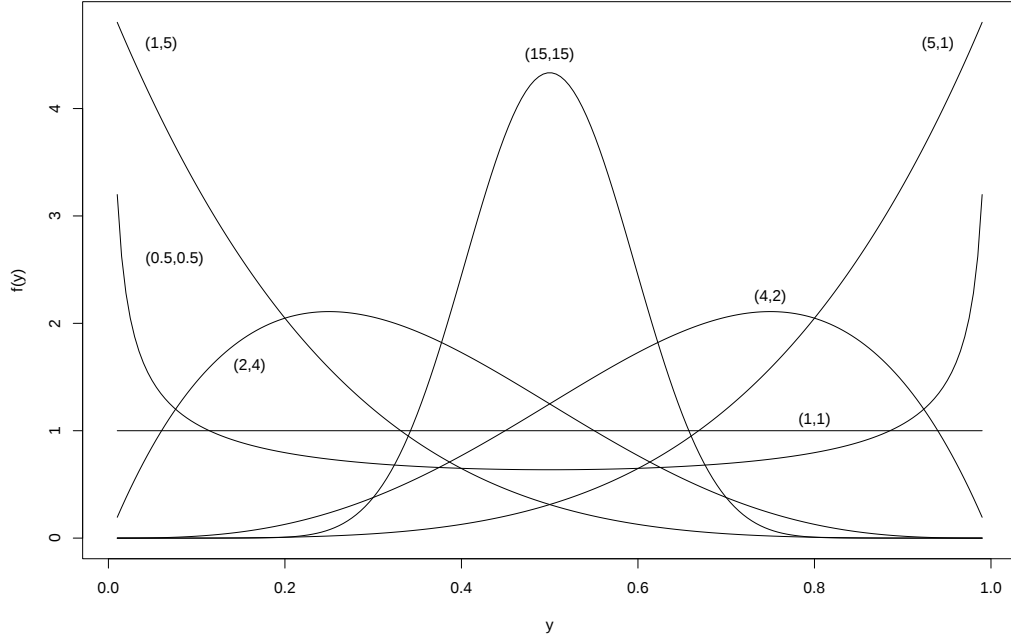


Figura 2.2.1: Gráfico da densidade beta para diferentes valores de p e q .

Se fixamos $q = 2$, no lado esquerdo da Figura 2.2.2, temos o gráfico de várias densidades beta geradas para diferentes escolhas de p ; fixando $p = 2$ no lado direito da Figura 2.2.2 temos o gráfico de várias densidades beta geradas para diferentes escolhas de q . Observamos claramente que ocorre uma reflexão em torno da reta $y = 1/2$ quando permutamos p e q , devido à expressão da densidade como função de y e $1 - y$.

Definimos a função beta completa $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$ e a função beta incompleta como

$$B_y(p, q) = \int_0^y t^{p-1}(1-t)^{q-1} dt, \quad 0 \leq y \leq 1.$$

A função de distribuição acumulada beta é dada pela expressão

$$\begin{aligned} \mathcal{B}(y; p, q) &= P(Y \leq y) = \int_{-\infty}^y f(t) dt = \int_0^y \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt \\ &= \frac{1}{B(p, q)} \int_0^y t^{p-1} (1-t)^{q-1} dt = \frac{B_y(p, q)}{B(p, q)} = I_y(p, q), \end{aligned}$$

onde

$$I_y(p, q) = \frac{y^p}{pB(p, q)} F(p, 1-q, 1+p; y), \quad 0 \leq y \leq 1,$$

a função $F(\alpha, \beta, \gamma; y)$ sendo conhecida como função hipergeométrica de Gauss, e sendo obtida como solução da equação diferencial

$$y(1-y) \frac{d^2 \omega}{dy^2} + (\gamma - (\alpha + \beta + 1)y) \frac{d\omega}{dy} - \alpha\beta\omega = 0$$

na forma

$$\omega(y) = F(\alpha, \beta, \gamma; y) = \sum_{j=0}^{\infty} b_j y^j = \sum_{j=0}^{\infty} c_j;$$

c_j dado por $c_0 = 1$, $c_{j+1} = (\alpha + j)(\beta + j)/(\gamma + j)(1 + j)yc_j$. Desta maneira, a análise da distribuição beta está diretamente relacionada com o estudo de equações diferenciais de segunda ordem. Uma discussão interessante desta relação, e sua aplicação na modelagem de dados hidrológicos pode ser vista em <http://www2.dnv.com/ocean/bk/c/a22/s0.htm>.

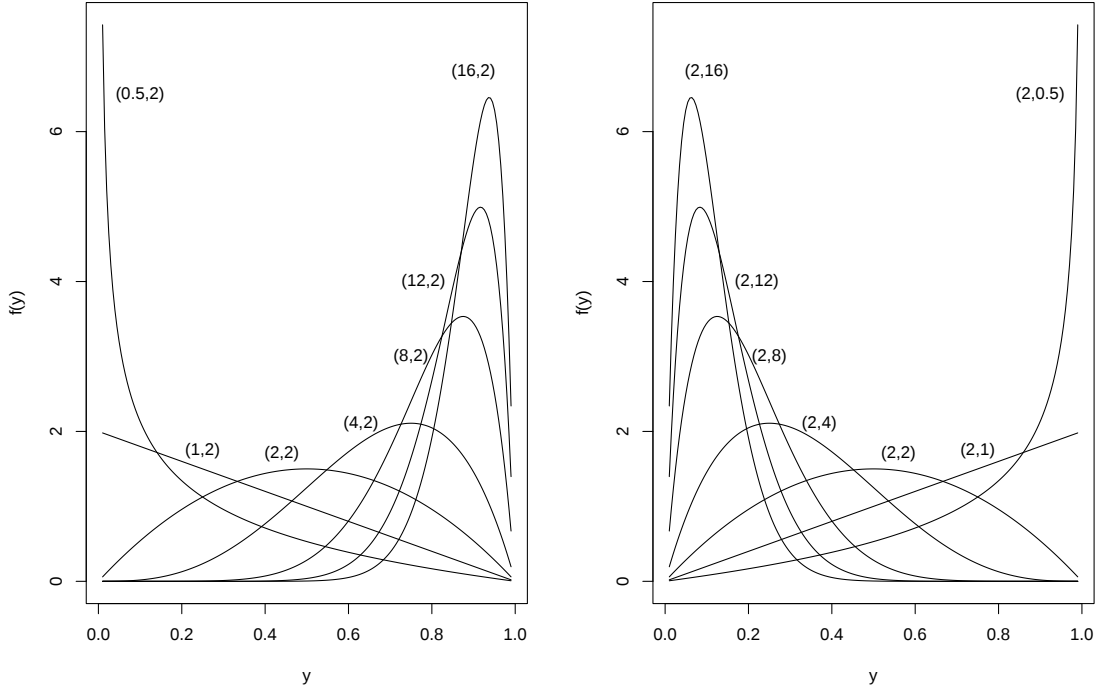


Figura 2.2.2: Gráfico da densidade beta para diferentes valores de p e q fixando $q = 2$ (esquerda) e $p = 2$ (direita).

A função geradora de momentos da distribuição beta é

$$\varphi(s) = \int_0^1 e^{st} f(t) dt = \frac{1}{B(p, q)} \int_0^1 e^{st} t^{p-1} (1-t)^{q-1} dt = M(q, p+q; -s),$$

onde a função $M(p, q; y)$ é chamada função hipergeométrica de confluência, a qual é solução da equação diferencial

$$y \frac{d^2 u}{dy^2} + (q - y) \frac{du}{dy} + pu = 0,$$

com

$$u(y) = M(p, q; y) = \sum_{j=0}^{\infty} b_j y^j = \sum_{j=0}^{\infty} c_j,$$

os c_j 's da forma

$$c_0 = 1 \quad \text{e} \quad c_{j+1} = \frac{(p+j)}{(q+j)(1+j)} y c_j.$$

É importante conhecer a média, variância, assimetria e curtose da distribuição beta para determinar seu comportamento sob diferentes escolhas dos parâmetros. O momento de ordem n ao redor de zero da distribuição $\mathcal{B}(p, q)$ é obtido através de

$$\begin{aligned} E(Y^n) = m_n &= \int_0^1 y^n f(y) dy = \int_0^1 y^n \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1} dy \\ &= \int_0^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{n+p-1} (1-y)^{q-1} dy \\ &= \frac{B(p+n, q)}{B(p, q)} = \frac{\Gamma(p+q)\Gamma(p+n)}{\Gamma(p+q+n)\Gamma(p)} \\ &= \frac{p(n)}{(p+q)_{(n)}} \quad \text{onde,} \quad a_{(n)} = a(a+1)(a+2) \cdots (a+n-1). \end{aligned}$$

Desta forma,

$$\begin{aligned} m_0 &= 1, \\ m_1 &= \frac{p(1)}{(p+q)_{(1)}} = \frac{p}{p+q} = EY, \\ m_2 &= \frac{p(2)}{(p+q)_{(2)}} = \frac{p(p+1)}{(p+q)(p+q+1)} = EY^2. \end{aligned} \tag{2.2.2}$$

A variância é

$$\text{Var}(Y) = EY^2 - (EY)^2 = \frac{p(p+1)}{(p+q)(p+q+1)} - \frac{p^2}{(p+q)^2} = \frac{p q}{(p+q)^2(p+q+1)}. \tag{2.2.3}$$

Através dos momentos, podem ser obtidos os valores da assimetria $E[(Y - \mu)]^3 = \mu_3$ e curtose $E[(Y - \mu)]^4 = \mu_4$, onde μ é a média da distribuição:

$$\begin{aligned} \mu_3 &= 2(q-p) \sqrt{p^{-1} + q^{-1} + (pq)^{-1}} (p+q+2)^{-1}, \\ \mu_4 &= \frac{3(p+q+1)\{2(p+q)^2 + p q(p+q-6)\}}{\{p q(p+q+2)(p+q+3)\}}. \end{aligned}$$

As quantidades μ_3 e μ_4 medem a assimetria e a grossura de caudas da distribuição, respectivamente. Para mais detalhes, ver Greene (2000, p. 64).

É de se notar que a densidade (2.2.1) é um produto de funções contínuas que admitem derivadas de todas as ordens com relação a p e q (Rudin, 1976) e sua log-densidade é da forma

$$\log \Gamma(p+q) - \log \Gamma(p) - \log \Gamma(q) + (p-1) \log y + (q-1) \log(1-y).$$

Assim, as primeiras derivadas da log-densidade da distribuição beta são

$$\begin{aligned}\frac{\partial \log f(y)}{\partial p} &= \psi(p+q) - \psi(p) + \log y, \\ \frac{\partial \log f(y)}{\partial q} &= \psi(p+q) - \psi(q) + \log(1-y),\end{aligned}$$

com $\psi(\cdot)$ denotando a função digama.³ As segundas derivadas da log-densidade são:

$$\frac{\partial^2 \log f(y)}{\partial p^2} = \psi'(p+q) - \psi'(p), \quad (2.2.4)$$

$$\frac{\partial^2 \log f(y)}{\partial p \partial q} = \psi'(p+q), \quad (2.2.5)$$

$$\frac{\partial^2 \log f(y)}{\partial q^2} = \psi'(p+q) - \psi'(q), \quad (2.2.6)$$

$\psi'(\cdot)$ sendo a função trigama. As terceiras derivadas da log-densidade são

$$\frac{\partial^3 \log f(y)}{\partial p^3} = \psi''(p+q) - \psi''(p),$$

$$\frac{\partial^3 \log f(y)}{\partial p^2 \partial q} = \psi''(p+q),$$

$$\frac{\partial^3 \log f(y)}{\partial p \partial q^2} = \psi''(p+q) - \psi''(q),$$

$$\frac{\partial^3 \log f(y)}{\partial q^3} = \psi''(p+q) - \psi''(q),$$

onde $\psi''(\cdot)$ é a função tetragama. Como se observa de (2.2.4) até (2.2.6), as derivadas de ordem maior ou igual a 2 da função de log-verossimilhança com respeito aos parâmetros não dependem da variável aleatória y , que verifica a condição de regularidade (v) da Seção 1.4. Adicionalmente, observamos através da Figura 2.2.3 que para diferentes escolhas dos parâmetros (p, q) , suas respectivas funções de distribuição são diferentes, i.e., a distribuição beta é totalmente identificável, valendo assim a condição de regularidade (i) da Seção 1.4. Além disso, observamos que o suporte da distribuição $\mathcal{C} = \{y : f(y; p, q) > 0\} = (0, 1)$, o qual não depende dos parâmetros (p, q) definidos no espaço paramétrico $\Theta = (0, \infty) \times (0, \infty)$.

Seja $\mathcal{Y} = (y_1, \dots, y_n)$ onde os y_i 's são variáveis aleatórias independentes com densidades $f(y_i; p, q) \sim \mathcal{B}(p, q)$, então, a função de verossimilhança é dada pela expressão

$$L(p, q) = \prod_{i=1}^n f(y_i; (p, q)) = \prod_{i=1}^n \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y_i^{p-1} (1-y_i)^{q-1}.$$

³ A função digama é uma das funções denominadas genericamente de poligama, que são definidas para $n = 0, 1, \dots$ como $\psi^{(n)}(x) = \frac{d^{n+1}}{dx^{n+1}} \log \Gamma(x)$, $x > 0$. Para maiores detalhes, ver Bury (1999, §14.8) e Dishon & Weiss (1980).

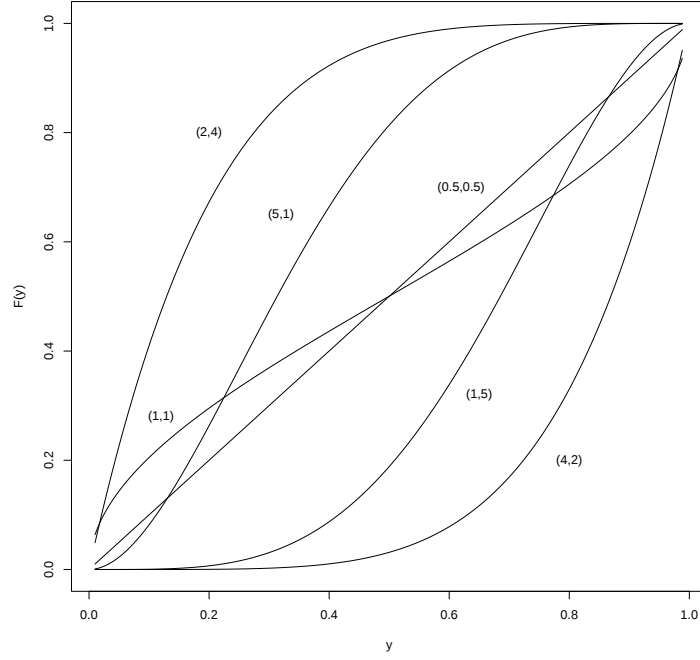


Figura 2.2.3: Gráfico da função de distribuição acumulada beta para diferentes valores de p e q .

Logo, a função de log-verossimilhança $\ell(p, q)$ é da forma

$$\ell(p, q) = \log(L(p, q)) = \sum_{i=1}^n \log \left\{ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y_i^{p-1} (1-y_i)^{q-1} \right\}.$$

Assim,

$$\begin{aligned} \ell(p, q) &= \sum_{i=1}^n \left\{ \log \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} + \log y_i^{p-1} + \log(1-y_i)^{q-1} \right\} \\ &= \sum_{i=1}^n \left\{ \log \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} + (p-1) \log(y_i) + (q-1) \log(1-y_i) \right\} \\ &= n \log \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} + (p-1) \sum_{i=1}^n \log(y_i) + (q-1) \sum_{i=1}^n \log(1-y_i) \\ &= n \left\{ \log \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} + (p-1) \log \left\{ \left(\prod_{i=1}^n y_i \right)^{1/n} \right\} + (q-1) \log \left\{ \left(\prod_{i=1}^n (1-y_i) \right)^{1/n} \right\} \right\}. \end{aligned}$$

Desta forma,

$$\begin{aligned} \ell(p, q) &= n \left\{ \log \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} + (p-1) \log g_1 + (q-1) \log g_2 \right\} \\ &= n \{ A(p, q) + (p-1) \log g_1 + (q-1) \log g_2 \}, \end{aligned} \tag{2.2.7}$$

onde $A(p, q) = \log\{\Gamma(p+q)[\Gamma(p)\Gamma(q)]^{-1}\}$ e $g_1 = (\prod_{i=1}^n y_i)^{1/n}$, $g_2 = (\prod_{i=1}^n (1 - y_i))^{1/n}$ são as médias geométricas dos y_i 's e dos $(1 - y_i)$'s, respectivamente. De (2.2.7) observamos, através do critério de fatoração, que g_1 e g_2 são estatísticas conjuntamente suficientes para p e q . O vetor escore

$$U(\underline{\theta}) = \nabla_{\underline{\theta}} \log(L(\underline{\theta})) = \begin{pmatrix} \psi(p) - \psi(p+q) - \log g_1 \\ \psi(q) - \psi(p+q) - \log g_2 \end{pmatrix},$$

tem suas componentes linearmente independentes, já que as estatísticas g_1 e g_2 como funções das variáveis aleatórias y_i 's são independentes. Logo, $U(\underline{\theta})$ é um vetor aleatório que admite densidade conjunta. através das densidades das estatísticas g_1 e g_2 .

Podemos observar que a função de log-verossimilhança obtida a partir de (2.2.1) admite derivadas de todas as ordens em relação à variável y e ao vetor de parâmetros $\underline{\theta} = (p, q)$. Pelo fato de ser $f(y; p, q) = f(y; \underline{\theta})$ uma densidade beta em $(0, 1)$, podemos mostrar que $\partial f(y, \underline{\theta})/\partial \underline{\theta}$ é contínua em Θ , $\int f(y, \underline{\theta}) dy$ existe e $\int |\partial f(y, \underline{\theta})/\partial \underline{\theta}| dy < M < \infty$ para todo $\underline{\theta}$ pertencente a uma vizinhança do valor verdadeiro $\underline{\theta}_0$. Assim, são válidas as condições do Teorema (1.3.2.) dado em Amemiya (1985, p.17). Desta forma, para a função de verossimilhança, podemos inverter a ordem das operações de diferenciação em relação a $\underline{\theta}$ e integração com respeito a y . Assim,

$$\begin{aligned} E_{\underline{\theta}}(U(\underline{\theta})) &= E_{\underline{\theta}}\left(\frac{\partial \log(L(\underline{\theta}))}{\partial \underline{\theta}}\right) = \int \frac{1}{L(\underline{\theta})} \frac{\partial L(\underline{\theta})}{\partial \underline{\theta}} L(\underline{\theta}) dy \\ &= \int \frac{\partial L(\underline{\theta})}{\partial \underline{\theta}} dy = \frac{\partial}{\partial \underline{\theta}} \int L(\underline{\theta}) dy = 0. \end{aligned}$$

A matriz $K(\underline{\theta})$ existe pela condição de regularidade (v) da Seção 1.4. e podemos mostrar que $K(\underline{\theta})$ é positiva-definida. Para verificar esta última condição, observamos que $K(\underline{\theta}) = \text{Cov}_{\underline{\theta}}(U(\underline{\theta})) \geq 0$ e suponhamos que $K(\underline{\theta})$ não é estritamente positiva, i.e., existe um vetor $\underline{a} = (a_1, a_2)^\top$ tal que $\underline{a}^\top K(\underline{\theta}) \underline{a} = 0$, que implica $\text{Var}(\underline{a}^\top U(\underline{\theta})) = 0$. Assim, $\underline{a}^\top U(\underline{\theta}) = k$ com k constante, o que significa que as componentes do vetor $U(\underline{\theta})$ e a função constante são linearmente dependentes, mas isto é uma contradição dada a independência linear das estatísticas g_1 e g_2 que formam as componentes de $U(\underline{\theta})$. Assim, é verificada a condição de regularidade (iv) da Seção 1.4.

Os estimadores de máxima verossimilhança $\hat{p}$ e $\hat{q}$ de p e q são obtidos pela maximização de (2.2.7), como soluções do sistema não-linear

$$\begin{aligned} \psi(p) - \psi(p+q) &= \log g_1, \\ \psi(q) - \psi(p+q) &= \log g_2. \end{aligned} \tag{2.2.8}.$$

Podemos mostrar que a densidade (2.2.1) pertence à família exponencial canônica, i.e.,

$$\begin{aligned} f(y; \underline{\theta}) &= \exp\{p \log(y) + q \log(1 - y) - \log(y) - \log(1 - y) + A(p, q)\} \\ &= h(y) \exp\{\underline{\theta}^\top T(y) - B(\underline{\theta})\}, \end{aligned}$$

sendo $h(y) = \{-\log(y) - \log(1 - y)\}$, $B(\underline{\theta}) = A(p, q)$ e $T(y) = (\log(y), \log(1 - y))^\top$. Então, ao considerar a amostra $\underline{y}$, o vetor

$$T_n(\underline{y}) = \left(\sum_{i=1}^n \log(y_i), \sum_{i=1}^n \log(1 - y_i) \right) = (n \log(g_1), n \log(g_2))$$

admite densidade conjunta em $(-\infty, 0) \times (-\infty, 0)$. Assim, $c_1 \log(g_1) + c_2 \log(g_2)$ admite densidade em $(-\infty, 0)$, $(0, \infty)$ ou $(-\infty, \infty)$ (dependendo dos sinais de c_1 e c_2), $\forall (c_1, c_2) \neq (0, 0)$. Desta forma, pelo Teorema (2.3.1) dado em Bickel & Doksum (2001, p.122) se garante que a solução de (2.2.8) é única; então, além das propriedades mencionadas anteriormente, garante-se que nosso modelo beta é um modelo regular. Sob as condições de regularidade da Seção 1.4, o valor esperado da função de escore é igual a zero e

$$\begin{aligned} E[\log g_1] &= \psi(p) - \psi(p + q), \\ E[\log g_2] &= \psi(q) - \psi(p + q). \end{aligned}$$

Através de (2.2.8) observamos que as estimativas de p e q não podem ser obtidas de forma analítica fechada, então é preciso que sejam calculadas numericamente através de um algoritmo iterativo.

2.3. O modelo de regressão beta

Em muitos casos temos conhecimento limitado sobre a relação entre variáveis envolvidas em um problema de interesse. Se consideramos os valores observados de tais variáveis como os resultados de um experimento, devemos ter, então, uma ferramenta teórica, um modelo matemático, através do qual estas variáveis estejam relacionadas, para atuar como base do processo gerador de dados. O objetivo desta seção é definir um modelo de regressão para modelar variáveis aleatórias que são regidas por uma distribuição de probabilidades beta, caracterizada pelos parâmetros p e q . Para obter uma estrutura de regressão que permita modelar a média de uma variável resposta sujeita a um parâmetro de precisão, trabalhamos com uma parametrização diferente daquela da densidade (2.2.1). Seja $\mu = p/(p + q)$ e $\phi = p + q$, i.e., $p = \mu\phi$ e $q = (1 - \mu)\phi$. De (2.2.2) e (2.2.3) se deduz que, se $Y \sim \mathcal{B}(p, q)$, então

$$E(Y) = \mu, \tag{2.3.1}$$

$$\text{Var}(Y) = \frac{V(\mu)}{1 + \phi}, \tag{2.3.2}$$

onde μ é a média da variável resposta, $V(\mu) = \mu(1 - \mu)$ denota a “função de variância”, e ϕ pode ser interpretado como um parâmetro de precisão, no sentido de que, para μ fixo, se o valor de ϕ aumenta, a variância da variável resposta diminui. Pode-se observar isto na

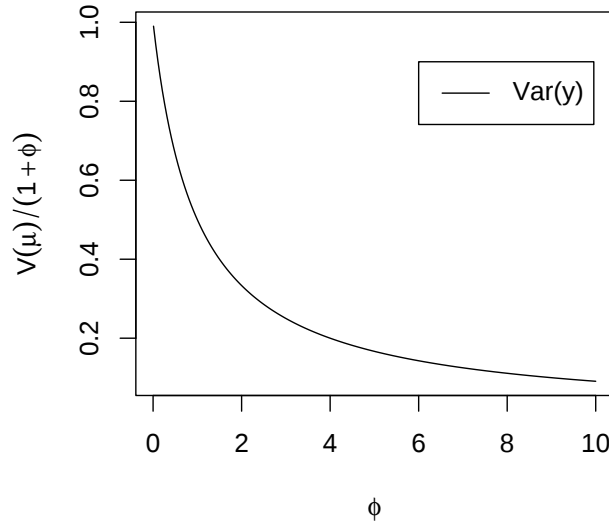


Figura 2.2.4: Comportamento da variância para μ fixo.

Figura 2.2.4. Com a mudança de parâmetros, a densidade beta da variável aleatória Y dada em (2.2.1) pode ser escrita como

$$f(y; \mu, \phi) = \begin{cases} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, & y \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases} \quad (2.3.3)$$

e sua função de log-densidade é dada por

$$\begin{aligned} \log f(y) = & \log \Gamma(\phi) - \log \Gamma(\mu\phi) - \log \Gamma((1-\mu)\phi) + (\mu\phi - 1) \log y \\ & + \{(1-\mu)\phi - 1\} \log(1-y), \end{aligned} \quad (2.3.4)$$

com $0 < \mu < 1$ e $\phi > 0$, já que para $p, q > 0$ e $0 < p < p + q$, o que implica que $0 < p/(p + q) < 1$ e $p + q > 0$. A Figura 2.2.5 apresenta diferentes densidades beta sob a nova parametrização, correspondentes a alguns valores dos parâmetros (μ, ϕ) . Podemos observar que sob as diferentes escolhas dos parâmetros a forma da densidade muda. Em particular, quando $\mu = 1/2$ as densidades se apresentam com forma simétrica; no entanto quando $\mu \neq 1/2$ as formas são não simétricas. Além disso, observamos que, para μ fixo, a dispersão da distribuição diminui quando ϕ aumenta, como era de se esperar a partir da relação (2.3.2). Podemos notar, no painel superior da Figura 2.2.5 a presença de duas densidades com forma de ‘J’ e duas outras com forma de ‘J’ invertido. Nós percebemos que no caso de $\mu = 1/2$ e $\phi = 2$ a densidade se reduz a uma distribuição uniforme padrão. De forma análoga, a densidade beta permite modelar densidades de forma de ‘U’.

Neste trabalho assumimos que a variável resposta está restrita ao intervalo $(0, 1)$. No caso em que a resposta assume valores no intervalo arbitrário (a, b) , com a, b conhecidos, podemos modelar a variável $(Y - a)/(b - a)$, a qual é uma padronização que permite trabalhar com a densidade (2.3.3) no intervalo $(0, 1)$. Seja $\underline{y} = (y_1, \dots, y_n)$ onde os y_i 's são v.a.'s independentes. Suponha que $y_t, t = 1, \dots, n$, segue uma densidade da forma (2.3.3) onde a média μ_t e o parâmetro de precisão ϕ são desconhecidos. Ferrari & Cribari-Neto (2003) propõem o seguinte modelo com estrutura de regressão beta. Suponhamos que a média de y_t satisfaz uma relação funcional da forma

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = \eta_t, \quad (2.3.5)$$

onde $\beta = (\beta_1, \dots, \beta_k)^\top$ é um vetor de parâmetros de regressão desconhecidos ($\beta \in \mathbb{R}^k$), η_t é um preditor linear e $x_{t1}, \dots, x_{tk}$ são observações de k covariáveis ($k < n$) conhecidas. Além disso, assumimos que a função de ligação $g : (0, 1) \rightarrow \mathbb{R}$ seja estritamente monótona e duas vezes diferenciável. Podemos escolher para o modelo de regressão beta diferentes funções de ligação, por exemplo, podemos usar a especificação logit $g(\mu) = \log\{\mu/(1-\mu)\}$, a função probit $g(\mu) = \Phi^{-1}(\mu)$, onde $\Phi(\cdot)$ representa a função de distribuição acumulada de uma variável aleatória normal padrão, a função de ligação log-log complementar $g(\mu) = \log\{-\log(1-\mu)\}$, a ligação $g(\mu) = -\log\{-\log(\mu)\}$, entre outras. Uma discussão detalhada destas funções de ligação pode ser encontrada em McCullagh & Nelder (1989, §4.3.1). A utilização de outras transformações como funções de ligação encontra-se descrita em Atkinson (1985, cap. 7).

Dada a estrutura do modelo de regressão beta estaríamos tentados a dizer que ele é um modelo linear generalizado. Como pode-se ver em McCullagh & Nelder (1989), a *componente aleatória* de um modelo linear generalizado considera que a resposta y_t tem densidade na família exponencial

$$\pi(y; \theta_t, \phi) = \exp\{\phi[y\theta_t - b(\theta_t) + c(y, \phi)]\},$$

onde $b(\cdot)$ e $c(\cdot, \cdot)$ são funções conhecidas para cada observação, $\phi > 0$ é um parâmetro conhecido de escala e θ_t é o *parâmetro natural* (ou canônico) que caracteriza a distribuição $\pi(y; \theta_t, \phi)$. Pode-se mostrar, neste caso, que

$$E(y_t) = \mu_t = \frac{db(\theta_t)}{d\theta_t},$$

$$\text{Var}(y_t) = \frac{d^2b(\theta_t)}{\phi d\theta_t^2},$$

de tal forma que $\text{Var}(y_t) = \phi^{-1}V_t$, onde $V_t = \frac{d\mu_t}{d\theta_t}$ é a *função de variância*, que depende unicamente da média e, portanto, pode ser expressa, em geral, por uma relação funcional unívoca da média a partir de $\theta = \int V^{-1}d\mu = q(\mu)$. Assim, especificando o parâmetro $\mu = q^{-1}(\theta)$, a distribuição $\pi(y; \theta_t, \phi)$ é univocamente determinada (Patil & Shorrock, 1965), e a relação funcional variância-média caracteriza a distribuição na família exponencial $\pi(y; \theta_t, \phi)$.

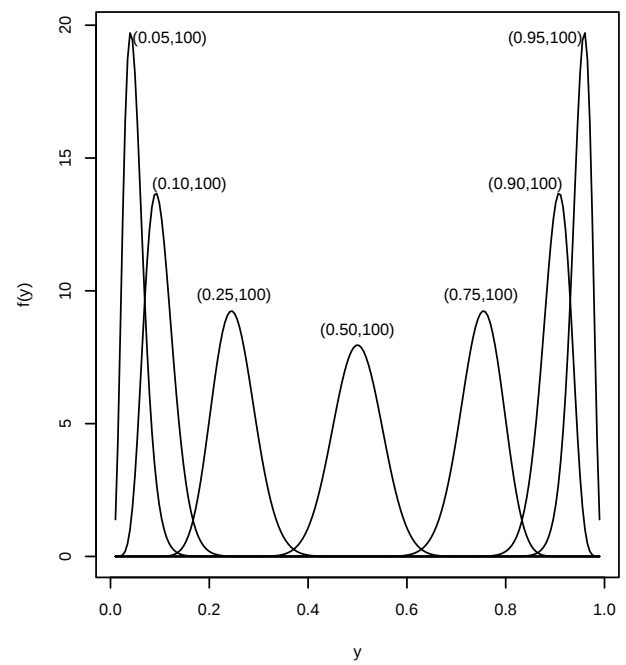
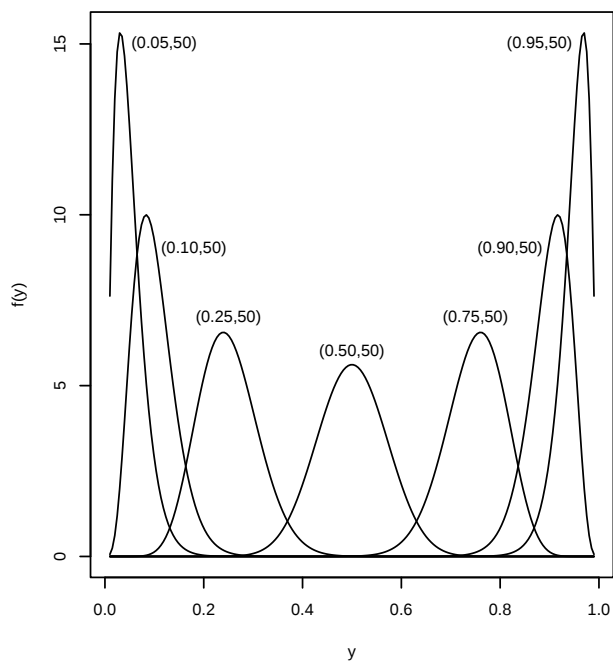
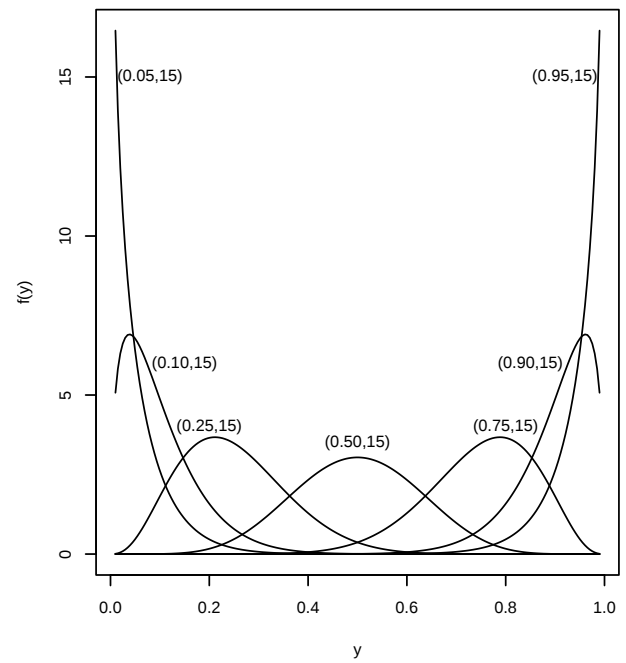
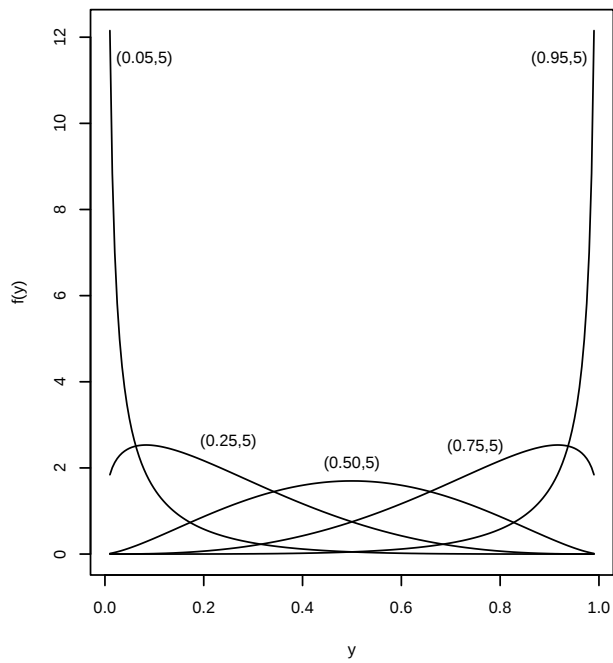


Figura 2.2.5: Densidades beta para diferentes valores de (μ, ϕ) .

Contudo, esta relação não caracteriza a distribuição beta (2.3.3) na família exponencial $\pi(y; \theta_t, \phi)$, já que

$$\begin{aligned} f(y; \mu, \phi) &= \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \\ &= \exp\{\phi[t(y)\theta - b(\theta) + c(y, \theta)]\}, \end{aligned}$$

onde $t(y) = \log(y/(1-y))$, $\theta = \mu$, $b(\theta) = \phi^{-1} \log(\Gamma(\mu\phi)\Gamma((1-\mu)\phi)\Gamma(\phi)^{-1})$ e $c(y, \theta) = \phi^{-1}\{(\log(y) + (\phi-1)\log(1-y))\}$. Além disso, se ϕ for conhecido obtemos de (2.3.2) que a função de variância corresponde à de um modelo binomial contradizendo assim, a relação únivoca entre a variância e a distribuição $\pi(y; \theta, \phi)$.

Uma interessante aplicação do modelo de regressão beta é obtida quando a função de ligação é logit; neste caso, a média μ_t de y_t pode ser escrita como

$$\mu_t = \frac{\exp(x_t^\top \beta)}{1 + \exp(x_t^\top \beta)},$$

onde $x_t^\top = (x_{t1}, \dots, x_{tk})$, $t = 1, \dots, n$. Neste contexto, os parâmetros de regressão têm uma importante interpretação. Suponhamos que ao valor do i -ésimo regressor são acrescidas δ unidades, mantendo fixos os valores das demais variáveis. Seja $\mu^\dagger$ a média da variável y sob o novo valor da covariável e seja μ a média da variável y sob os valores das covariáveis originais. Um simples cálculo mostra

$$\exp(\delta\beta_i) = \frac{\mu^\dagger/(1-\mu^\dagger)}{\mu/(1-\mu)},$$

i.e., $\exp(\delta\beta_i)$ é a razão de chances.

Com isto em mente, consideremos o conjunto de dados apresentados em Prater (1956). A variável dependente deste estudo é a proporção de petróleo cru que é transformado em gasolina após o processo de destilação e fracionamento. As potenciais covariáveis que tentam modelar tal proporção são: gravidade do petróleo cru (graus API), pressão do vapor de petróleo cru (lbf/in²), petróleo cru a 10% do ponto ASTM (i.e., a temperatura em que 10% do petróleo cru é vaporizado) e temperatura (graus F) em que toda a gasolina se evapora. O conjunto de dados contém 32 observações da variável resposta e das variáveis independentes. Definimos a chance de converter petróleo cru em gasolina como o número de unidades de petróleo cru, menos de 10 unidades, que em média são convertidas em gasolina dividido pelo número de unidades que não são convertidas. Assim, como ilustração, se em média 20% do petróleo cru é transformado em gasolina, então a razão de conversão é igual a 2/8. Suponha que a temperatura em que a gasolina é vaporizada aumenta em 50° F, i.e., $\delta = 50$. Daí, 50 vezes o parâmetro associado a esta covariável pode ser interpretado como o logaritmo da razão de chances de converter petróleo cru em gasolina sob a nova condição relativa à condição antiga, quando todas as outras variáveis permanecem constantes.

Continuando com nossa análise, a função de log-verossimilhança para o modelo de regressão beta proposto neste trabalho é da forma

$$\ell(\beta, \phi) = \sum_{t=1}^n \ell_t(\mu_t, \phi) \quad (2.3.6)$$

onde

$$\begin{aligned} \ell_t(\mu_t, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t \\ &\quad + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t), \end{aligned}$$

e μ_t , definida através de (2.3.5) como $\mu_t = g^{-1}(\eta_t)$, é uma função de β . A função de escore é obtida pela diferenciação da função de log-verossimilhança (2.3.6) em relação aos parâmetros desconhecidos. De acordo com o que foi discutido na Seção (2.2) podemos assumir que o modelo de regressão beta como função de (2.2.1) é um modelo regular, e seus estimadores de máxima verossimilhança são únicos. Como é apresentado no Apêndice A.1, para $r = 1, \dots, k$,

$$\frac{\partial \ell(\beta, \phi)}{\partial \beta_r} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} = \sum_{t=1}^n \phi[y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} x_{tr} \quad (2.3.7)$$

e

$$\frac{\partial \ell(\beta, \phi)}{\partial \phi} = \sum_{t=1}^n \{\mu_t[y_t^* - \mu_t^*] + \log(1 - y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)\}, \quad (2.3.8)$$

onde

$$\begin{aligned} y_t^* &= \log\left(\frac{y_t}{1 - y_t}\right), \\ \mu_t^* &= \psi(\mu_t \phi) - \psi((1 - \mu_t)\phi). \end{aligned}$$

Denotamos por $\text{diag}(a_t)$ a matriz diagonal de dimensão $n \times n$ com elemento típico a_t , $t = 1, \dots, n$. O sistema de equações em μ_t e ϕ ; definido a partir de (2.3.7) e (2.3.8), onde igualando as derivadas a zero, pode ser expresso matricialmente. Seja X a matriz formada das cováriaveis do modelo, com dimensão $n \times k$ e t -ésimo vetor linha $x_t^\top$. Definimos ainda os vetores $y^* = (y_1^*, \dots, y_n^*)$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)$, e as matrizes $\tilde{\mu} = \text{diag}(\mu_t)$ e $T = \text{diag}\left(\frac{d\mu_t}{d\eta_t}\right)$. De forma análoga definimos $h^\top = (\log(1 - y_1) + \psi(\phi) - \psi((1 - \mu_1)\phi), \dots, \log(1 - y_n) + \psi(\phi) - \psi((1 - \mu_n)\phi))$ e seja $\mathbb{1} = (1, \dots, 1)^\top$ o vetor de uns de dimensão $n \times 1$. Consideramos o vetor de parâmetros conjuntos $\underline{\theta} = (\beta^\top, \phi)^\top$, de tal forma que o vetor escore de dimensão $(k + 1) \times 1$ é da forma $U(\underline{\theta}) = (U_\beta(\beta, \phi), U_\phi(\beta, \phi))^\top$, onde

$$U_\beta(\beta, \phi) = \phi X^\top T(y^* - \mu^*), \quad (2.3.9)$$

$$U_\phi(\beta, \phi) = (\{(y^* - \mu^*)^\top \text{diag}(\mu_t)\} - h^\top) \mathbb{1}. \quad (2.3.10)$$

Os estimadores de máxima verossimilhança de β e ϕ são obtidos pela solução do sistema $U(\underline{\theta}) = 0$, i.e.,

$$U_\beta(\beta, \phi) = 0,$$

$$U_\phi(\beta, \phi) = 0,$$

e observamos que tais estimadores não possuem forma fechada. Assim, eles têm que ser obtidos numericamente pela maximização da função de log-verossimilhança usando um algoritmo de otimização não-linear, tal como o algoritmo de Newton (Newton-Raphson, Fisher's scoring, BHHH, etc.) ou um algoritmo quasi-Newton (BFGS). Neste trabalho utilizamos o algoritmo BFGS, o qual foi discutido na Seção 1.7.1. Em geral, estes algoritmos precisam da especificação de um valor inicial para poderem usar o esquema iterativo. Ferrari & Cribari-Neto (2003) sugerem usar como vetor inicial a estimativa do vetor β obtida a partir de uma regressão linear da variável resposta transformada $g(y_1), \dots, g(y_n)$ em X , i.e., $(X^\top X)^{-1} X^\top z$, onde o vetor z é da forma $z = (g(y_1), \dots, g(y_n))^\top$. Além disso, a partir de (2.3.2) obtém-se que $\phi = \{\mu_t(1 - \mu_t)/\text{Var}(y_t)\} - 1$, e, ao expandir até primeira ordem a função $g(y_t)$ em série de Taylor ao redor do ponto μ_t e tomando a variância, deduz-se que

$$\text{Var}(g(y_t)) \approx \text{Var}\{g(\mu_t) + (y_t - \mu_t)g'(\mu_t)\} = \text{Var}(y_t)\{g'(\mu_t)\}^2,$$

i.e.,

$$\text{Var}(y_t) \approx \text{Var}(g(y_t))/\{g'(\mu_t)\}^2.$$

Assim, a sugestão de uma estimativa inicial para ϕ pode ser dada por

$$\frac{1}{n} \sum_{t=1}^n \frac{\check{\mu}_t(1 - \check{\mu}_t)}{\check{\sigma}_t^2} - 1,$$

onde $\check{\mu}_t$ é obtido aplicando $g^{-1}(\cdot)$ ao t -ésimo valor ajustado da regressão linear de z em X , i.e., $\check{\mu}_t = g^{-1}(x_t^\top (X^\top X)^{-1} X^\top z)$, e $\check{\sigma}_t^2 = \check{e}^\top \check{e} / [(n - k)\{g'(\check{\mu}_t)\}^2]$, sendo $\check{e} = z - (X^\top X)^{-1} X^\top z$ o vetor de resíduos de mínimos quadrados da regressão linear sob a variável resposta transformada.

Para determinar o comportamento da variabilidade dos estimadores dos parâmetros do modelo de regressão beta, podemos obter uma fórmula para a matriz de informação de Fisher $K(\theta)$. Como se pode observar no Apêndice A.1, os momentos de segunda ordem são

$$\begin{aligned} \kappa_{rs} &= -\phi^2 \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr}, \\ \kappa_{r\phi} &= -\sum_{t=1}^n c_t \frac{d\mu_t}{d\eta_t} x_{tr}, \\ \kappa_{\phi\phi} &= -\sum_{t=1}^n d_t, \end{aligned}$$

onde $w_t = [\psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi)]$, o vetor $c = (c_1, \dots, c_n)^\top$ é de dimensão $n \times 1$ com $c_t = \phi[\mu_t w_t - \psi'((1 - \mu_t)\phi)]$, e, analogamente, $d_t = (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t\phi) - \psi'(\phi)$.

Definimos ainda

$$\begin{aligned} W_{\beta\beta} &= \text{diag}\left(\left(\phi \frac{d\mu_t}{d\eta_t}\right)^2 w_t\right), \\ W_{\beta\phi} &= Tc, \\ W_{\phi\beta} &= W_{\beta\phi}^\top, \\ W_{\phi\phi} &= \text{tr}(\text{diag}(d_t)), \end{aligned}$$

onde $\text{tr}(\cdot)$ é o operador traço. Definimos a matriz $\widetilde{W}$ de dimensão $(n+1) \times (n+1)$ por

$$\widetilde{W} = \begin{pmatrix} W_{\beta\beta} & W_{\beta\phi} \\ W_{\phi\beta} & W_{\phi\phi} \end{pmatrix}.$$

Para facilitar nossos cálculos definimos $\widetilde{X}$ como a matriz aumentada de dimensão $(n+1) \times (k+1)$ da forma

$$\widetilde{X} = \begin{pmatrix} X & 0 \\ 0 & 1 \end{pmatrix}, \quad (2.3.11)$$

sendo X a matriz de variáveis explicativas de dimensão $n \times k$.

Sob a notação anterior temos que a matriz de informação de Fisher é dada por

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \quad (2.3.12)$$

onde $K_{\beta\beta} = X^\top W_{\beta\beta} X$, $K_{\beta\phi} = X^\top W_{\beta\phi} = K_{\phi\beta}^\top$ e $K_{\phi\phi} = W_{\phi\phi}$. Se consideramos o vetor de parâmetros conjuntos $\underline{\theta} = (\beta^\top, \phi)^\top$, a matriz de informação de Fisher pode ser escrita como

$$K(\underline{\theta}) = K(\beta, \phi) = \widetilde{X}^\top \widetilde{W} \widetilde{X}. \quad (2.3.13)$$

Notamos que $K_{\beta\phi} = K_{\phi\beta}^\top \neq 0$, o que indica que os parâmetros β e ϕ não são ortogonais, em contraste ao que se verifica na classe de modelos lineares generalizados (McCullagh & Nelder, 1989).

Podemos obter aproximações das quantidades w_t , c_t , d_t , $t = 1, \dots, n$ que permitem o cálculo da matriz de informação de Fisher $K(\underline{\theta})$, quando $\mu_t\phi$ e $(1-\mu_t)\phi$ são grandes. Pode-se mostrar (Abramowitz & Stegun, p.259), que para $z \rightarrow \infty$,

$$\begin{aligned} \psi(z) &= \log(z) - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} + \dots, \\ \psi'(z) &= \frac{1}{z} + \frac{1}{2z^2} + \frac{1}{6z^3} - \frac{1}{30z^5} + \dots \end{aligned}$$

Desta forma, para $t = 1, \dots, n$ se $\mu_t\phi$ e $(1-\mu_t)\phi$ são suficientemente grandes, i.e., se ϕ é grande, já que $0 < \mu_t < 1$, então, ao utilizar as expansões anteriores para a função digama e sua derivada, temos as aproximações

$$\begin{aligned} w_t &\approx \frac{1}{\phi\mu_t(1-\mu_t)} = \frac{1}{\phi V(\mu_t)}, \\ c_t &\approx \phi \left[\mu_t \frac{1}{\phi\mu_t(1-\mu_t)} - \frac{1}{\phi(1-\mu_t)} \right] = 0, \\ d_t &\approx \frac{(1-\mu)}{\phi} + \frac{\mu}{\phi} - \frac{1}{\phi} = 0. \end{aligned}$$

Logo, $K_{\beta\beta} = X^\top W_{\beta\beta} X$ com $W_{\beta\beta} \approx \text{diag}\left(\left(\frac{d\mu_t}{d\eta_t}\right)^2 \frac{\phi}{V(\mu_t)}\right)$, $K_{\beta\phi} = K_{\phi\beta}^\top \approx \underline{0}$, e $K_{\phi\phi} \approx 0$, onde $\underline{0}$ é um vetor de zeros de dimensão $k \times 1$. Desta forma, à medida que ϕ aumenta os blocos $K_{\beta\phi}$ e $K_{\phi\beta}$ tendem a zero, assim como $K_{\phi\phi}$. Portanto, quando $\phi \rightarrow \infty$, a derivada de (2.3.10) com respeito a ϕ vá para zero e a curvatura tende para zero, i.e., a função de log-verossimilhança é menos curva, o que faz com que ϕ seja difícil de identificar. Assim a informação sobre ϕ é pequena, sendo desta forma estimável unicamente o bloco $K_{\beta\beta} = \phi X^\top W_{\beta\beta} X$, que para ϕ suficientemente grande toma a forma da matriz de informação de Fisher de um modelo linear generalizado, (McCullagh & Nelder 1989).

Usando a expressão padrão para a inversa de matrizes particionadas (ver, por exemplo, Rao, 1973, p. 33), deduz-se que a inversa da matriz de informação de Fisher (2.3.12) é

$$K(\underline{\theta})^{-1} = K(\beta, \phi)^{-1} = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix} \quad (2.3.14)$$

com

$$\begin{aligned} K^{\beta\beta} &= (X^\top W_{\beta\beta} X)^{-1} \left\{ I_k + \frac{X^\top T c c^\top T^\top X (X^\top W_{\beta\beta} X)^{-1}}{\gamma} \right\} \\ &= K_{\beta\beta}^{-1} \left\{ I_k + \frac{K_{\beta\phi} K_{\phi\beta} K_{\beta\beta}^{-1}}{\gamma} \right\}, \end{aligned}$$

com $\gamma = \text{tr}(\text{diag}(d_t)) - c^\top T^\top X (X^\top W_{\beta\beta} X)^{-1} X^\top T c = \text{tr}(\text{diag}(d_t)) K_{\phi\beta} K_{\beta\beta}^{-1} K_{\beta\phi}$. Além disso,

$$K^{\beta\phi} = (K^{\phi\beta})^\top = -\frac{1}{\gamma} (X^\top W_{\beta\beta} X)^{-1} X^\top T c = -\frac{1}{\gamma} K_{\beta\beta}^{-1} K_{\beta\phi}$$

e $K^{\phi\phi} = \frac{1}{\gamma}$. Aqui I_k representa a matriz identidade de dimensão $k \times k$. Desta forma, como se discutiu anteriormente, sob as condições de regularidade da Seção 1.4 se garante a normalidade assintótica do estimador de máxima verossimilhança:

$$\widehat{\underline{\theta}} = \begin{pmatrix} \widehat{\beta} \\ \widehat{\phi} \end{pmatrix} \overset{A}{\sim} \mathcal{N}_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K(\underline{\theta})^{-1} \right),$$

sendo $\widehat{\beta}$ e $\widehat{\phi}$ os estimadores de máxima verossimilhança de β e ϕ , respectivamente, e $\mathcal{N}_{k+1}$ uma distribuição normal $(k+1)$ -variada.

Além disso, pela normalidade assintótica e dada a consistência do estimador de máxima verossimilhança $\widehat{\underline{\theta}}$, podemos construir intervalos de confiança assintóticos da forma (1.6.2) para os parâmetros do modelo de regressão beta. Assim, para $r = 1, \dots, k$,

$$\left(\widehat{\beta}_r - z_{1-\frac{\alpha}{2}} (K(\widehat{\underline{\theta}})^{rr})^{1/2}, \widehat{\beta}_r + z_{1-\frac{\alpha}{2}} (K(\widehat{\underline{\theta}})^{rr})^{1/2} \right) \quad (2.3.15)$$

e

$$\left(\widehat{\phi} - z_{1-\frac{\alpha}{2}} (K(\widehat{\underline{\theta}})^{\phi\phi})^{1/2}, \widehat{\phi} + z_{1-\frac{\alpha}{2}} (K(\widehat{\underline{\theta}})^{\phi\phi})^{1/2} \right) \quad (2.3.16)$$

são intervalos de confiança assintóticos (ICA) para β_r e ϕ de tamanho $100(1 - \alpha)\%$, respectivamente. As variâncias assintóticas de $\hat{\beta}_r$ e $\hat{\phi}$ são $K(\hat{\theta})^{rr}$ e $K(\hat{\theta})^{\phi\phi}$, respectivamente, sendo $K(\hat{\theta})^{rr}$ o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em $\hat{\theta}$. Adicionalmente, se em nosso modelo de regressão beta a função de ligação é logit, então ao usar o método delta⁴ de forma adequada, podemos construir um intervalo de confiança assintótico de tamanho $100(1 - \alpha)\%$ para a razão de chances $e^{c\beta_r}$ da forma

$$\left(\exp \left\{ c \left(\hat{\beta}_r - z_{1-\frac{\alpha}{2}} (K(\hat{\theta})^{rr})^{1/2} \right) \right\}, \exp \left\{ c \left(\hat{\beta}_r + z_{1-\frac{\alpha}{2}} (K(\hat{\theta})^{rr})^{1/2} \right) \right\} \right). \quad (2.3.17)$$

Finalmente, um intervalo de confiança assintótico de tamanho $100(1 - \alpha)\%$ para a resposta média μ , dado um vetor de covariáveis x , pode ser calculado a partir de (1.6.2) e do método delta, sendo dado por

$$\left(g^{-1} \left(\hat{\eta} - z_{1-\frac{\alpha}{2}} \text{e.p.}(\hat{\eta}) \right), g^{-1} \left(\hat{\eta} + z_{1-\frac{\alpha}{2}} \text{e.p.}(\hat{\eta}) \right) \right), \quad (2.3.18)$$

onde g^{-1} é a inversa da função de ligação e $\hat{\eta} = x^\top \hat{\beta}$ o valor do erro padrão $\text{e.p.}(\hat{\eta}) = \sqrt{(x^\top \hat{K}^{\beta\beta} x)}$, sendo $\hat{K}^{\beta\beta} = K^{\beta\beta}(\hat{\theta})$ a inversa da matriz de informação de Fisher de β avaliada em $\hat{\theta}$. Temos que ressaltar que este intervalo somente é válido se a função de ligação é estritamente crescente, pois, caso contrário não se garante manter a relação de ordem entre os extremos do intervalo de confiança.

Uma pergunta natural que surge ao estudar o modelo de regressão beta e suas estimativas de máxima verossimilhança é se estas são adequadas estatisticamente. Ferrari & Cribari-Neto (2003) descrevem algumas técnicas que são úteis para realizar inferências sobre os parâmetros da regressão. Entre os procedimentos que eles consideram está o teste da razão de verossimilhanças. Suponhamos que o objetivo é testar a hipótese nula $\mathcal{H}_0 : \beta_1 = \beta_1^{(0)}$ contra a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq \beta_1^{(0)}$, onde $\beta_1 = (\beta_1, \dots, \beta_m)^\top$ e $\beta_1^{(0)} = (\beta_1^{(0)}, \dots, \beta_m^{(0)})^\top$, para $m < k$ e $\beta_1^{(0)}$ dado. A estatística da razão de verossimilhanças é

$$w_1 = 2 \left\{ \ell(\hat{\beta}, \hat{\phi}) - \ell(\tilde{\beta}, \tilde{\phi}) \right\},$$

onde $\ell(\hat{\beta}, \hat{\phi})$ é a função de log-verossimilhança e $(\tilde{\beta}^\top, \tilde{\phi}^\top)^\top$ é o estimador restrito de máxima verossimilhança de $(\beta^\top, \phi)^\top$ obtido impondo-se a hipótese nula. Sob as condições de regularidade da Seção 1.4 e sob $\mathcal{H}_0$, temos que

$$w_1 \overset{A}{\sim} \chi_m^2;$$

desta forma, pode-se testar a hipótese nula de interesse usando como aproximação os valores críticos da distribuição assintótica χ_m^2 .

⁴ O método delta afirma que se T_n é um estimador consistente de θ e $\sqrt{n}[T_n - \theta] \overset{A}{\sim} \mathcal{N}(0, \tau^2)$, então $\sqrt{n}[g(T_n) - g(\theta)] \overset{A}{\sim} \mathcal{N}(0, (g'(\theta))^2 \tau^2)$, assumindo que $g'(\theta)$ existe e não é zero. Para maiores detalhes, ver Lehmann & Casella (1998).

Capítulo 3

Correção de viés

3.1 Revisão sobre correção de viés

Os estimadores de máxima verossimilhança (EMV) em modelos estatísticos são em geral viesados para os valores verdadeiros dos parâmetros. Usualmente o viés é ignorado na prática sob a suposição de que este é pequeno ao ser comparado com o erro padrão do estimador do parâmetro. De fato, pode mostrar-se que numa amostra de tamanho n o viés em geral é de ordem $O(n^{-1})$ e o desvio padrão assintótico é de ordem $O(n^{-1/2})$. Entretanto, para alguns modelos o viés pode ser consideravelmente maior que o correspondente desvio padrão em amostras de tamanho não “muito grande”. Portanto, é útil obtermos expressões para calcular os vieses de segunda ordem, as quais permitem determinar a qualidade das estimativas, sobretudo para amostras de tamanho pequeno ou moderado. Historicamente, o artigo pioneiro sobre o cálculo do viés surgiu com um trabalho de Bartlett (1953) sobre intervalos de confiança aproximados baseados nos EMV no caso uniparamétrico, fornecendo uma expressão simples para o viés de ordem $O(n^{-1})$. Haldane (1953) e Haldane & Smith (1956) discutiram expansões assintóticas para os EMV em amostras aleatórias de um ou dois parâmetros desconhecidos, encontrando expressões de ordem $O(n^{-1})$ para os primeiros quatro cumulantes. Shenton & Wallington (1962) desenvolveram uma metodologia para obter o viés de ordem $O(n^{-1})$ dos EMV no contexto biparamétrico. Mais tarde, Shenton & Bowman (1963) desenvolveram os quatro primeiros momentos amostrais dos EMV até ordem $O(n^{-4})$. Fórmulas para o viés dos EMV até ordem $O(n^{-2})$ e covariâncias de mesma ordem, ainda para o caso multiparamétrico, foram obtidas por Bowman & Shenton (1965) e Shenton & Bowman (1977).

Posteriormente, Cox & Snell (1968) apresentaram uma expressão geral para o viés de ordem $O(n^{-1})$ dos EMV nos casos uniparamétrico e multiparamétrico. Neste artigo, os autores trabalham com o modelo $Y = X\beta + \epsilon$, onde X é uma matriz conhecida de dados, β um vetor de parâmetros desconhecidos e ϵ é um vetor $n \times 1$ de variáveis aleatórias não observáveis independentes e distribuídas com média zero e variância constante. Definindo R tal que $Y = X\hat{\beta} + R$, eles examinaram as propriedades dos R 's e encontraram expressões para as esperanças e covariâncias dos R 's, através de uma expansão em série de Taylor de $R_i - \epsilon_i$ em termos de $\hat{\beta} - \beta$. As fórmulas encontradas, em particular a expressão do viés de segunda ordem de $\hat{\beta}$, são úteis, dado que a maioria das propriedades dos R 's são similares às dos ϵ 's quando o número de parâmetros é pequeno comparado com o tamanho de amostra. Um inconveniente na aplicação da metodologia de Cox & Snell em modelos multiparamétricos é o cálculo de inúmeros cumulantes de derivadas da função de log-verossimilhança. Assim, se p é

o número de parâmetros do modelo, torna-se necessário calcular $2^p - (p+1)$ cumulantes para encontrar os vieses dos EMV desses parâmetros. Um método geral para o cálculo de vieses de segunda ordem numa classe de problemas de mínimos quadrados não-lineares foi apresentado por Box (1971), que considerou, entre outros aspectos, o cálculo de viés por estimação bayesiana, embora os exemplos ilustrativos tenham sido para o caso de uma distribuição a priori uniforme, na qual o estimador é de máxima verossimilhança; seu enfoque é aplicado para determinar os vieses do estimador de máxima verossimilhança dos parâmetros da distribuição gama. Robertson & Fryer (1970) desenvolveram um método geral para encontrar os vieses e covariâncias de ordem $O(n^{-2})$ dos estimadores de máxima verossimilhança. Sowden (1972) comparou os métodos de estimação por máxima verossimilhança, χ^2 mínimo e χ^2 mínimo modificado, num modelo de resposta quantal “dose resposta”, em termos dos seus vieses de segunda ordem, mostrando que o método de máxima verossimilhança é um procedimento mais eficiente em relação aos outros métodos. Fryer & Robertson (1972) comparam os vieses dos estimadores de momentos, de máxima verossimilhança, multinomial e χ^2 mínimo até ordem $O(n^{-1})$ e erro quadrático médio de ordem $O(n^{-2})$ para uma série de distribuições misturadas.

A partir da fórmula desenvolvida por Cox & Snell (1968), Anderson & Richardson (1979), McLanchlan (1980) e Copas (1988) encontraram os vieses dos EMV em problemas de discriminação logística. Uma análise feita por Cook, Tsai & Wei (1986) mostra o cálculo dos vieses dos EMV em modelos de regressão não-lineares com erro normal, ou seja, com função de resposta $y_i = f(x_i, \beta) + \epsilon_i$, $i = 1, \dots, n$, onde f é uma função não-linear duas vezes diferenciável em β . Eles mostram que o viés tende a zero à medida que a diferença esperada δ entre as aproximações linear e quadrática de $f(\beta)$ se aproxima de zero, de tal forma que o modelo é essencialmente linear, ou se δ é ortogonal ao espaço coluna de V onde V é a matriz de elementos $f_i^r = \partial f_i / \partial \beta_r$, $i = 1, \dots, n$, $r = 1, \dots, p$ e o viés b é expresso na forma $b = (V^T V)^{-1} V^T \delta$. Além disso, eles mostram como o viés afeta os métodos de diagnóstico e se estabelece uma relação entre viés e curvatura. Young & Bakir (1987) apresentam estimadores corrigidos em modelos de regressão log-gama. Mais recentemente, Cordeiro & Klein (1994) desenvolvem fórmulas matriciais para corrigir os EMV em modelos ARMA. Isto foi possível graças às propriedades de invariância dos cumulantes em relação à permutação de parâmetros (McCullagh, 1987).

Paula (1992) generaliza trabalhos sobre cálculo de viés até ordem $O(n^{-1})$ e sua forma matricial, encontrados na literatura estatística. Cordeiro (1993) apresenta fórmulas gerais em notação matricial para o vício de ordem $O(n^{-1})$ dos EMV de dois modelos de regressão heteroscedásticos. Cordeiro & McCullagh (1991) e Cordeiro & Cribari-Neto (1993) desenvolvem expressões simples para o cálculo do viés com aplicações em modelos econométricos. Paula & Cordeiro (1995) derivam fórmulas para os vieses de ordem $O(n^{-1})$ dos estimadores dos parâmetros em modelos não-lineares da família exponencial. Ferrari, Botter, Cordeiro & Cribari-Neto (1996) derivam o vício de ordem $O(n^{-1})$ em modelos uniparamétricos e comparam o estimador corrigido com o estimador de máxima verossimilhança usual em

termos do erro quadrático médio. Cribari–Neto, Botter, Cordeiro & Ferrari (1998) obtêm expressões em forma fechada para o viés de segunda ordem dos EMV para um número de distribuições da família exponencial uniparamétrica, mostrando graficamente como cada vício e erro quadrático do EMV e de sua versão corrigida variam de acordo com o parâmetro que indexa a distribuição. Vasconcellos & Cordeiro (1997a) generalizam os trabalhos de Box (1971) e Cook et al. (1986) ao obterem fórmulas gerais em notação matricial para a correção de viés em modelos multivariados não-lineares heteroscedásticos. Aubin & Cordeiro (1997) e Cordeiro & Vasconcellos (1997) obtêm fórmulas matriciais para os vieses de ordem $O(n^{-1})$ dos parâmetros da matriz de covariância, num modelo de regressão linear com erros normais e matriz de covariância desconhecida. Vasconcellos & Cordeiro (1997b) deduzem uma fórmula geral para os vieses de segunda ordem dos EMV nos modelos SUR (“Seemingly Unrelated Regressions”), os quais apresentam correlação serial entre os erros nas diferentes equações de um conjunto de equações de regressão. As fórmulas obtidas mostram como o viés pode ser expresso como o vetor das estimativas de mínimos quadrados de uma regressão linear ponderada auxiliar.

Cordeiro, Vasconcellos & Santos (1998) derivam fórmulas para os vieses de segunda ordem dos EMV dos parâmetros de uma regressão não-linear com erros t -Student e dos parâmetros de dispersão e precisão. Cordeiro & Cribari–Neto (1998) estudam os vieses dos EMV em alguns modelos lineares generalizados e modelos não lineares de regressão da família exponencial, focalizando os estimadores dos parâmetros que compõem os preditores lineares em alguns modelos de regressão importantes. Botter & Cordeiro (1998) obtêm fórmulas para os vieses de segunda ordem dos EMV em modelos lineares generalizados com parâmetros de dispersão, e tais fórmulas podem ser vistas como generalizações dos trabalhos de Cordeiro & McCullagh (1991) e Cordeiro (1993). Posteriormente, Vasconcellos & Cordeiro (1999) obtêm expressões em forma matricial para os vieses de ordem $O(n^{-1})$ dos EMV dos parâmetros do vetor de locação em modelos de regressão multivariada com erros t -Student. Cordeiro & Vasconcellos (1999) discutem o melhoramento dos EMV em modelos de regressão von Mises, obtendo expressões matriciais gerais para os vieses de segunda ordem dos EMV dos parâmetros que determinam a locação da distribuição.

Além do método analítico para a correção de viés dos EMV, outros métodos são conhecidos; por exemplo, Firth (1993) propõe um método “preventivo” de redução do viés dos EMV por meio de uma modificação da função escore, mostrando que este método não depende da existência de restrições na maximização da função de log-verossimilhança. Mais recentemente, Ferrari & Cribari–Neto (1998) exploram a relação entre o método analítico, baseado em expansões de Edgeworth e a técnica de reamostragem bootstrap, sendo estes dois métodos equivalentes para a correção de viés dos EMV. Por outro lado, utilizam o método bootstrap para obter expressões gerais para correção de viés de terceira ordem dos EMV em modelos multiparamétricos. Recentes trabalhos sobre correção de viés do EMV foram feitos por Cordeiro, Rocha, Rocha & Cribari–Neto (1997) e Cribari–Neto & Vasconcellos (2002) para os dois parâmetros que indexam a distribuição beta. Entretanto, estes resultados não

são aplicados a modelos com estrutura de regressão, daí a idéia de Vasconcellos & Cribari-Neto (2003) de obter estimativas melhoradas dos EMV dos parâmetros da distribuição beta, onde tais parâmetros são modelados através de uma estrutura de regressão. Além disso, eles encontram expressões de forma fechada para o viés de segunda ordem.

3.2. Correção de Cox & Snell

Cox & Snell (1968) desenvolveram a técnica descrita a seguir para corrigir estimadores de máxima verossimilhança. Consideremos o modelo estatístico $f(\underline{y}, \underline{\theta})$, com $\underline{\theta} \in \mathbb{R}^k$, e seja $\hat{\underline{\theta}}$ o estimador de máxima verossimilhança de $\underline{\theta}$ obtido como solução da equação (1.4.4), com $\hat{U}_r = 0$, para $r = 1, \dots, k$, sendo $\hat{U}_r$ obtido avaliando-se U_r em $\hat{\underline{\theta}}$. Suponhamos válidas todas as condições de regularidade da Seção 1.4. Expandindo a componente $\hat{U}_r = 0$ até primeira ordem em torno de $\underline{\theta}$, tem-se $\hat{U}_r = U_r + \sum U_{rs}(\hat{\theta}_s - \theta_s) + O_p(1)$, que equivale a $U_r + \sum U_{rs}(\hat{\theta}_s - \theta_s) + O_p(1) = 0$. Em notação matricial, $U = J(\hat{\underline{\theta}} - \underline{\theta}) + O_p(1)$. De (1.4.1) e (1.4.4) observamos que $J = K + O_p(n^{1/2})$, desta forma $U = K(\hat{\underline{\theta}} - \underline{\theta}) + O_p(1)$. Daí,

$$\hat{\underline{\theta}} - \underline{\theta} = K^{-1}U + O_p(n^{-1}) \quad (3.2.1)$$

A fórmula anterior é importante para calcular os momentos e cumulantes de ordens superiores dos estimadores de máxima verossimilhança. Usando-se novamente as condições de regularidade, expandimos $\hat{U}_r$ até segunda ordem, obtendo

$$\hat{U}_r = U_r + \sum_s U_{rs}(\hat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{s,t} U_{rst}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t) + o_p(1).$$

Tomado o valor esperado,

$$\sum_s E\{U_{rs}(\hat{\theta}_s - \theta_s)\} + \frac{1}{2} \sum_{s,t} E\{U_{rst}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t)\} + o(1) = 0,$$

o que equivale a

$$\sum_s \kappa_{rs} E(\hat{\theta}_s - \theta_s) + \sum_s \text{Cov}(U_{rs}, \hat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{s,t} \kappa_{rst} (-\kappa^{st}) + o(1) = 0, \quad (3.2.2)$$

com $-\kappa^{rs} = \kappa^{r,s}$ representando o elemento (r, s) da inversa K^{-1} da matriz de informação. Usando (3.2.1) segue-se que, até ordem $o(1)$,

$$\text{Cov}(U_{rs}, \hat{\theta}_s - \theta_s) = \text{Cov}\left(U_{rs}, -\sum_t \kappa^{st} U_t\right) = -\sum_t \text{Cov}(U_{rs}, U_t) \kappa^{st} = -\sum_t \kappa_{rs,t} \kappa^{st}. \quad (3.2.3)$$

Definindo o viés de ordem $O(n^{-1})$ de $\hat{\theta}_r$ por $B(\hat{\theta}_r)$ e substituindo (3.2.3) em (3.2.2) vem

$$\sum_s \kappa_{rs} B(\hat{\theta}_s) = \sum_{s,t} \kappa^{st} \left(\kappa_{rs,t} + \frac{1}{2} \kappa_{rst} \right) + o(1),$$

cujas inversões produzem até ordem $O(n^{-1})$, para $r = 1, \dots, k$,

$$B(\hat{\theta}_r) = \sum_{s,t,u} \kappa^{rs} \kappa^{tu} \left(\kappa_{st,u} + \frac{1}{2} \kappa_{stu} \right), \quad (3.2.4)$$

que é uma fórmula geral para determinar o viés de ordem $O(n^{-1})$ em estimadores de máxima verossimilhança em modelos multiparamétricos. Em muitas situações torna-se conveniente substituir, como consequência das identidades de Bartlett, $\kappa_{st,u} + \frac{1}{2} \kappa_{stu}$ por $\kappa_{st}^{(u)} - \frac{1}{2} \kappa_{stu}$, o que facilita expressar o viés em notação matricial. A grande utilidade de (3.2.4) é definir um estimador de máxima verossimilhança corrigido até ordem $O(n^{-1})$ dada por $\tilde{\theta}_r = \hat{\theta} - \hat{B}(\hat{\theta}_r)$, onde $\hat{B}(\cdot)$ é o viés $B(\cdot)$ avaliado em $\hat{\theta}$. Este novo estimador $\tilde{\theta}_r$ tem viés de ordem $O(n^{-2})$, pois $E(\tilde{\theta}) = \theta + O(n^{-2})$, e pode ser preferido em relação a $\hat{\theta}_r$, cujo viés é de ordem $O(n^{-1})$. O cálculo de momentos e cumulantes dos estimadores de máxima verossimilhança de ordem superior, como, por exemplo, $E(\hat{\theta})$ e $\text{Var}(\hat{\theta})$, até ordem $O(n^{-2})$ é bastante complexo, já que precisamos incluir outros termos de ordem superior na expansão (3.2.2). Para uma discussão detalhada sobre a expansão de (3.2.2) até ordens maiores que $O(n^{-2})$, ver Shenton & Bowman (1977).

Observa-se que $B(\hat{\theta})$ pode ser expresso na forma

$$B(\hat{\theta}) = K(\theta)^{-1} A \text{vec} \left(K(\theta)^{-1} \right) + O(n^{-2}), \quad (3.2.5)$$

onde $A = \left[A^{(1)} | \dots | A^{(k)} \right]$ é uma matriz $k \times k^2$ particionada, com $A^{(l)} = (a_{jm}^{(l)})$ tendo elemento típico

$$a_{jm}^{(l)} = \kappa_{jm}^{(l)} - \frac{1}{2} \kappa_{jml}, \quad j, m, l = 1, \dots, k,$$

e $\text{vec}(\cdot)$ o operador que transforma uma matriz em vetor coluna sobrepondo as colunas desta matriz.

3.2.1 Correção de viés dos estimadores de máxima verossimilhança do modelo de regressão beta

Um dos objetivos deste trabalho é obter uma expressão para calcular o viés de segunda ordem usando a fórmula (3.2.4) de Cox & Snell (1968) com o fim de obter estimativas corrigidas dos parâmetros de nosso modelo de regressão beta. Introduzimos a seguinte notação. As derivadas da função de log-verossimilhança com respeito aos parâmetros desconhecidos são indicadas por índices, onde as letras, $r, s, \dots$ correspondem às derivadas em relação às componentes β e a letra ϕ corresponde às derivadas com respeito ao parâmetro ϕ . Logo,

$U_r = \partial\ell/\partial\beta_r$, $U_\phi = \partial\ell/\partial\phi$, $U_{\phi s} = \partial^2\ell/\partial\phi\partial\beta_s$, $U_{rs\phi} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\phi$ e assim por diante. A notação para os momentos destas derivadas é dada por Lawley (1956): $\kappa_{rs} = E(U_{rs})$, $\kappa_{\phi\phi} = E(U_\phi U_\phi)$, $\kappa_{rs,\phi} = E(U_{rs}U_\phi)$, $\kappa_{rst} = E(U_{rst})$, etc., onde todos os κ 's são momentos sob a amostra, e em geral são de ordem $O(n)$. Também as derivadas são denotadas por $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{r\phi}^{(\phi)} = \partial\kappa_{r\phi}/\partial\phi$, etc. Temos que a fórmula (3.2.4) de Cox & Snell (1968) para calcular o viés de segunda ordem do estimador de máxima verossimilhança da a -ésima componente do vetor $\hat{\beta}$ é

$$\begin{aligned} B(\hat{\beta}_a) = & \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} \right\} + \sum_{\phi,s,u} \kappa^{a\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ & + \sum_{r,\phi,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} + \sum_{r,s,\phi} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\ & + \sum_{\phi,\phi,u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} + \sum_{\phi,s,\phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\ & + \sum_{r,\phi,\phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} + \sum_{\phi,\phi,\phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\}, \quad (3.2.1.1) \end{aligned}$$

que pode ser reescrita como

$$\begin{aligned} B(\hat{\beta}_a) = & \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(t)} - \frac{1}{2} \kappa_{rsu} \right\} + \kappa^{a\phi} \sum_{s,u} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ & + \sum_{r,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} + \sum_{r,s} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\ & + \kappa^{a\phi} \sum_u \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} + \kappa^{a\phi} \sum_s \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\ & + \kappa^{\phi\phi} \sum_r \kappa^{ar} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} + \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\}; \quad (3.2.1.2) \end{aligned}$$

para o estimador $\hat{\phi}$ do parâmetro ϕ temos

$$\begin{aligned} B(\hat{\phi}) = & \sum_{r,s,u} \kappa^{\phi r} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} \right\} + \sum_{\phi,s,u} \kappa^{\phi\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ & + \sum_{r,\phi,u} \kappa^{\phi r} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} + \sum_{r,s,\phi} \kappa^{\phi r} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\ & + \sum_{\phi,\phi,u} \kappa^{\phi\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} + \sum_{\phi,s,\phi} \kappa^{\phi\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\ & + \sum_{r,\phi,\phi} \kappa^{\phi r} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} + \sum_{\phi,\phi,\phi} \kappa^{\phi\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\}, \quad (3.2.1.3) \end{aligned}$$

onde $B(\widehat{\beta}_a)$ e $B(\widehat{\phi})$ denotam os vieses de segunda ordem dos estimadores de máxima verossimilhança de $\widehat{\beta}_a$ e $\widehat{\phi}$, respectivamente. Pode-se observar de (2.3.12) que as entradas da matriz $W_{\beta\phi}$ não são nulas, o que leva a uma complicação do cálculo, indicando que todos os termos da expansão (3.2.1.1) devem ser calculados. São deduzidas, através dos momentos e cumulantes obtidos no Apêndice A.1, as parcelas (A.1.30) até (A.1.37) da fórmula de correção de viés de Cox & Snell (1968). A partir de algum algebrismo e usando notação matricial derivamos as parcelas das expressões (3.2.1.2) e (3.2.1.3), de tal forma que, para nosso modelo de regressão beta,

$$\begin{aligned} B(\widehat{\beta}_a) = & \mathcal{L}_a^\top K^{\beta\beta} X^\top W_1 \delta_{\beta\beta} + \mathcal{L}_a^\top K^{\beta\phi} \text{tr}(W_3 X K^{\beta\beta} X^\top) \\ & + \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_3 X) K^{\beta\phi} + \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_2 X) K^{\beta\phi} \\ & + \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_5) X K^{\beta\phi} + \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_4) X K^{\beta\phi} \\ & + \mathcal{L}_a^\top K^{\beta\beta} X^\top \text{vdiag}(W_4)^\top K^{\phi\phi} + \mathcal{L}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(S), \end{aligned}$$

onde $\text{vdiag}(\cdot)$ é o vetor linha formado pela diagonal principal de uma matriz quadrada. Fatorando adequadamente, $B(\widehat{\beta}_a)$ se simplifica na forma

$$\begin{aligned} & \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_1 \delta_{\beta\beta} + (X^\top W_3 X) K^{\beta\phi} + (X^\top W_2 X) K^{\beta\phi} + X^\top \text{vdiag}(W_4)^\top K^{\phi\phi}) \\ & + \mathcal{L}_a^\top K^{\beta\phi} (\text{tr}(W_3 X K^{\beta\beta} X^\top) + \text{vdiag}(W_5) X K^{\beta\phi} + \text{vdiag}(W_4) X K^{\beta\phi} + K^{\phi\phi} \text{tr}(S)). \end{aligned}$$

Assim,

$$\begin{aligned} B(\widehat{\beta}_a) = & \mathcal{L}_a^\top K^{\beta\beta} X^\top (W_1 \delta_{\beta\beta} + (W_3 + W_2) X K^{\beta\phi} + \text{vdiag}(W_4)^\top) \\ & + \mathcal{L}_a^\top K^{\beta\phi} (\text{tr}(W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(S) + \text{vdiag}(W_5 + W_4) X K^{\beta\phi}). \end{aligned}$$

Logo, o viés de segunda ordem do vetor $\widehat{\beta}$ é

$$\begin{aligned} B(\widehat{\beta}) = & K^{\beta\beta} X^\top (W_1 \delta_{\beta\beta} + (W_3 + W_2) X K^{\beta\phi} + \text{vdiag}(W_4)^\top K^{\phi\phi}) \\ & + K^{\beta\phi} (\text{tr}(W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(S) + \text{vdiag}(W_5 + W_4) X K^{\beta\phi}). \end{aligned}$$

Dada a matriz aumentada $\widetilde{X}$ de dimensão $(n+1) \times (k+1)$ definida em (2.3.11), definimos o vetor auxiliar $\widetilde{\delta}$ de dimensão $(n+1) \times 1$ como

$$\widetilde{\delta} = \begin{pmatrix} W_1 \delta_{\beta\beta} + (W_3 + W_2) X K^{\beta\phi} + \text{vdiag}(W_4)^\top \\ \text{tr}(W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(S) + \text{vdiag}(W_5 + W_4) X K^{\beta\phi} \end{pmatrix}.$$

Além disso, definimos

$$K^{\beta*} = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \end{pmatrix},$$

o bloco superior da matriz $K(\varrho)^{-1}$ de dimensão $k \times (k+1)$, obtida em (2.3.14). Desta forma, o viés de segunda ordem de $\widehat{\beta}$ pode ser escrito como

$$B(\widehat{\beta}) = K^{\beta*} \widetilde{X}^\top \widetilde{\delta}. \quad (3.2.1.4)$$

De forma análoga aos cálculos anteriores, deduz-se que

$$\begin{aligned} B(\widehat{\phi}) &= K^{\phi\beta} X^\top W_1 \delta_{\beta\beta} + K^{\phi\phi} \text{tr}(W_3 X K^{\beta\beta} X^\top) \\ &\quad + K^{\phi\beta} (X^\top W_3 X) K^{\beta\phi} + K^{\phi\beta} (X^\top W_2 X) K^{\beta\phi} \\ &\quad + K^{\phi\phi} \text{vdiag}(W_5) X K^{\beta\phi} + K^{\phi\phi} \text{vdiag}(W_4) X K^{\beta\phi} \\ &\quad + K^{\phi\beta} X^\top \text{vdiag}(W_4)^\top K^{\phi\phi} + K^{\phi\phi} K^{\phi\phi} \text{tr}(S), \end{aligned}$$

e $B(\widehat{\phi})$ se simplifica na forma

$$\begin{aligned} &K^{\phi\beta} (X^\top W_1 \delta_{\beta\beta} + (X^\top W_3 X) K^{\beta\phi} + (X^\top W_2 X) K^{\beta\phi} + X^\top \text{vdiag}(W_4)^\top K^{\phi\phi}) \\ &+ K^{\phi\phi} (\text{tr}(W_3 X K^{\beta\beta} X^\top) + \text{vdiag}(W_5) X K^{\beta\phi} + \text{vdiag}(W_4) X K^{\beta\phi} + K^{\phi\phi} \text{tr}(S)). \end{aligned}$$

Desta forma,

$$\begin{aligned} B(\widehat{\phi}) &= K^{\phi\beta} X^\top (W_1 \delta_{\beta\beta} + (W_3 + W_2) X K^{\beta\phi} + \text{vdiag}(W_4)^\top K^{\phi\phi}) \\ &\quad + K^{\phi\phi} (\text{tr}(W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(S) + \text{vdiag}(W_5 + W_4) X K^{\beta\phi}). \end{aligned}$$

Definimos

$$K^{\phi*} = \begin{pmatrix} K^{\phi\beta} & K^{\phi\phi} \end{pmatrix},$$

o bloco inferior da matriz $K(\underline{\theta})^{-1}$ de dimensão $1 \times (k+1)$, Assim, o viés de segunda ordem de $\widehat{\phi}$ pode ser escrito como

$$B(\widehat{\phi}) = K^{\phi*} \widetilde{X}^\top \widetilde{\delta}. \quad (3.2.1.5)$$

Agora, se consideramos o vetor conjunto $\underline{\theta} = (\beta^\top, \phi)^\top$, podemos escrever a partir de (3.2.1.4) e (3.2.1.5) uma expressão para o viés de segunda ordem do estimador de máxima verossimilhança de $\underline{\theta}$ da forma

$$B(\widehat{\underline{\theta}}) = K(\underline{\theta})^{-1} \widetilde{X}^\top \widetilde{\delta},$$

ou, alternativamente, a partir de (2.3.13) como

$$B(\widehat{\underline{\theta}}) = (\widetilde{X}^\top \widetilde{W} \widetilde{X})^{-1} \widetilde{X}^\top \widetilde{\delta},$$

que pode ser escrita como

$$B(\widehat{\underline{\theta}}) = (\widetilde{X}^\top \widetilde{W} \widetilde{X})^{-1} \widetilde{X}^\top \widetilde{W} \widetilde{\xi},$$

onde definimos $\widetilde{\xi} = \widetilde{W}^{-1} \widetilde{\delta}$. Desta forma, os coeficientes do vetor $B(\widehat{\underline{\theta}})$ podem ser estimados a partir de uma regressão de mínimos quadrados generalizados dada pela equação anterior; com este resultado, definimos um estimador de máxima verossimilhança corrigido $\widetilde{\underline{\theta}}$ da forma

$$\widetilde{\underline{\theta}} = \widehat{\underline{\theta}} - \widehat{B}(\widehat{\underline{\theta}}), \quad (3.2.1.6)$$

onde $\widehat{B}(\widehat{\underline{\theta}})$ denota o estimador de máxima verossimilhança de $B(\widehat{\underline{\theta}})$, i.e., os parâmetros desconhecidos são substituídos por suas respectivas estimativas de máxima verossimilhança. Diremos que $\widetilde{\underline{\theta}}$ é um estimador de máxima verossimilhança corrigido “corretivamente”.

Sabemos que sob as condições de regularidade da Seção 1.4 o estimador de máxima verossimilhança $\hat{\theta}$ é consistente para θ e pela expressão (3.2.4) o viés $B(\hat{\theta})$ de segunda ordem converge para zero quando o tamanho de amostra n tende para infinito. Assim, o estimador corrigido $\tilde{\theta}$ é um estimador consistente para θ e

$$E(\tilde{\theta}) = \theta + O(n^{-2}).$$

Além disso, dada a convergência em distribuição do estimador de máxima verossimilhança a uma normal $(k+1)$ -variada da forma (1.4.5) e usando o teorema de Slutsky, o estimador corrigido satisfaz

$$\tilde{\theta} \stackrel{A}{\sim} \mathcal{N}_{k+1}(\theta, K^{-1}(\theta)).$$

Na Figura 3.2.1.1, EMV denota o estimador de máxima verossimilhança e ECCS denota o estimador de máxima verossimilhança corrigido por viés de Cox & Snell (1968), para um exemplo em particular.

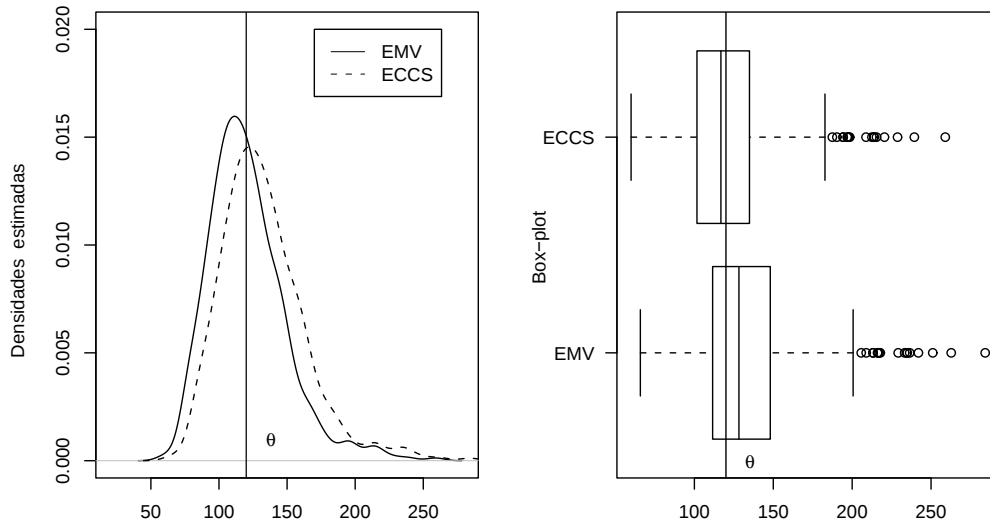


Figura 3.2.1.1: Comparação das densidades dos estimadores de máxima verossimilhança e corrigido por viés de Cox & Snell (1968)

Podemos utilizar de forma natural como estimador consistente da matriz de informação de Fisher $K(\theta)$, a matriz $K(\tilde{\theta})$, para tamanhos de amostra moderados, ou $K(\hat{\theta})$, já que em princípio, a correção do estimador $\hat{\theta}$ por viés só deveria deslocar sua distribuição na direção do verdadeiro valor do parâmetro sem afetar muito a estrutura de variância, como pode ser visto na Figura 3.2.1.1. Com esta idéia em mente, para um tamanho de amostra suficientemente grande, podemos construir intervalos de confiança assintóticos da forma (1.6.1).

Assim, para $r = 1, \dots, k$,

$$\left(\tilde{\beta}_r - z_{1-\frac{\alpha}{2}}(K(\tilde{\theta})^{rr})^{1/2}, \quad \tilde{\beta}_r + z_{1-\frac{\alpha}{2}}(K(\tilde{\theta})^{rr})^{1/2} \right), \quad (3.2.1.7)$$

$$\left(\tilde{\phi}_1 - z_{1-\frac{\alpha}{2}}(K(\tilde{\theta})^{\phi\phi})^{1/2}, \quad \tilde{\phi}_1 + z_{1-\frac{\alpha}{2}}(K(\tilde{\theta})^{\phi\phi})^{1/2} \right), \quad (3.2.1.8)$$

ou

$$\left(\tilde{\beta}_r - z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{rr})^{1/2}, \quad \tilde{\beta}_r + z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{rr})^{1/2} \right) \quad (3.2.1.9)$$

e

$$\left(\tilde{\phi}_1 - z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{\phi\phi})^{1/2}, \quad \tilde{\phi}_1 + z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{\phi\phi})^{1/2} \right), \quad (3.2.1.10)$$

que são intervalos de confiança assintóticos para β_r e ϕ de tamanho $100(1 - \alpha)\%$, respectivamente. Para os intervalos (3.2.1.7) e (3.2.1.8), as variâncias estimadas de $\tilde{\beta}_r$ e $\tilde{\phi}$ são $K(\tilde{\theta})^{rr}$ e $K(\tilde{\theta})^{\phi\phi}$, respectivamente, com $K(\tilde{\theta})^{rr}$ o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em $\tilde{\theta}$. No caso dos intervalos (3.2.1.9) e (3.2.1.10) as variâncias estimadas para $\tilde{\beta}_r$ e $\tilde{\phi}$ são $K(\hat{\theta})^{rr}$ e $K(\hat{\theta})^{\phi\phi}$, respectivamente, com $K(\hat{\theta})^{rr}$ o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em $\hat{\theta}$. Chamamos neste trabalho os intervalos da forma (3.2.1.7) a (3.2.1.10) de intervalos assintóticos corrigidos corretivamente (IACC). Dado que os estimadores corrigidos são consistentes para os parâmetros do modelo, esperamos que os valores das variâncias assintóticas de cada estimativa corrigida sejam razoavelmente menores para tamanhos moderados de amostra que suas correspondentes estimativas de máxima verossimilhança. Assim, os intervalos (3.2.1.7) e (3.2.1.8) comparados com os intervalos (3.2.1.9) e (3.2.1.10) devem em princípio possuir um comprimento menor, proporcionando estimativas intervalares mais precisas. Contudo, como estamos trabalhando com um resultado de tipo assintótico, percebemos que o efeito da correção não deve ser alto em termos da estrutura de variância, i.e., esperamos que os intervalos (3.2.1.7), (3.2.1.9) e (3.2.1.8), (3.2.1.10) sejam muito parecidos em termos de sua simetria e comprimento. .

3.2.2 O viés do vetor de médias

Consideremos o vetor de médias de nosso modelo de regressão, $\mu = (\mu_1, \dots, \mu_n)^\top$, onde $\mu_t = g^{-1}(x_t^\top \beta)$, $t = 1, \dots, n$ e g^{-1} é a função inversa da função de ligação. Assim, o vetor de médias é uma função dos parâmetros de regressão da forma $\mu = f(\beta)$ com $f = (f_1, \dots, f_n)$, e $f_t = g^{-1}(x_t^\top \beta)$, $t = 1, \dots, n$. Pelo princípio de invariância o estimador de máxima verossimilhança de μ é $\hat{\mu} = f(\hat{\beta})$. Para $t = 1, \dots, n$, definimos $d_{tr} = \frac{\partial f_t}{\partial \beta_r}$ e $v_{trs} = \frac{\partial^2 f_t}{\partial \beta_r \partial \beta_s}$, com $r, s = 1, \dots, k$. Além disso, definimos D a matriz de dimensão $n \times k$ com elemento típico $(t, r) = d_{tr}$ e $V = [V_1 | \dots | V_n]^\top$ é a matriz de dimensão $n \times k \times k$ com elemento

$(t, r, s) = v_{trs}$. Se tomamos uma expansão em serie de Taylor de f até o terceiro termo, numa vizinhança de $\hat{\beta}$, obtemos

$$f(\hat{\beta}) \approx f(\beta) + D(\hat{\beta} - \beta) + \frac{1}{2}(\hat{\beta} - \beta)^\top V(\hat{\beta} - \beta), \quad (3.2.2.1)$$

onde $\frac{1}{2}(\hat{\beta} - \beta)^\top V(\hat{\beta} - \beta)$ é um vetor de dimensão $n \times 1$ com t -ésima componente $\frac{1}{2}(\hat{\beta} - \beta)^\top V_t(\hat{\beta} - \beta)$ e V_t a t -ésima matriz quadrada de dimensão $k \times k$ da forma

$$V_t = \begin{pmatrix} v_{t11} & \cdots & v_{t1k} \\ \vdots & \ddots & \vdots \\ v_{tk1} & \cdots & v_{tkk} \end{pmatrix}.$$

Tomando a esperança para os dois membros da igualdade (3.2.2.1) temos

$$E(f(\hat{\beta}) - f(\beta)) \approx DE(\hat{\beta} - \beta) + \frac{1}{2}E((\hat{\beta} - \beta)^\top V(\hat{\beta} - \beta)).$$

Para $t = 1, \dots, n$, e usando-se (3.2.4), temos que

$$\begin{aligned} \frac{1}{2}E((\hat{\beta} - \beta)^\top V_t(\hat{\beta} - \beta)) &= \frac{1}{2}\left\{\text{tr}[V_t \text{Cov}(\hat{\beta} - \beta)] + B(\hat{\beta})^\top V_t B(\hat{\beta})\right\} \\ &= \frac{1}{2}\text{tr}\{K^{\beta\beta} V_t\} + \frac{1}{2}\{O(n^{-2})O(n^2)O(n^{-2})\} \\ &= \frac{1}{2}\text{tr}\{K^{\beta\beta} V_t\} + O(n^{-2}). \end{aligned}$$

Desta forma,

$$\begin{aligned} B(\hat{\mu}) &= DB(\hat{\beta}) + \frac{1}{2}E((\hat{\beta} - \beta)^\top V(\hat{\beta} - \beta)) \\ &= DB(\hat{\beta}) + \frac{1}{2}\text{tr}(V \text{Cov}(\hat{\beta} - \beta)) + O(n^{-2}) \\ &= DB(\hat{\beta}) + \frac{1}{2}\text{tr}(V K^{\beta\beta}) + O(n^{-2}). \end{aligned}$$

Portanto, o viés de ordem $O(n^{-2})$ de $\hat{\mu}$ obtido através do viés de ordem $O(n^{-2})$ de $\hat{\beta}$ é da forma

$$B(\hat{\mu}) = DB(\hat{\beta}) + \frac{1}{2}\text{tr}(V K^{\beta\beta}).$$

De forma análoga ao estimador definido em (3.2.1.6), podemos construir um estimador corrigido de máxima verossimilhança do vetor de média $\hat{\mu}$ da forma

$$\tilde{\mu} = \hat{\mu} - \hat{B}(\hat{\mu}),$$

de tal forma que

$$E(\tilde{\mu}) = \mu + O(n^{-2}).$$

3.3 Correção preventiva

Sob as condições de regularidade da seção (1.4), o viés do estimador de máxima verossimilhança $\hat{\theta}$ de θ pode ser expresso como uma série de potências de $1/n$ de forma

$$B(\hat{\theta}) = \frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2} + \dots, \quad (3.3.1)$$

sendo n o tamanho de amostra. David Firth, em 1993, desenvolveu um método alternativo para corrigir o viés de ordem $O(n^{-1})$ dos estimadores de máxima verossimilhança. O método, chamado de “preventivo”, consiste numa modificação conveniente da função escore que permite remover o termo n^{-1} da expansão (3.3.1), sob a suposição de que o modelo é regular e o estimador de máxima verossimilhança definido através da equação (1.4.1) é único. A idéia de Firth é simples, já que ao introduzir um pequeno viés na função escore o viés de $\hat{\theta}$ pode ser reduzido. Se $\hat{\theta}$ apresenta viés positivo $B(\theta)$ e o gradiente local esperado $K(\theta)$ da função escore define sua inclinação, então, a função escore é deslocada pela quantidade $K(\theta)B(\theta)$, como é ilustrado na Figura 3.3.1. A quantidade $K(\theta)$ é a matriz de informação. Este deslocamento define uma função escore modificada dada por

$$U^*(\theta) = U(\theta) + K(\theta)B(\theta). \quad (3.3.2)$$

Assim, a solução θ^* de $U^* = 0$ produz uma estimativa modificada corrigida.

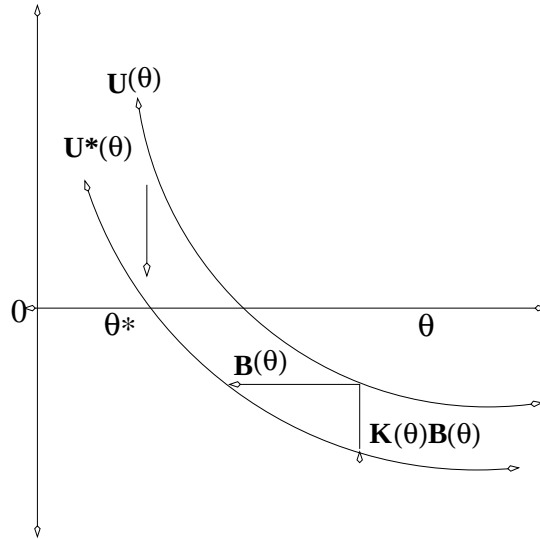


Figura 3.3.1: Função escore modificada.

No caso de um vetor de parâmetros, U^* é vista como uma equação vetorial. O viés em (3.3.2) é em geral desconhecido, mas é possível obter o termo de ordem n^{-1} utilizando a fórmula de Cox & Snell (3.2.4), de tal forma que se $\theta = (\theta_1, \dots, \theta_k)$,

$$\frac{B(\hat{\theta}_r)}{n} = \sum_{s,t,u} \kappa^{rs} \kappa^{tu} \left(\kappa_{st,u} + \frac{1}{2} \kappa_{stu} \right) \quad (3.3.3)$$

onde r, s, t, u indexam o espaço paramétrico. Usamos a notação padrão para os momentos das derivadas da função de log-verossimilhança, dada por Lawley (1956) a qual foi apresentada na Seção (1.5). Podemos observar, que para nosso modelo de regressão beta, ao substituir $B(\underline{\theta})$ por $B(\hat{\underline{\theta}}) = K(\underline{\theta})^{-1} \tilde{X}^\top \tilde{\delta}$, a função escore modificada (3.3.2) para nosso modelo de regressão beta é

$$\begin{aligned} U^*(\underline{\theta}) &= U(\underline{\theta}) + K(\underline{\theta})B(\hat{\underline{\theta}}) \\ &= U(\underline{\theta}) + K(\underline{\theta})K(\underline{\theta})^{-1} \tilde{X}^\top \tilde{\delta} \\ &= U(\underline{\theta}) + \tilde{X}^\top \tilde{\delta}. \end{aligned}$$

No caso em que $E[U_r U_{st}] = 0$ para todo r, s, t , podemos verificar que θ^* é obtido ao maximizar-se uma função log-verossimilhança modificada $\ell^* = \ell + \frac{1}{2} \log |K|$. Isto ocorre, por exemplo, quando as derivadas de segunda ordem da função de log-verossimilhança com respeito aos parâmetros não dependem de valores amostrais. Segundo Firth (1993) e de forma análoga ao método corretivo discutido na seção anterior, o estimador θ^* apresenta as mesmas propriedades que o estimador de máxima verossimilhança, i.e., θ^* é consistente para θ , e sua distribuição assintótica é tal que

$$\theta^* \overset{A}{\sim} \mathcal{N}(\theta, K^{-1}(\theta)),$$

como está mostrado na Figura 3.3.2. Além disso, se o modelo em análise pertence à família exponencial completa, se garante que

$$E(\theta^*) = \theta + O(n^{-2}).$$

No gráfico, EMV denota o estimador de máxima verossimilhança e DFirth denota o estimador preventivo de David Firth.

Se utilizamos, como estimador consistente da matriz de informação de Fisher $K(\theta)$ a matriz $K(\theta^*)$, podemos construir intervalos de confiança assintóticos da forma (1.6.1) para os parâmetros do modelos de regressão beta. Assim, para $r = 1, \dots, k$,

$$\left(\beta_r^* - z_{1-\frac{\alpha}{2}}(K(\theta^*)^{rr})^{1/2}, \quad \beta_r^* + z_{1-\frac{\alpha}{2}}(K(\theta^*)^{rr})^{1/2} \right) \quad (3.3.4)$$

e

$$\left(\phi^* - z_{1-\frac{\alpha}{2}}(K(\theta^*)^{\phi\phi})^{1/2}, \quad \phi^* + z_{1-\frac{\alpha}{2}}(K(\theta^*)^{\phi\phi})^{1/2} \right) \quad (3.3.5)$$

são intervalos de confiança assintóticos para β_r e ϕ de tamanho $100(1-\alpha)\%$, respectivamente. As variâncias estimadas de β_r^* e ϕ^* são $K(\theta^*)^{rr}$ e $K(\theta^*)^{\phi\phi}$, respectivamente, onde $K(\theta^*)^{rr}$ é o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em θ^* . Analogamente ao discutido na Seção 3.2.1, podemos construir intervalos assintóticos da forma (1.6.2), utilizando como estimador consistente da matriz de informação de Fisher a matriz $K(\hat{\theta})$. Desta forma, os intervalos para β_r e ϕ de tamanho $100(1-\alpha)\%$ utilizando as estimativas corrigidas preventivas serão respectivamente

$$\left(\beta_r^* - z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{rr})^{1/2}, \quad \beta_r^* + z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{rr})^{1/2} \right), \quad (3.3.6)$$

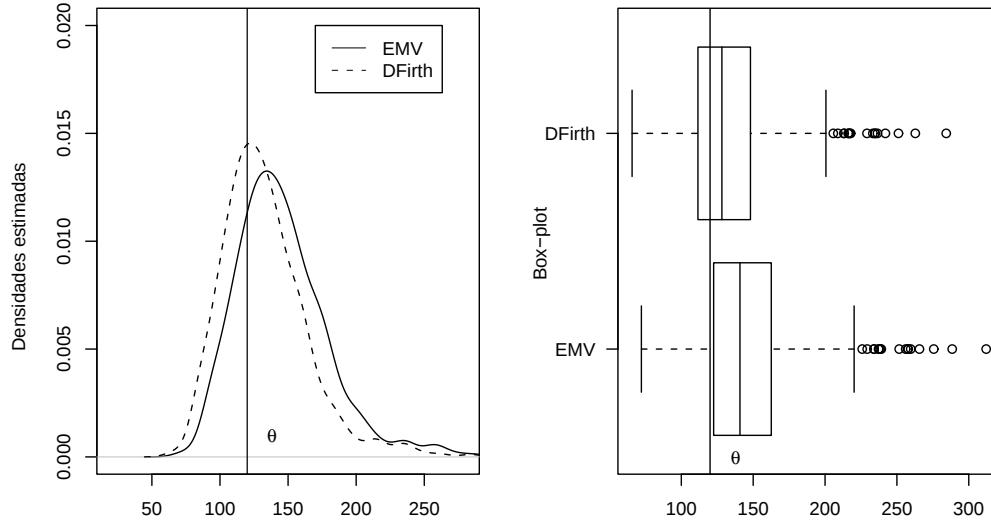


Figura 3.3.2: Comparação das densidades dos estimadores de máxima verossimilhança e corrigido preventivo.

$$\left(\phi^* - z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{\phi\phi})^{1/2}, \quad \phi^* + z_{1-\frac{\alpha}{2}}(K(\hat{\theta})^{\phi\phi})^{1/2} \right), \quad (3.3.7)$$

onde as variâncias estimadas de β_r^* e ϕ^* são $K(\hat{\theta})^{rr}$ e $K(\hat{\theta})^{\phi\phi}$, respectivamente. $K(\hat{\theta})^{rr}$ denota o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em $\hat{\theta}$. Neste trabalho, chamamos os intervalos da forma (3.3.4) até (3.3.7) de intervalos assintóticos corrigidos preventivamente (IACP).

3.4 Correção por Bootstrap

Os métodos estatísticos computacionais surgiram exatamente da necessidade de simular um experimento ou um fenômeno diversas vezes, com o objetivo de obter resultados que a teoria matemática não pode fornecer de maneira rápida e eficiente. O bootstrap é um método computacional de inferência estatística introduzido por Efron (1979) capaz de responder a questões reais sem a necessidade de exaustivos, complicados e muitas vezes inviáveis cálculos analíticos. Efron & Tibshirani (1993, p. 2) escrevem: “The basic ideas of statistics haven’t changed, but their implementation has. The modern computer lets us apply these ideas flexibly, quickly, easily, and with a minimum of mathematical assumptions.”

A metodologia bootstrap permite a estimação de variâncias, intervalos de confiança, p -valores e outras quantidades de interesse da inferência estatística através de reamostragem direta dos dados, os quais são tratados como se fossem a própria população. Tipicamente, o

método bootstrap fornece aproximações para a distribuição de uma estatística de interesse que podem ser consideravelmente mais precisas que as aproximações assintóticas de primeira ordem. A aplicação de bootstrap para modelos de regressão foi estudada detalhadamente por Wu (1986) e, mais recentemente, muitos autores têm investigado o uso deste método em econometria, entre eles Berkowitz & Kilian (2000), Horowitz (1997), Jeong & Maddala (1996) e Vinod (1993).

Consideremos uma amostra aleatória $\mathcal{Y} = (y_1, \dots, y_n)$ de uma variável aleatória populacional Y cuja distribuição está determinada completamente por sua função de distribuição acumulada P . Seja $\theta = t(P)$ uma função de P denominada parâmetro e $\hat{\theta} = S(\mathcal{Y})$ um estimador de θ . Se P é desconhecida, algumas propriedades distribucionais de $\hat{\theta}$ podem ser obtidas utilizando teoria assintótica de primeira ordem. No entanto, tal estimação pode ser muito difícil e até mesmo inviável de se obter. A aplicação de bootstrap, neste caso, consiste em obter, a partir da amostra original $\mathcal{Y}$, um grande número de amostras $y^* = (y_1^*, \dots, y_n^*)$, calcular as respectivas réplicas bootstrap de $\hat{\theta}$, $\hat{\theta}^* = S(y^*)$, e, com base na distribuição empírica de $\hat{\theta}^*$, estimar a função de distribuição de $\hat{\theta}$. Existem duas versões do método bootstrap que se diferenciam unicamente na forma de obtenção das amostras. Na primeira versão, a amostra bootstrap y^* é obtida de uma estimativa não-paramétrica de P , onde a função de distribuição empírica da amostra original $\mathcal{Y}$ é dada por

$$\hat{P}(\iota) = \frac{\#\{y_i \leq \iota\}}{n},$$

a qual atribui probabilidade $1/n$ a cada valor y_i , $i = 1, \dots, n$. Se y^* é uma amostra aleatória de $\hat{P}$, então ela é formada por n observações extraídas aleatoriamente com reposição da amostra $\mathcal{Y}$, que é tratada como se fosse a população. Neste caso trata-se do bootstrap não-paramétrico, que é uma aplicação direta do princípio denominado “plug in”, o qual utiliza $\hat{P}$ no lugar da distribuição desconhecida populacional. Logo, se $\theta = t(P)$, a respectiva estimativa “plug in” é $\hat{\theta} = t(\hat{P})$. A estimativa bootstrap não-paramétrica de qualquer característica de P é uma estimativa “plug in”, uma vez que todo processo de estimação parte das amostras y^* extraídas de $\hat{P}$. Na segunda versão é necessário que P pertença a uma família paramétrica dimensionalmente finita e conhecida, P_η .

A obtenção de um estimador consistente para η viabiliza uma estimativa paramétrica de P , representada por $P_{\hat{\eta}}$. É importante ressaltar que a existência do modelo paramétrico não exclui a análise não-paramétrica, já que esta última pode ser muito útil no momento de avaliar a veracidade das conclusões obtidas a partir do método paramétrico. A utilização do método bootstrap se justifica, por exemplo, quando a teoria assintótica fornece aproximações pouco precisas da distribuição de uma estatística de teste. Neste caso, as diferenças entre o nível empírico do teste (realizado com base em valores críticos assintóticos) e o nível verdadeiro podem ser substanciais. A aplicação de bootstrap, neste caso, é de grande relevância, uma vez que o método pode reduzir consideravelmente, ou até eliminar, as distorções de tamanho de testes estatísticos em amostras finitas. No entanto, nestas circunstâncias, onde a distribuição assintótica pode ser obtida, não se pode esperar que o método bootstrap forneça

aproximações mais precisas, a não ser que a distribuição da estatística em questão seja assintoticamente pivotal, i.e., independente de parâmetros desconhecidos.

Segundo Horowitz (1994), procedimentos de bootstrap simples fornecem aproximações melhoradas para a distribuição de estatísticas assintoticamente pivotaes, mas não para a distribuição de estatísticas que não apresentam esta propriedade. Beran (1986) mostra que se a distribuição assintótica da estatística, sob a hipótese nula, é pivotal, então, sob algumas condições de regularidade, os tamanhos de testes bootstrap apresentam erros de ordem menor, i.e., que convergem mais rapidamente para zero, que os erros dos testes baseados na teoria assintótica. Denotamos o viés do estimador $\hat{\theta} = S(\underline{y})$ por $B_P(\theta, \hat{\theta})$, da forma

$$B_P(\theta, \hat{\theta}) = E_P[S(\underline{y})] - t(P), \quad (3.4.1)$$

onde o subscrito P indica que a esperança matemática é obtida com base na medida P . Os estimadores de bootstrap do viés nas versões paramétrica e não-paramétrica são definidos substituindo a verdadeira distribuição P , que gerou a amostra original, pelas distribuições P_η e $\hat{P}$, respectivamente, na expressão (3.2.3.1). Assim, os vieses nestas duas formas de bootstrap são expressos por

$$\begin{aligned} B_{P_\eta}(\theta, \hat{\theta}) &= E_{P_\eta}[S(\underline{y})] - t(P_\eta), \\ B_{\hat{P}}(\theta, \hat{\theta}) &= E_{\hat{P}}[S(\underline{y})] - t(\hat{P}). \end{aligned}$$

Gerando R amostras bootstrap $(y_1^*, \dots, y_R^*)$ de forma independente e calculando as respectivas réplicas bootstrap $(\theta_1^*, \dots, \theta_R^*)$, pode-se aproximar as esperanças bootstrap $E_{P_\eta}[S(\underline{y})]$ e $E_{\hat{P}}[S(\underline{y})]$ pela média

$$\hat{\theta}^*(\cdot) = \frac{1}{R} \sum_{i=1}^R \theta_i^*.$$

Desta forma, os vieses estimados por bootstrap baseados nas R réplicas do estimador $\hat{\theta}$ são

$$\begin{aligned} \hat{B}_{P_\eta}(\theta, \hat{\theta}) &= \hat{\theta}^*(\cdot) - S(\underline{y}), \\ \hat{B}_{\hat{P}}(\theta, \hat{\theta}) &= \hat{\theta}^*(\cdot) - S(\underline{y}). \end{aligned}$$

A diferença entre as estimativas anteriores baseia-se na forma de geração das amostras bootstrap para obtenção das R estimativas $\hat{\theta}^*(\cdot)$, $1 \leq i \leq R$. Tipicamente, o cálculo de estimativas do viés requer entre 50 e 200 amostras de bootstrap. Utilizando as estimativas bootstrap do viés definem-se os estimadores corrigidos até segunda ordem por bootstrap como

$$\bar{\theta}_1 = S(\underline{y}) - \hat{B}_{P_\eta}(\theta, \hat{\theta}) = 2S(\underline{y}) - \hat{\theta}^*(\cdot), \quad (3.4.2)$$

$$\bar{\theta}_2 = S(\underline{y}) - \hat{B}_{\hat{P}}(\theta, \hat{\theta}) = 2S(\underline{y}) - \hat{\theta}^*(\cdot), \quad (3.4.3)$$

onde as estimativas $\bar{\theta}_1$ e $\bar{\theta}_2$ são denominadas estimativas CBC (Constant-Bias-Corrections), MacKinnon & Smith (1998). No contexto deste trabalho utilizaremos como o estimador

corrigido por bootstrap de $\hat{\theta}$ a quantidade $\overline{\theta}$ na versão paramétrica (3.4.2). Uma discussão detalhada sobre a correção de viés de segunda ordem por bootstrap e sua relação com a correção analítica pode ser encontrada em Ferrari & Cribari–Neto (1998).

3.5 Intervalos de confiança bootstrap

A grande preocupação de muitos estatísticos recai em encontrar métodos que permitam reduzir o erro de cobertura de intervalos da forma (1.6.1). Alguns deles utilizam refinamentos da aproximação normal através das chamadas *expansões de Edgeworth*, mas tais expansões, em geral, são de difícil obtenção e não muito úteis na prática. Entretanto, a construção de intervalos de confiança usando bootstrap fornece intervalos de confiança com níveis de cobertura mais próximos da verdadeira probabilidade de cobertura. Singh (1981) e Bickel & Freedman (1980) estudam o método bootstrap e mostram como as propriedades deste estão intimamente relacionadas às expansões de Edgeworth. Se utilizamos o método bootstrap para obter a distribuição empírica $\hat{P}$ de $\hat{\theta}$, podemos construir o denominado **intervalo percentil** (ICP), com um nível aproximado de cobertura de $1 - \alpha$ definido pelos percentis $\alpha/2$ e $1 - \alpha/2$ de $\hat{P}$, i.e., o intervalo é da forma

$$(\hat{P}^{-1}(\alpha/2), \hat{P}^{-1}(1 - \alpha/2)). \quad (3.4.4)$$

Utilizando a metodologia bootstrap descrita na seção anterior, se ordenamos as R réplicas bootstrap de $\hat{\theta}$, os limites inferior e superior do intervalo de confiança percentil serão as réplicas $R(\alpha/2)$ e $R(1 - \alpha/2)$, respectivamente, assumindo que $R(\alpha/2)$ e $R(1 - \alpha/2)$ sejam inteiros e $0 < \alpha < 0.5$. Assim, este intervalo não necessariamente é simétrico em relação ao valor $\hat{\theta}$. Seja k o maior inteiro menor ou igual ao número $(R + 1)\alpha/2$, no caso de $R(\alpha/2)$ e $R(1 - \alpha/2)$ não sejam inteiros tomamos como limites inferior e superior do intervalo percentil, as réplicas k e $(R + 1 - k)$, respectivamente. Como observamos, esta construção garante a não inclusão de valores impróprios para o parâmetro de interesse no intervalo de confiança. Neste trabalho, ainda consideramos intervalos de confiança de tipo percentil para θ utilizando a distribuição empírica $\hat{P}$ dos estimadores corrigidos $\tilde{\theta}$ e θ^* , os quais chamamos de intervalos de confiança percentil corrigido corretivamente (ICCC) e preventivamente (ICCP), respectivamente.

Suponhamos que $\hat{\theta}$ não possui distribuição normal em amostras finitas e que g é uma função que normaliza a distribuição, i.e.,

$$\hat{\rho} \sim \mathcal{N}(\rho, \tau^2),$$

com $\hat{\rho} = g(\hat{\theta})$, onde τ é algum desvio padrão constante. Desta forma, podemos construir um intervalo de confiança exato com probabilidade de cobertura $1 - \alpha$ para o parâmetro ρ da forma

$$\left(\hat{\rho} - z_{1-\frac{\alpha}{2}}\tau, \hat{\rho} + z_{1-\frac{\alpha}{2}}\tau \right).$$

Outra vantagem do método percentil é que o intervalo percentil de $\hat{\theta}$ pode ser expresso por

$$\left(g^{-1}(\hat{\rho} - z_{1-\frac{\alpha}{2}}\tau), g^{-1}(\hat{\rho} + z_{1-\frac{\alpha}{2}}\tau) \right),$$

sem a necessidade de conhecer a função normalizadora, já que, por exemplo, a quantidade $g^{-1}(\hat{\rho} + z_{1-\frac{\alpha}{2}}\tau)$ é estimada pela estimativa bootstrap $g^{-1}(\hat{\rho} + [g(\hat{P}(1-\alpha/2)) - \hat{\rho}])$. Embora este método seja útil, ele pode ter o problema de subestimar as caudas da distribuição bootstrap, podendo assim prejudicar o desempenho nos níveis de cobertura. (Efron & Tibshirani, 1993).

Uma forma de generalizar o método anterior é permitir que a função normalizadora e estabilizadora da variância apresente um viés constante v_0 , i.e.,

$$(\hat{\rho} - \rho)/\tau \sim \mathcal{N}(-v_0, 1).$$

Desta forma, um intervalo de confiança exato com probabilidade de cobertura $1 - \alpha$ para o parâmetro ρ é dado por

$$\left(\hat{\rho} + \tau v_0 - z_{1-\frac{\alpha}{2}}\tau, \hat{\rho} + \tau v_0 + z_{1-\frac{\alpha}{2}}\tau \right),$$

o qual podemos transformar num intervalo de confiança para θ aplicando a transformação inversa g^{-1} aos seus limites. Desta forma, o intervalo de confiança com nível de cobertura de aproximadamente $1 - \alpha$ para θ é da forma

$$(\hat{P}^{-1}(\Phi(z_{[\alpha/2]})), \hat{P}^{-1}(\Phi(z_{[1-\alpha/2]}))). \quad (3.4.5)$$

Efron (1981), denominou o intervalo anterior de BC (“Bias-corrected”), sendo $z_{\alpha/2}$ e $z_{1-\alpha/2}$ os percentis $\alpha/2$ e $(1 - \alpha/2)$ da distribuição normal padrão, $\Phi(\cdot)$ a função de distribuição acumulada normal padrão e

$$\begin{aligned} z_{[\alpha/2]} &= 2v_0 + z_{\alpha/2}, \\ z_{[1-\alpha/2]} &= 2v_0 + z_{1-\alpha/2}. \end{aligned}$$

Como o valor de v_0 é desconhecido, podemos utilizar uma estimativa $\hat{v}_0$ da medida do viés da mediana da distribuição com relação a $\hat{\theta}$ através de

$$\hat{v}_0 = \Phi^{-1}\left(\frac{\#\hat{\theta}_i^* < \hat{\theta}}{R}\right), \quad (3.4.6)$$

onde $\Phi^{-1}(\cdot)$ é a inversa da função de distribuição normal padrão.

Uma extensão do método BC é considerar a possibilidade de variação do erro padrão para diferentes valores do parâmetro θ . Assim, o intervalo de confiança BCa (“Bias-Corrected and accelerated”) é baseado numa suposição mais geral (DiCiccio & Tibshirani, 1987). Suponhamos que

$$\hat{\rho} - \rho \sim \mathcal{N}(-v_0(1 + a\rho), (1 + a\rho)^2),$$

onde $\rho = g(\theta)$ é uma transformação monótona. A estatística padronizada é normal, com algum viés e provavelmente com erro padrão variando linearmente com o parâmetro ρ . Podemos notar que ao tomar $a = 0$ no método BCa sua forma distribucional é idêntica à forma

distribucional do método BC, exceto pela constante τ . Embora aparentemente os métodos sejam parecidos, eles apresentam uma grande diferença, já que no método BC a transformação estabiliza e normaliza a variância de $\hat{\theta}$, enquanto no método BCa a transformação só normaliza a distribuição de $\hat{\theta}$. Desta forma, o intervalo de confiança BCa com probabilidade de cobertura de aproximadamente $1 - \alpha$ para θ é da forma

$$(\hat{P}^{-1}(\Phi(z_{[\alpha/2]}^*)), \hat{P}^{-1}(\Phi(z_{[1-\alpha/2]}^*))), \quad (3.4.7)$$

onde

$$z_{[\alpha/2]}^* = v_0 + \frac{v_0 + z_{\alpha/2}}{1 - a(v_0 + z_{\alpha/2})},$$

$$z_{[1-\alpha/2]}^* = v_0 + \frac{v_0 + z_{1-\alpha/2}}{1 - a(v_0 + z_{1-\alpha/2})},$$

e a é uma constante que mede a razão de mudança do erro padrão de $\hat{\theta}$ com respeito ao verdadeiro valor do parâmetro, chamada de constante de aceleração. Esta constante não é conhecida, mas pode ser estimada através da expressão

$$\hat{a} = \frac{1}{6} \text{Skew}(\dot{\ell}_{\theta}(\hat{\theta}))|_{\theta=\hat{\theta}},$$

onde $\text{Skew}(\cdot)$ denota o coeficiente de assimetria de uma distribuição dada e $\dot{\ell}_{\theta}(\hat{\theta})$ é a derivada da função de log-verossimilhança avaliada no ponto $\hat{\theta}$. Portanto, de forma análoga ao método percentil, os limites do intervalo BCa com nível de cobertura aproximadamente $1 - \alpha$ são os percentis δ_1 e δ_2 definidos através da distribuição bootstrap de $\hat{\theta}$ por

$$\delta_1 = \Phi\left(\hat{v}_0 + \frac{\hat{v}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{v}_0 + z_{\alpha/2})}\right),$$

$$\delta_2 = \Phi\left(\hat{v}_0 + \frac{\hat{v}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{v}_0 + z_{1-\alpha/2})}\right).$$

A principal desvantagem do método BCa é a grande quantidade de réplicas bootstrap (em geral, são usadas entre 1000 e 2000 réplicas bootstrap) de $\hat{\theta}$, aumentando assim o custo computacional. Os métodos anteriores estão resumidos na Tabela 3.5.1. Para uma discussão mais detalhada sobre a construção de intervalos de confiança bootstrap ver Efron & Tibshirani (1993) e Davison & Hinkley (1997).

Tabela 3.5.1. Quatro métodos de construção de intervalos de confiança para um parâmetro θ .

Método	Quantil $\alpha/2$	Corrigido sim
Padrão	$\hat{\theta} + \text{e.p.}(\hat{\theta})z_{\alpha/2}$	$\hat{\theta} \sim \mathcal{N}(\theta, \text{Var}(\hat{\theta}))$
Percentil	$\hat{P}^{-1}(\alpha/2)$	Existe uma transformação monótona $\hat{\rho} = g(\hat{\theta})$ tal que: $\hat{\rho} \sim \mathcal{N}(\rho, \tau^2)$
Bias-Corrected	$\hat{P}^{-1}(\Phi(z_{[\alpha/2]}))$	$(\hat{\rho} - \rho)/\tau \sim \mathcal{N}(-v_0, 1)$
BCa	$\hat{P}^{-1}(\Phi(z_{[\alpha/2]}^*))$	$\hat{\rho} - \rho \sim \mathcal{N}(-v_0(1 + a\rho), (1 + a\rho)^2)$

Nota: Os termos constantes τ , v_0 e a são definidos de acordo com o discutido nesta seção. A tabela foi tomada de Efron & Tibshirani (1986, Tabela 6).

Capítulo 4

Avaliação numérica

4.1 Introdução

Os métodos de simulação estocástica permitem reproduzir o comportamento de um sistema ou de um processo físico real que pode ser muito complexo. Esta reprodução requer a geração de uma história artificial do comportamento de tal sistema, como, por exemplo, a geração de variáveis pseudo-aleatórias com uma certa distribuição. É a partir da história gerada que se realizam inferências relativas às características do sistema real que é representado, além de ser possível descrever e analisar o comportamento do sistema real. Usando este artifício, consegue-se responder a algumas perguntas que auxiliam desenvolvimento de sistemas reais. Esta metodologia permite ao investigador tomar decisões sobre o desenvolvimento destes sistemas antes de suas implementações. Assim, a simulação, quando não resolve os problemas, ajuda a identificá-los, avaliando qualitativamente as soluções alternativas. Os modelos de simulação são necessários porque a experimentação de certos sistemas pode vir a ser muito custosa, por envolver tempos inaceitáveis de execução, ou simplesmente ser impossível de ser realizada. Definitivamente, a simulação é apropriada quando o modelo matemático que descreve o sistema é muito complicado, ou quando o sistema envolve variáveis de dinâmica que o tornam muito complexo de observar por um período razoável de tempo, ou simplesmente quando contém variáveis estocásticas que perturbam a dinâmica do sistema original. Neste contexto, aparece o método de Monte Carlo, o qual consiste na simulação de um sistema cuja solução aproximada é baseada na aleatoriedade da amostragem. O método de Monte Carlo associa a uma grande quantidade de métodos numéricos uma característica comum, que é a amostragem de variáveis aleatórias.

Nas últimas décadas ampliou-se a quantidade de problemas em que esta técnica pode ser aplicada, convertendo-se ela numa ferramenta muito útil em diversas áreas das ciências. Os métodos numéricos que são empregados na simulação estocástica utilizam um gerador de números pseudo-aleatórios uniformes e um algoritmo de geração que permita construir números pseudo-aleatórios de outras distribuições a partir dele. Um ponto que é muito importante neste caso é que o gerador de uniformes tenha boas propriedades, tais como um período elevado, ou seja, que a sucessão de números pseudo-aleatórios tarde muito a se repetir, como acontece no caso do gerador “multiply-with-carry” (GM), cujo período é de 2^{60} ; ver Marsaglia (1997). Como exemplo, consideremos a função $g(x)$ e suponhamos que desejamos conhecer $\theta = \int_0^1 g(x)dx$. Como podemos ver, este é um problema determinístico, onde

podemos considerar a seguinte solução: Seja $U \sim U(0, 1)$, então, $E[g(U)] = \int_0^1 g(u)f(u)du$, se tomamos $f(u) = 1$, $g(u) = g(x)$ e $du = dx$ temos que $E[g(U)] = \int_0^1 g(x)dx = \theta$, o qual pode ser aproximado por $\sum_{i=1}^k \frac{g(u_i)}{k} \approx E[g(U)] = \theta$, quando $k \rightarrow \infty$; onde $u_1, \dots, u_k$ são selecionados da variável U . Desta forma, podemos solucionar um problema determinístico usando o cálculo do valor esperado de uma amostra grande. O algoritmo de resolução aproximada deste problema através de simulação seria o seguinte:

- 1) Valores iniciais, $s_1 = 0$.
- 2) Gerar $u_i \sim U(0, 1)$.
- 3) Calcular $g(u_i)$.
- 4) Calcular $s_1 = s_1 + g(u_i)$.
- 5) Repetir o cálculo k -vezes.
- 6) Calcular $\hat{\theta} = \frac{s_1}{k}$

4.2 Detalhes metodológicos

Através de simulações de Monte Carlo, avaliamos o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta e suas versões corrigidas em amostras de tamanho finito e sob diferentes critérios. Todo o procedimento de cálculo foi programado na linguagem de programação `0x` (Cribari-Neto & Zarkos, 2003; Doornik, 2001). Consideramos o modelo de regressão beta com estrutura

$$g(\mu_t) = \beta_0 + \beta_1 x_t, \quad t = 1, \dots, n \quad (4.2.1)$$

onde g é a função de ligação logit; os valores da variável explicativa x_t , para cada observação são conhecidos. Os parâmetros β_0, β_1 e ϕ de nosso modelo são desconhecidos e definem o vetor $\underline{\theta} = (\beta_0, \beta_1, \phi)^\top$.

Nosso experimento Monte Carlo pode ser descrito da seguinte maneira. Definimos valores verdadeiros dos parâmetros, $\beta_0 = 1.5$, $\beta_1 = 1.8$, $\phi = 120$; os valores da covariável x_t são selecionados de uma variável aleatória distribuída exponencialmente, $\exp(3)$, com a qual formamos a matriz de regressores X . Ressaltamos que os valores de X permanecem constantes em todo o experimento. Para nosso experimento consideramos os tamanhos de amostra $n = 20, 40$ e 60 , o número de réplicas de Monte Carlo $R = 5000$ e número de réplicas bootstrap $B = 600$.

Para cada réplica de Monte Carlo geramos uma amostra aleatória da variável resposta $\underline{y} = (y_1, \dots, y_n)$ com $y_t, t = 1, \dots, n$ obtido através da parametrização (2.3.4) da distribuição

Beta, i.e., $y_t \sim \mathcal{B}(\mu_t, \phi)$, onde

$$\mu_t = \frac{\exp(\beta_0 + \beta_1 x_t)}{(1 + \exp(\beta_0 + \beta_1 x_t))}.$$

Dados os valores do vetor resposta e a matriz de regressores, avaliamos a regressão auxiliar

$$(X^\top X)^{-1} X^\top z,$$

onde z é o vetor da variável resposta transformado $z = (g(y_1), \dots, g(y_n))^\top$. Tomamos as estimativas dos parâmetros obtidas da regressão auxiliar como valores iniciais para o processo de maximização da função de log-verossimilhança (2.3.6). A função de log-verossimilhança é maximizada através do método BFGS discutido na Seção 1.7.1, onde os valores do ponto de máximo desta função são as estimativas $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\phi}$ de máxima verossimilhança dos parâmetros do modelo proposto. Através da formula (3.2.1.6) e dos valores das estimativas de máxima verossimilhança, avaliamos a estimativa do viés $\hat{B}(\hat{\theta})$ de segunda ordem dada pela formula (3.2.1.4) de Cox & Snell (1968). Conseqüentemente, calculamos o estimador corrigido de máxima verossimilhança $\tilde{\theta} = \hat{\theta} - \hat{B}(\hat{\theta})$. Calculado $\hat{B}(\hat{\theta})$ e avaliando a inversa da matriz de informação de Fisher no vetor $\tilde{\theta}$ encontramos a estimativa corrigida preventivamente θ^* , ao solucionar a equação não-linear (3.3.2), através da função SolveNLE, a qual está implementada na linguagem de programação `0x` e é utilizada no programa de simulação apresentado no apêndice B.

Adicionalmente, para cada réplica de Monte Carlo, i.e., para cada estimativa $\hat{\theta}$, geramos B réplicas bootstrap do modelo de forma paramétrica, i.e., geramos $\mathcal{Y}_1^*, \dots, \mathcal{Y}_B^*$ onde $\mathcal{Y}_r^* = (y_1^*, \dots, y_t^*)$, $r = 1, \dots, B$ é formado por $y_t^* \sim \mathcal{B}(\mu_t^*, \hat{\phi})$, $t = 1, \dots, n$, onde

$$\mu_t^* = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_t)}{(1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_t))}.$$

Com estas réplicas bootstrap determinamos a estimativa bootstrap do viés de $\hat{\theta}$ e calculamos assim a estimativa corrigida por bootstrap $\bar{\theta} = (\bar{\beta}_0, \bar{\beta}_1, \bar{\phi})$. Também, para cada réplica de Monte Carlo e cada uma das estimativas de máxima verossimilhança dos parâmetros do modelo foram consideradas estimativas intervalares de tipo assintótico ICA, IACC, IACP; bootstrap percentil ICP, ICC, ICCP e bootstrap BCa. Todos os intervalos foram construídos a duas caudas. Nesta simulação, cada tamanho de amostra envolve pelo menos $(5000 \times (n + 1))$ maximizações não-lineares + $(5000 \times (n + 1))$ resoluções de sistemas não-lineares mais $+(5000 \times 600 \times (n + 1))$ maximizações não-lineares. Quando $n = 60$, isto equivale a quase 10 milhões de maximizações não-lineares e 10 milhões de soluções de sistemas não-lineares. Portanto, nossa simulação é computacionalmente intensiva.

Para a análise de resultados da estimação pontual, foram calculados para cada tamanho de amostra: a média, estimativas do viés, viés relativo, variância, erro médio quadrático e a estimativa da raiz do erro médio quadrático das 5000 estimativas. Para a análise dos

resultados de estimação intervalar, são apresentados as probabilidades de cobertura nominais, a média dos 5000 limites superior e inferior, médias dos comprimentos, as probabilidades de cobertura observadas, as probabilidades observadas do limite inferior do intervalo ser maior que o valor do parâmetro e do limite superior do intervalo ser menor que o valor do parâmetro.

4.3 Resultados e discussão

4.3.1. Resultados da correção de viés

Nesta seção apresentamos os resultados de simulação referentes à correção de viés das estimativas de máxima verossimilhança dos parâmetros β_0 , β_1 e ϕ do modelo de regressão beta (4.2.1). Na tabela 4.3.1 se encontra a informação obtida da simulação para o parâmetro de intercepto β_0 do modelo. Podemos observar que, em módulo, as estimativas dos vieses dos estimadores corrigidos corretivo e de bootstrap, $\tilde{\beta}_0$ e $\bar{\beta}_0$, foram consideravelmente menores que o viés estimado da estimativa de máxima verossimilhança $\hat{\beta}_0$ para todos os tamanhos de amostra considerados, mostrando assim a efetividade da correção corretiva e por bootstrap. Neste caso, a correção por bootstrap é ligeiramente mais forte comparada com a correção analítica de Cox & Snell (1968). Por outro lado, ocorre o contrário com a estimativa corrigida preventiva β_0^* , que piora seu desempenho tanto com relação ao viés quanto aos valores da variância, considerando os diferentes tamanhos de amostra.

Percebemos que as estimativas do viés, viés relativo, variância, EQM e $\sqrt{\text{EQM}}$ de todas as estimativas do parâmetro β_0 diminuem aumentando o tamanho de amostra n , como era de se esperar, já que a variância assintótica como função de n é decrescente. Podemos verificar que as estimativas da variância, EQM e $\sqrt{\text{EQM}}$ dos estimadores corrigidos, incluindo o estimador corrigido preventivo, não apresentam grandes variações em relação às estimativas da variância, EQM e $\sqrt{\text{EQM}}$ do estimador de máxima verossimilhança $\hat{\beta}_0$; porque a correção por viés em princípio deve deslocar a distribuição do estimador de máxima verossimilhança na direção do verdadeiro valor do parâmetro sem afetar muito sua estrutura de variância. Por exemplo, para $n = 20$, o EQM das estimativas corrigidas de β_0 e da estimativa de máxima verossimilhança estão em torno do valor 0.0066, como pode-se observar na tabela 4.3.1. Esta mesma estabilidade permanece para os diferentes tamanhos de amostra considerados.

As variâncias assintóticas verdadeiras do estimador de máxima verossimilhança do parâmetro β_0 para $n = 20, 40$ e 60 são 0.0063906, 0.0033593 e 0.0023587, respectivamente. Assim, os valores das estimativas das variâncias dos estimadores corrigidos e de máxima verossimilhança estão próximos a estes valores, manifestando-se uma estabilidade nos valores das variâncias das estimativas corrigidas e não corrigida para os diferentes tamanhos de amostra utilizados. Ao comparar a média, viés e viés relativo de $\hat{\beta}_0$ com as médias, os vieses e vieses relativos dos estimadores corrigidos, percebemos que estas não diferem muito, o que faz pensar que o ganho de precisão dos estimadores corrigidos não é muito grande.

Na tabela 4.3.2. se encontram os resultados de simulação do parâmetro de inclinação β_1 do modelo (4.2.1). Observamos, que de forma análoga ao que acontece com o parâmetro de intercepto β_0 , as estimativas corrigidas corretiva e de bootstrap apresentam os melhores desempenhos em termos do módulo do viés e viés relativo, sendo que a correção analítica de Cox & Snell (1968) é ligeiramente mais forte comparada com a correção por bootstrap; o que difere do resultado obtido para o parâmetro de intercepto. Desta forma, $\tilde{\beta}_1$ melhora o ajuste do parâmetro de inclinação β_1 em termos do viés e viés relativo. Também este mesmo comportamento se apresenta para os diferentes tamanhos de amostra considerados. Portanto, a correção por viés do estimador de máxima verossimilhança é efetiva.

Tabela 4.3.1. Resultados de simulação para as estimativas de máxima verossimilhança do parâmetro $\beta_0 = 1.5$ e suas versões corrigidas.

n	Estimador	Média	Viés	Viés Rel.	Variância	EQM	$\sqrt{\text{EQM}}$
20	$\hat{\beta}_0$	1.50310	0.00310	0.00206	0.006654	0.006664	0.081634
	$\tilde{\beta}_0$	1.50198	0.00198	0.00132	0.006653	0.006657	0.081588
	β_0^*	1.50465	0.00465	0.00310	0.006664	0.006685	0.081763
	$\bar{\beta}_0$	1.50153	0.00153	0.00102	0.006653	0.006656	0.081583
40	$\hat{\beta}_0$	1.501338	0.001338	0.000892	0.003423	0.003425	0.058524
	$\tilde{\beta}_0$	1.500165	0.000165	0.000110	0.003422	0.003422	0.058494
	β_0^*	1.502564	0.002564	0.001709	0.003427	0.003433	0.058593
	$\bar{\beta}_0$	1.500458	0.000458	0.000306	0.003426	0.003426	0.058536
60	$\hat{\beta}_0$	1.500627	0.000627	0.000418	0.002387	0.002387	0.048859
	$\tilde{\beta}_0$	1.499857	-0.000143	-0.000095	0.002386	0.002386	0.048850
	β_0^*	1.501424	0.001424	0.000949	0.002388	0.002390	0.048884
	$\bar{\beta}_0$	1.500083	0.000083	0.000055	0.002390	0.002390	0.048890

Verificamos que o estimador corrigido preventivo β_1^* , de forma análoga ao estimador β_0^* piora seu desempenho em todas as medidas analisadas e considerando todos os tamanhos de amostra, o que difere de resultados de simulação dados em Cribari–Neto & Vasconcellos (2002) para o caso de variáveis aleatórias independentes e identicamente distribuídas. Assintoticamente as estimativas corrigidas têm a mesma estrutura de variância assintótica. De fato, percebemos que as estimativas da variância, EQM e $\sqrt{\text{EQM}}$ do estimador de máxima verossimilhança β_1 e suas versões corrigidas apresentam um equilíbrio em torno de um ponto. Por exemplo, se $n = 20$ as variâncias dos estimadores estão em torno do valor 0.98, para $n = 40$ estão em torno de 0.29 e no caso de $n = 60$ ao redor do ponto 0.19. Além disso, observamos que as estimativas do viés, variância, EQM e $\sqrt{\text{EQM}}$ diminuem com o aumento do tamanho de amostra n , pois à medida que n aumenta, as variâncias assintóticas dos estimadores analisados diminuem, fazendo com que o viés tenda para zero e as distribuições do

estimador de máxima verossimilhança e suas versões corrigidas tendam a concentrar-se ao redor do verdadeiro valor do parâmetro β_1 . As variâncias assintóticas verdadeiras da estimativa de máxima verossimilhança do parâmetro β_1 para $n = 20, 40$ e 60 são $0.94753, 0.28273$ e 0.19820 , respectivamente. De forma análoga à análise feita para β_0 , as variâncias das estimativas corrigidas e não corrigida para os diferentes tamanhos de amostra utilizados estão próximas. Ao compararmos a média, o viés e o viés relativo estimado de $\hat{\beta}_1$ com as médias, os vieses e vieses relativos das estimativas corrigidas, notamos que o ganho de precisão pela correção no estimador $\hat{\beta}_1$ não é muito grande, como ocorre de forma análoga com os estimadores corrigidos de β_0 .

Assim, em geral, percebemos que para este modelo de regressão beta, os estimadores corrigidos dos parâmetros da regressão não apresentam um ganho de ajuste no modelo muito grande em relação aos estimadores de máxima verossimilhança, em termos das medidas analisadas. Isto nos leva a pensar que as estimativas de máxima verossimilhança dos parâmetros de regressão têm boas propriedades amostrais, i.e., os valores estimados dos parâmetros de regressão se aproximam dos verdadeiros valores dos parâmetros da regressão apesar de os tamanhos de amostra considerados serem relativamente pequenos.

Tabela 4.3.2. Resultados de simulação para as estimativas de máxima verossimilhança do parâmetro $\beta_1 = 1.8$ e suas versões corrigidas.

n	Estimador	Média	Viés	Viés Rel.	Variância	EQM	$\sqrt{\text{EQM}}$
20	$\hat{\beta}_1$	1.790476	-0.009524	-0.005291	0.979154	0.979245	0.989568
	$\tilde{\beta}_1$	1.805542	0.005542	0.003079	0.980788	0.980819	0.990363
	β_1^*	1.779916	-0.020084	-0.011158	0.979601	0.980004	0.989952
	$\bar{\beta}_1$	1.786805	-0.013195	-0.007331	0.977723	0.977897	0.988887
40	$\hat{\beta}_1$	1.790608	-0.009392	-0.005218	0.291355	0.291444	0.539855
	$\tilde{\beta}_1$	1.804120	0.004120	0.002289	0.291850	0.291867	0.540248
	β_1^*	1.778862	-0.021138	-0.011743	0.291086	0.291533	0.539938
	$\bar{\beta}_1$	1.789677	-0.010323	-0.005735	0.291422	0.291529	0.539934
60	$\hat{\beta}_1$	1.790686	-0.009314	-0.005174	0.194875	0.194962	0.441545
	$\tilde{\beta}_1$	1.799312	-0.000688	-0.000382	0.195062	0.195063	0.441659
	β_1^*	1.782815	-0.017185	-0.009547	0.194753	0.195048	0.441642
	$\bar{\beta}_1$	1.789645	-0.010355	-0.005753	0.194921	0.195028	0.441620

Na tabela 4.3.3 apresentamos os resultados da simulação do parâmetro de precisão ϕ do modelo (4.2.1). Observamos na tabela 4.3.3. que as estimativas de máxima verossimilhança se encontram muito distantes do verdadeiro valor do parâmetro $\phi = 120$, i.e., apresentam vieses muito altos considerando os diferentes tamanhos de amostra. Por exemplo, para

$n = 20$, o viés estimado de $\hat{\phi}$ é 28.95312. Desta forma, a correção por viés neste caso se torna muito importante, já que se a estimativa do parâmetro ϕ é muito alta comparada com o verdadeiro valor do parâmetro, a variância da variável resposta pode estar sendo sub-estimada, o que em eventuais testes de hipótese assintóticos poderia levar a rejeitar os parâmetros da regressão quando na realidade eles são estatisticamente significativos. De forma análoga aos resultados de estimação dos parâmetros de regressão, a correção de viés por Cox & Snell (1968) e a correção por bootstrap melhoram o desempenho do estimador de máxima verossimilhança, sendo que, em módulo, o viés e viés relativo estimado da estimativa corrigida corretivamente apresentam desempenho superior que as das outras estimativas, considerando os diferentes tamanhos de amostra n . Contudo, a estimativa corrigida por bootstrap apresenta a menor estimação da variância para todos os tamanhos de amostra.

As variâncias assintóticas verdadeiras do estimador de máxima verossimilhança do parâmetro ϕ para $n = 20, 40$ e 60 são 1434.9, 717.71 e 478.48 respectivamente. De forma contrária ao que ocorre com os parâmetros estimados da regressão, as variâncias dos estimadores corrigidos e não corrigidos para os diferentes tamanhos de amostra utilizados são muito diferentes, sendo que os valores das estimativas da variância do estimador de bootstrap são os mais próximos às verdadeiras variâncias assintóticas considerando os diferentes tamanhos de amostra. Por outro lado, a estimativa corrigida preventiva ϕ^* piora seu desempenho tanto com relação ao viés quanto aos valores da variância, considerando todos os tamanhos de amostra. Como observamos, a estimativa $\hat{\phi}$ é maior que o verdadeiro valor do parâmetro ϕ e desta forma a variância da variável resposta diminui, concentrando muito mais a distribuição do vetor estimado de máxima verossimilhança $\hat{\theta}$ e fazendo assim com que a quantidade $\tilde{X}\tilde{\delta}$ avaliada em $\hat{\theta}$ não seja adequada para deslocar a função escore (3.3.4). Logo, para todos os parâmetros estimados do modelo, as estimativas corrigidas preventivas são piores em desempenho comparadas às respectivas estimativas de máxima verossimilhança.

Dai, poderíamos selecionar como estimador para o parâmetro ϕ o estimador corrigido por bootstrap incrementando levemente o viés ao ser comparado com o estimador corrigido por Cox & Snell (1968), mas facilitando seu cômputo. Além disso, percebemos uma forte diminuição da variância estimada do estimador $\hat{\phi}$ em relação as estimativas corrigidas $\tilde{\phi}$ e $\bar{\phi}$. Por exemplo, para $n = 20$ a variância estimada de $\hat{\phi}$ passa de 3103.57 a 1986.26 e 1758.78 para as estimativas corrigidas por Cox & Snell (1968) e por bootstrap respectivamente. Este mesmo fenômeno, é apresentado para o EQM e $\sqrt{\text{EQM}}$ considerando os diferentes tamanhos de amostra. Assim, para o parâmetro de ajuste ϕ recomendamos fortemente a correção do viés do estimador de máxima verossimilhança $\hat{\phi}$, em particular, a correção analítica de Cox & Snell (1968), ou a estimação corrigida por bootstrap que apresenta um desempenho similar.

4.3.2. Resultados dos intervalos de confiança

Nesta seção apresentamos os resultados de simulação referentes à construção dos in-

intervalos de confiança definidos ao longo deste trabalho, i.e., os intervalos assintóticos (ICA, IACC, IACP), percentil (ICP, ICC, ICCP) e BCa para os parâmetros do modelo de regressão beta (4.2.1). Foram avaliados um total de 243 intervalos com níveis nominais de cobertura $1 - \alpha$ iguais a 0.90, 0.95 e 0.99 para os tamanhos de amostra $n = 20, 40$ e 60. Todos os intervalos de confiança foram construídos de tal forma que contivessem o valor verdadeiro do parâmetro de interesse com probabilidade teoricamente $1 - \alpha$, com probabilidade $\alpha/2$ do limite inferior ser maior que o valor verdadeiro do parâmetro e com probabilidade $\alpha/2$ do limite superior ser menor do que este valor, para $0 < \alpha/2 < 0.5$.

Tabela 4.3.3. Resultados de simulação para as estimativas de máxima verossimilhança do parâmetro $\phi = 120$, e suas versões corrigidas.

n	Estimador	Média	Viés	Viés Rel.	Variância	EQM	$\sqrt{\text{EQM}}$
20	$\hat{\phi}$	148.9531	28.9531	0.24127	3103.577	3941.860	62.7842
	$\tilde{\phi}$	119.1283	-0.87162	-0.00726	1986.266	1987.025	44.5760
	ϕ^*	186.4067	66.4067	0.55338	4849.477	9259.328	96.2254
	$\bar{\phi}$	111.6602	-8.33972	-0.06949	1758.780	1828.332	42.7589
40	$\hat{\phi}$	134.0331	14.0331	0.11694	1096.540	1293.468	35.9648
	$\tilde{\phi}$	120.6078	0.60787	0.00506	888.1870	888.5571	29.8086
	ϕ^*	148.9867	28.9867	0.24155	1353.774	2194.007	46.8402
	$\bar{\phi}$	119.0727	-0.92725	-0.00772	866.9244	867.7842	29.4581
60	$\hat{\phi}$	128.8105	8.81059	0.07342	613.7862	691.4128	26.2947
	$\tilde{\phi}$	120.2088	0.20888	0.00174	534.6695	534.7131	23.1238
	ϕ^*	138.0418	18.0418	0.15034	704.6131	1030.120	32.0954
	$\bar{\phi}$	119.5717	-0.42822	-0.00356	530.2472	530.4306	23.0310

Para cada um dos métodos de estimação intervalar utilizados, a probabilidade de cobertura observada, denotada por “cobertura” é calculada a partir da frequência com que os 5000 intervalos construídos possuíram o verdadeiro valor do parâmetro. A probabilidade observada do limite inferior de confiança ser maior do que o verdadeiro valor do parâmetro, mostrada nas tabelas desta seção por “% Esquerdo”, foi estimada a partir da frequência com que os limites inferiores dos 5000 intervalos de confiança excederam este valor. Analogamente, a probabilidade observada do limite superior de confiança ser menor do que o verdadeiro valor do parâmetro, mostrada nas tabelas desta seção por “% Direito”, foi estimada a partir da frequência com que os limites superiores dos 5000 intervalos de confiança não excederam este valor.

As Figuras 4.3.2.1 até 4.3.2.9 contêm histogramas construídos a partir das 5000 estimativas de máxima verossimilhança dos parâmetros do modelo em análise, considerando os tamanhos de amostra $n = 20, 40$ e 60. Os segmentos de reta representam cada intervalo de confiança considerado, onde sua longitude simboliza o tamanho esperado. Os limites

superiores indicam o “% Esquerda” e os limites inferiores representam o “% Direita”. A maneira de exemplo, no apêndice B.4 apresentamos o programa para gerar o histograma com os intervalos de confiança do parâmetro β_0 quando $n = 40$. As tabelas 4.3.4.a e 4.3.4.b apresentam os intervalos de confiança para o parâmetro de intercepto $\beta_0 = 1.5$, e os intervalos assintóticos de tipo IACC e IACP para β_0 da forma (3.2.1.7) e (3.3.4) respectivamente. Notamos que os limites inferior e superior médios de cada intervalo de confiança considerado são muito parecidos entre eles. Por exemplo, para um nível nominal de confiança de 95% os limites inferiores do IACC e ICCS são 1.3396 e 1.3395 respectivamente. Observamos que, em geral, para os três níveis nominais de confiança e os diferentes tamanhos de amostra, a média dos comprimentos “Tamanho” dos 5000 intervalos de confiança do tipo ICCP é menor que as respectivas médias dos outros intervalos, seguido do intervalo de confiança IACP. Em geral, para os diferentes níveis de confiança, o intervalo IACC apresenta a melhor cobertura seguido do intervalo ICCS, considerando todos os tamanhos de amostra.

As Figuras 4.3.2.1, 4.3.2.2 e 4.3.2.3 mostram que todos os intervalos de confiança considerados para β_0 são aproximadamente simétricos ao redor de β_0 e balanceados, i.e., as probabilidades observadas “% Esquerdo” e “% Direito” coincidem aproximadamente nas caudas da densidade do estimador de máxima verossimilhança de β_0 considerando os diferentes níveis de confiança e tamanhos de amostra; isto é verificado nos valores da tabela 4.3.4.b. Também observamos que, quando o tamanho de amostra aumenta, a simetria torna-se mais forte para todos os intervalos de confiança e os valores percentuais “% Direito” e “% Esquerdo” tendem a concentrar-se em torno de uma quantidade. Por exemplo, para um nível nominal de confiança de 90%, quando o tamanho de amostra aumenta, os valores de “% Direito” e “% Esquerdo” se aproximam em média ao valor 5.6. Além disso, o aumento do tamanho de amostra melhora o desempenho de todos os intervalos de confiança, i.e., as coberturas aumentam, e dado que a variância diminui, os comprimentos são menores.

Um fato interessante ao observar as tabelas 4.3.4.a e 4.3.4.b é que, em geral, para os diferentes tamanhos de amostra considerados, os intervalos de confiança ICA, ICP e BCa apresentam desempenhos similares em “Tamanho”, “cobertura” e as porcentagens “% Esquerdo” e “% Direito”, o que nos leva a pensar que a distribuição empírica de $\hat{\beta}_0$ e sua distribuição assintótica são muito parecidas, apresentando só pequenas diferenças nos percentis uma da outra. De fato, como foi discutido na seção anterior, este resultado confirma nossa suspeita de que o estimador $\hat{\beta}_0$ possui boas propriedades assintóticas, apesar de que os tamanhos de amostra considerados são pequenos.

Este mesmo fenômeno ocorre também para os intervalos de confiança IACC, ICCS, IACP e ICCP já que, em princípio, as distribuições empíricas de $\tilde{\beta}_0$ e β_0^* são só deslocções da distribuição empírica de $\hat{\beta}_0$. A tabela 4.3.4.c apresenta os intervalos de confiança assintóticos para o parâmetro de intercepto $\beta_0 = 1.5$, sendo que os intervalos assintóticos de tipo IACC e IACP para β_0 são da forma (3.2.1.9) e (3.3.6) respectivamente. Percebemos que, de forma análoga ao analisado para os intervalos de confiança anteriores, os intervalos de confiança assintóticos para β_0 têm comportamento similar aos dos intervalos IACC e IACP para β_0 da

Tabela 4.3.4.a. Comportamento das estimativas intervalares do parâmetro β_0 .

α	Medidas	Inferior			Superior			Tamanho		
		$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	Intervalo									
	ICA	1.3814	1.4079	1.4225	1.6285	1.5928	1.5790	0.2473	0.1847	0.1567
	IACC	1.3660	1.4018	1.4191	1.6418	1.5965	1.5809	0.2756	0.1941	0.1616
	IACP	1.3959	1.4138	1.4259	1.6172	1.5894	1.5772	0.2212	0.1752	0.1515
	ICP	1.3809	1.4097	1.4226	1.6286	1.5942	1.5790	0.2477	0.1845	0.1563
	ICCC	1.3661	1.4037	1.4193	1.6286	1.5942	1.5790	0.2625	0.1905	0.1597
	ICCP	1.3951	1.4154	1.4260	1.6168	1.5907	1.5772	0.2216	0.1752	0.1511
5%	BCa	1.3787	1.4081	1.4215	1.6283	1.5939	1.5790	0.2495	0.1858	0.1574
	ICA	1.3577	1.3902	1.4075	1.6522	1.6105	1.5941	0.2947	0.2209	0.1865
	IACC	1.3396	1.3832	1.4036	1.6682	1.6152	1.5964	0.3289	0.2321	0.1927
	IACP	1.3747	1.3970	1.4114	1.6384	1.6062	1.5917	0.2632	0.2094	0.1802
	ICP	1.3573	1.3919	1.4074	1.6522	1.6116	1.5938	0.2948	0.2197	0.1864
	ICCC	1.3395	1.3849	1.4035	1.6690	1.6165	1.5964	0.3295	0.2316	0.1929
	ICCP	1.3738	1.3985	1.4114	1.6379	1.6073	1.5915	0.2640	0.2088	0.1801
1%	BCa	1.3550	1.3900	1.4060	1.6530	1.6120	1.5943	0.2980	0.2220	0.1883
	ICA	1.3115	1.3556	1.3782	1.6985	1.6451	1.6233	0.3177	0.2379	0.2005
	IACC	1.2879	1.3467	1.3732	1.7198	1.6516	1.6267	0.4314	0.3040	0.2538
	IACP	1.3332	1.3556	1.3831	1.6798	1.6391	1.6201	0.3468	0.2746	0.2365
	ICP	1.3091	1.3551	1.3760	1.6964	1.6441	1.6212	0.3194	0.2390	0.2029
	ICCC	1.2859	1.3461	1.3712	1.7185	1.6504	1.6249	0.3427	0.2480	0.2077
	ICCP	1.3304	1.3633	1.3814	1.6772	1.6381	1.6183	0.2863	0.2273	0.1958
1%	BCa	1.3023	1.3499	1.3713	1.7020	1.6479	1.6243	0.3263	0.2442	0.2077

Tabela 4.3.4.b. Comportamento das estimativas intervalares do parâmetro β_0 .

α	Medidas	Cobertura			% Esquerdo			% Direito		
		$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	Intervalo									
	ICA	86.38	87.92	88.76	7.68	5.98	5.84	5.94	6.10	5.40
	IACC	90.26	89.76	89.90	5.36	4.82	5.06	4.38	5.42	5.04
	IACP	82.54	86.20	87.46	10.04	7.26	6.74	7.42	6.54	5.80
	ICP	86.00	87.68	88.62	7.64	6.56	5.74	6.36	5.76	5.64
	ICCC	90.06	89.12	89.80	5.40	5.68	4.86	4.54	5.20	5.34
	ICCP	81.46	85.40	87.60	10.24	8.06	6.40	8.30	6.54	6.00
5%	BCa	86.62	87.86	88.84	6.94	6.26	5.52	6.44	5.88	5.64
	ICA	92.26	93.16	93.92	4.22	3.3	3.08	3.52	3.54	3.01
	IACC	94.82	94.44	94.76	2.68	2.64	2.50	2.51	2.92	2.74
	IACP	88.74	91.82	93.02	6.50	4.06	3.62	4.76	4.12	3.36
	ICP	91.98	92.96	93.74	4.42	3.62	3.14	3.60	3.42	3.12
	ICCC	90.06	89.12	89.80	3.14	2.76	2.68	2.60	2.74	2.88
	ICCP	81.46	85.40	87.60	6.60	4.54	3.58	4.94	3.92	3.58
1%	BCa	91.78	93.36	93.88	4.50	3.36	3.04	3.72	3.28	3.08
	ICA	97.50	98.40	98.54	1.34	0.88	0.62	1.16	0.72	0.84
	IACC	98.76	98.78	98.74	0.64	0.58	0.52	0.60	0.64	0.74
	IACP	95.82	97.98	98.18	2.38	1.12	0.84	1.81	0.90	0.98
	ICP	97.32	98.04	98.36	1.38	0.96	0.74	1.30	1.00	0.90
	ICCC	98.58	98.62	98.68	0.76	0.64	0.56	0.66	0.74	0.76
	ICCP	95.40	97.62	97.96	2.44	1.12	0.92	2.16	1.26	1.12
1%	BCa	97.44	98.24	98.46	1.32	0.84	0.68	1.24	0.92	0.86

forma (3.2.1.7) e (3.3.4). Ao compararmos os intervalos IACC e IACP nas versões (3.2.1.9) e (3.3.6) com respeito aos outros intervalos, notamos que o intervalo ICCP apresenta o menor comprimento e o IACC a melhor cobertura para os diferentes níveis e os diferentes tamanhos de amostra considerados. Este resultado é o mesmo quando analisamos os intervalos IACC e IACP nas versões (3.2.1.7) e (3.3.4) respectivamente. O interessante destes novos intervalos é que eles apresentam os mesmos comprimentos que o intervalo de confiança IAC, diferindo somente na sua localização para os diferentes tamanhos de amostra considerados, o que era de se esperar, já que, para sua construção, eles utilizam os mesmos erros padrões do estimador de máxima verossimilhança.

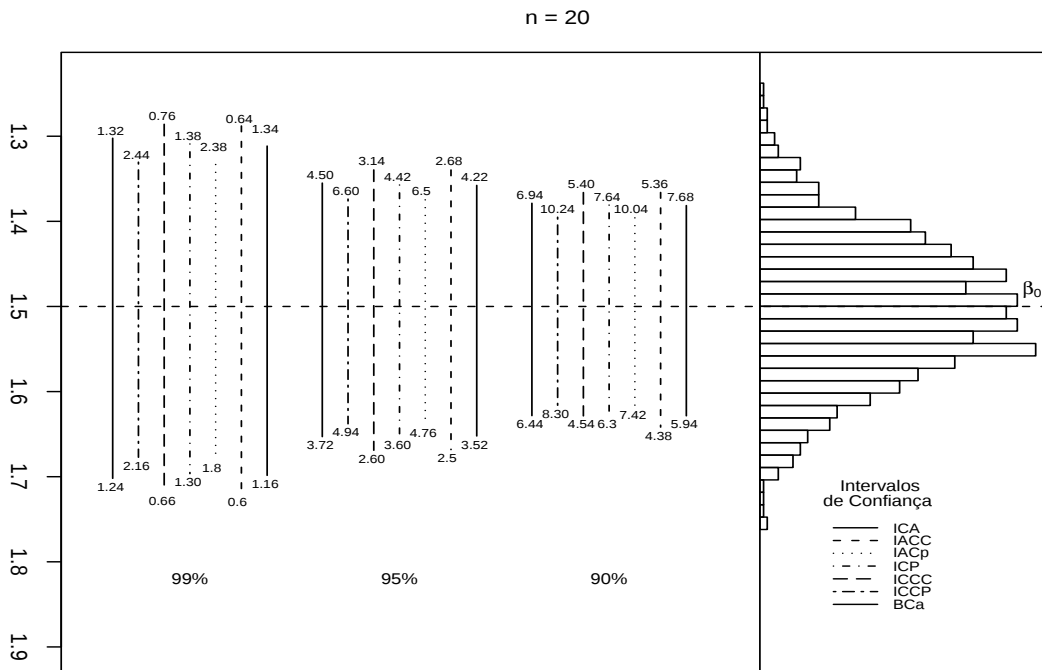


Figura 4.3.2.1: Estimação intervalar para β_0 com $n = 20$.

As tabelas 4.3.5.a e 4.3.5.b apresentam os intervalos de confiança para o parâmetro de inclinação $\beta_1 = 1.8$; os intervalos assintóticos de tipo IACC e IACP para β_1 são da forma (3.2.1.7) e (3.3.4) respectivamente. Percebemos que, em geral, para os três níveis nominais de confiança e os diferentes tamanhos de amostra, a média dos comprimentos “Tamanho” dos 5000 intervalos de confiança do tipo ICCP é menor que as respectivas médias dos outros intervalos, seguida da média do intervalo de confiança IACP; isto acontece de forma análoga com os intervalos de confiança para β_0 .

Em geral, para os diferentes níveis de cobertura, o intervalo IACC apresenta a melhor cobertura, considerando todos os tamanhos de amostra. As Figuras 4.3.2.4, 4.3.2.5 e 4.3.2.6 mostram que todos os intervalos de confiança considerados para os diferentes níveis de cobertura para β_1 são aproximadamente simétricos ao redor de β_1 , mas um pouco não-balanceados,

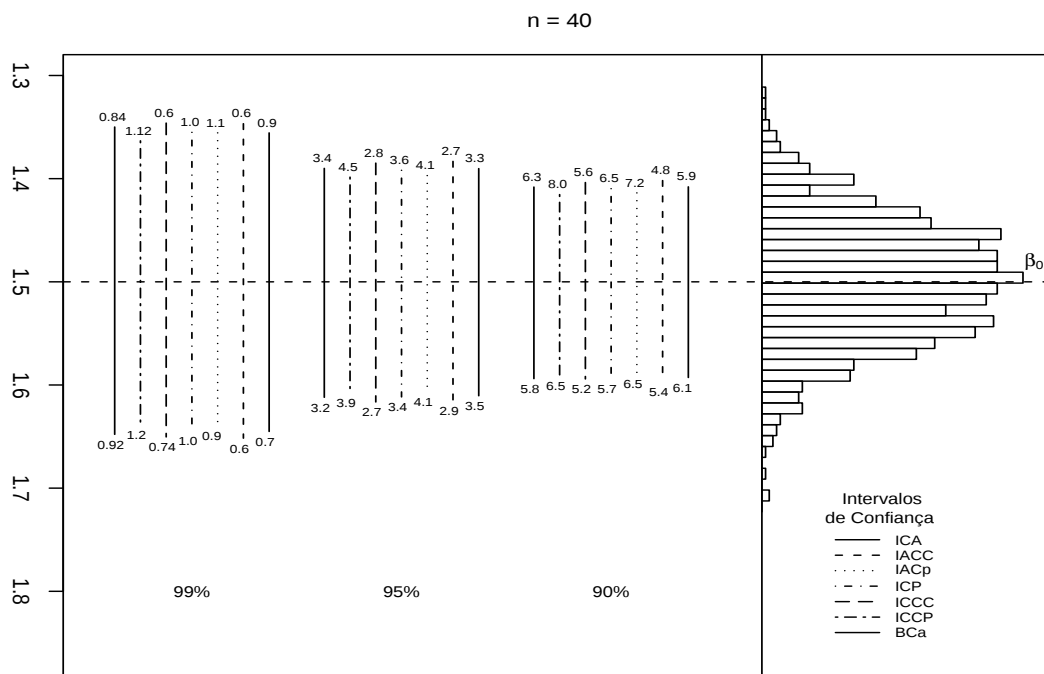


Figura 4.3.2.2: Estimação intervalar para β_0 com $n = 40$.

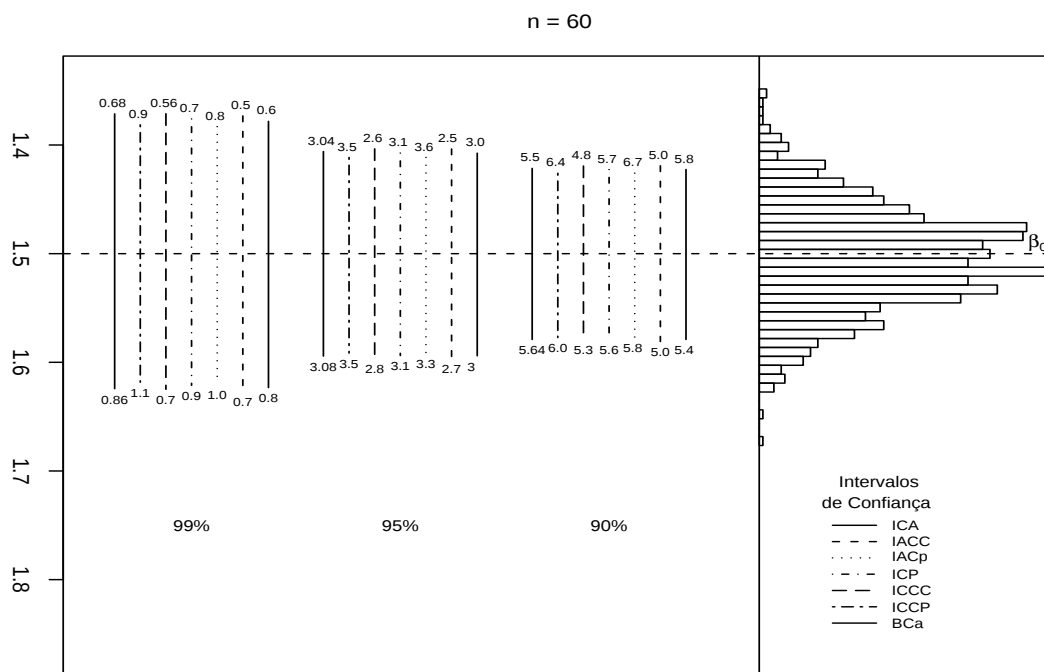


Figura 4.3.2.3: Estimação intervalar para β_0 com $n = 60$.

Tabela 4.3.4.c. Comportamento das estimativas intervalares assintóticas do parâmetro β_0 .

Medidas		Inferior			Superior			Tamanho		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	1.3794	1.4091	1.4224	1.6267	1.5934	1.5787	0.2473	0.1847	0.1567
	IACC	1.3783	1.4080	1.4217	1.6256	1.5923	1.5780	0.2473	0.1847	0.1567
	IACP	1.3810	1.4104	1.4232	1.6282	1.5947	1.5795	0.2473	0.1847	0.1567
5%	ICA	1.3557	1.3915	1.4075	1.6504	1.6111	1.5937	0.2947	0.2209	0.1865
	IACC	1.3546	1.3903	1.4067	1.6493	1.6099	1.5929	0.2947	0.2209	0.1865
	IACP	1.3573	1.3927	1.4083	1.6519	1.6123	1.5945	0.2947	0.2209	0.1865
1%	ICA	1.3094	1.3570	1.3782	1.6967	1.6456	1.6229	0.3177	0.2379	0.2005
	IACC	1.3083	1.3558	1.3774	1.6956	1.6444	1.6222	0.3877	0.2879	0.2005
	IACP	1.3110	1.3582	1.3790	1.6982	1.6468	1.6237	0.3877	0.2879	0.2005
Medidas		Cobertura			% Esquerdo			% Direito		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	86.12	87.68	88.86	7.34	6.44	5.46	6.54	5.88	5.68
	IACC	86.06	87.80	88.84	7.24	6.10	5.26	6.70	6.10	5.90
	IACP	86.20	87.76	88.80	7.50	6.66	5.62	6.30	5.58	5.58
5%	ICA	92.04	93.06	93.86	4.18	3.56	3.02	3.78	3.38	3.12
	IACC	92.08	93.24	93.88	4.10	3.26	2.92	3.82	3.50	3.20
	IACP	91.90	93.02	93.86	4.44	3.74	3.12	3.66	3.24	3.02
1%	ICA	97.24	98.36	98.48	1.50	0.9	0.82	1.26	0.74	0.70
	IACC	97.30	98.32	98.56	1.40	0.9	0.72	1.30	0.78	0.72
	IACP	97.16	98.30	98.48	1.60	0.96	0.86	1.24	0.74	0.66

Nota: Neste caso os intervalos assintóticos IACC e IACP são construídos com o mesmo erro padrão do ICA.

já que as probabilidades observadas “% Direito” são levemente superiores às probabilidades observadas “% Esquerdo”, sendo que este comportamento é reduzido com o aumento do tamanho de amostra. Analisando a tabela 4.3.5.a; observamos que quando o tamanho de amostra aumenta, a simetria torna-se mais forte para todos os intervalos de confiança e os valores percentuais “% Direito” e “% Esquerdo” tendem a concentrar-se em torno de uma mesma quantidade; de forma análoga ao que acontece para os intervalos de confiança de β_0 . Além disso, o aumento do tamanho de amostra melhora o desempenho de todos os intervalos de confiança, dado que as coberturas aumentam e os comprimentos diminuem.

Como ocorre também com os intervalos de confiança para β_0 , constatamos ao observar as tabelas 4.3.5.a e 4.3.5.b, para os diferentes tamanhos de amostra e níveis de cobertura considerados, que os intervalos de confiança ICA, ICP e BCa apresentam desempenhos similares tanto em comprimento, cobertura e porcentagens “% Esquerdo” e “% Direito”. Esta situação se manifesta para os intervalos de confiança IACC com ICCP, e IACP com ICCP, pois que, em princípio, as distribuições empíricas de $\tilde{\beta}_1$ e β_1^* são só deslocamentos da distribuição empírica de $\tilde{\beta}_1$. A tabela 4.3.5.c apresenta os intervalos de confiança assintóticos para o parâmetro de intercepto $\beta_1 = 1.8$, sendo que os intervalos assintóticos de tipo IACC e IACP para β_1 são da forma (3.2.1.9) e (3.3.6) respectivamente. De forma análoga ao discutido para os intervalos de confiança de β_0 , os intervalos de confiança assintóticos da forma (3.2.1.9) e (3.3.6) de β_1 têm comportamento similar aos dos intervalos IACC e IACP para β_1 da forma (3.2.1.7) e (3.3.4).

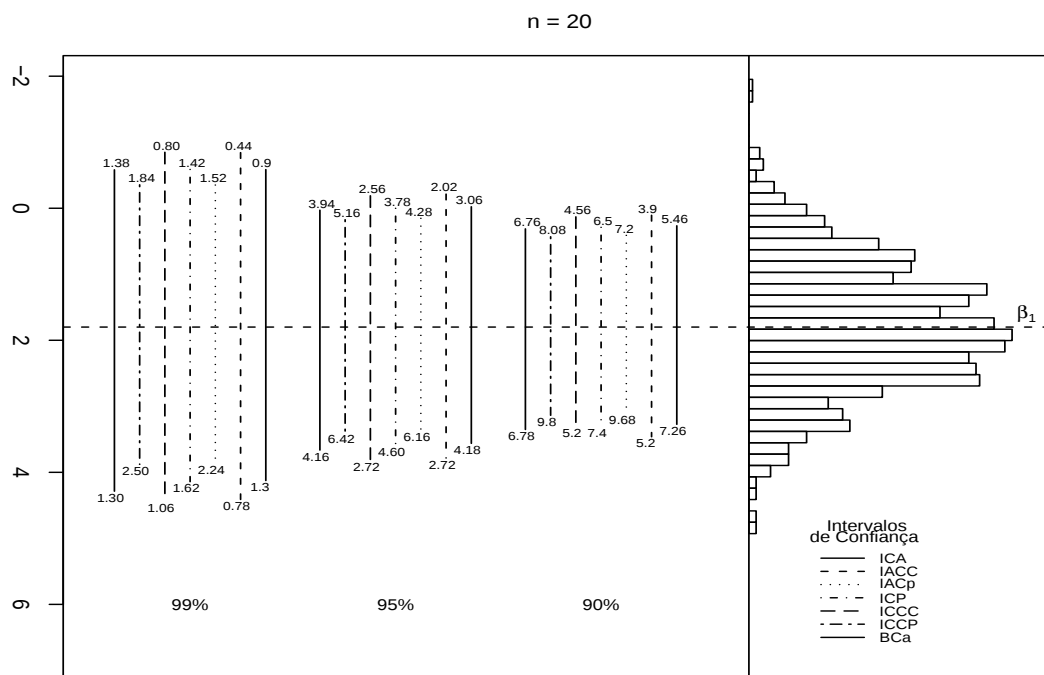


Figura 4.3.2.4: Estimação intervalar para β_1 com $n = 20$.

Tabela 4.3.5.a. Comportamento das estimativas intervalares do parâmetro β_1 .

Medidas		Inferior			Superior			Tamanho		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	0.2650	0.9579	1.0872	3.2734	2.6549	2.5221	3.0083	1.6970	1.4349
	IACC	0.1066	0.9263	1.0711	3.4618	2.7137	2.5555	3.3551	1.7874	1.4844
	IACP	0.4109	0.9890	1.1032	3.1064	2.6001	2.4902	2.6954	1.6110	1.3869
	ICP	0.2891	0.9485	1.0774	3.3063	2.6430	2.5108	3.0172	1.6945	1.4334
	ICC	0.1299	0.9169	1.0607	3.3063	2.6430	2.5108	3.1764	1.7260	1.4501
	ICCP	0.4313	0.9796	1.0926	3.1371	2.5884	2.4790	2.7058	1.6087	1.3864
	BCa	0.3117	0.9644	1.0858	3.3516	2.6724	2.5296	3.0399	1.7079	1.4438
5%	ICA	-0.0231	0.7953	0.9497	3.5616	2.8175	2.6595	3.5847	2.0221	1.7098
	IACC	-0.2147	0.7551	0.9289	3.7832	2.8850	2.6977	3.9979	2.1298	1.7688
	IACP	0.1527	0.8347	0.9704	3.3646	2.7544	2.6230	3.2118	1.9196	1.6526
	ICP	0.0009	0.7882	0.9410	3.5967	2.8085	2.6491	3.5957	2.0203	1.7081
	ICC	-0.1894	0.7489	0.9192	3.8234	2.8771	2.6871	4.0128	2.1282	1.7678
	ICCP	0.1735	0.8282	0.9608	3.3983	2.7444	2.6126	3.2247	1.9162	1.6518
	BCa	0.0288	0.8069	0.9504	3.6631	2.8492	2.6759	3.6342	2.0423	1.7254
1%	ICA	-0.5863	0.4776	0.6811	4.1248	3.1352	2.9282	3.8597	2.1773	1.8409
	IACC	-0.8428	0.4205	0.6510	4.4113	3.2196	2.9756	5.2541	2.7991	2.3246
	IACP	-0.3518	0.5331	0.7107	3.8692	3.0560	2.8827	4.2210	2.5228	2.1719
	ICP	-0.5877	0.4626	0.6638	4.1421	3.1175	2.9085	3.8941	2.1804	1.8470
	ICC	-0.8485	0.4100	0.6321	4.4333	3.2038	2.9542	4.1549	2.2329	1.8787
	ICCP	-0.3555	0.5170	0.6923	3.8798	3.0403	2.8622	3.4926	2.0714	1.7867
	BCa	-0.5825	0.4665	0.6587	4.2865	3.1990	2.9665	3.8888	2.1765	1.8521

Tabela 4.3.5.b. Comportamento das estimativas intervalares do parâmetro β_1 .

α	Medidas	Cobertura			% Esquerdo			% Direito		
		$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	Intervalo									
	ICA	87.28	88.68	88.86	5.46	5.06	5.44	7.26	6.26	5.70
	IACC	90.90	90.62	89.88	3.90	4.34	5.18	5.20	5.04	4.94
	IACP	83.14	86.74	87.58	7.18	5.72	5.88	9.68	7.54	6.54
	ICP	86.06	87.24	89.20	6.50	6.02	4.94	7.44	6.74	5.86
	ICCC	90.22	89.12	90.46	4.56	5.28	4.52	5.22	5.60	5.02
	ICCP	82.10	85.42	87.98	8.08	6.62	5.10	9.82	7.96	6.92
5%	BCa	86.46	87.24	89.58	6.76	6.30	5.08	6.78	6.46	5.34
	ICA	92.76	94.16	94.12	3.06	2.54	2.82	4.18	3.31	3.06
	IACC	95.26	95.18	94.70	2.02	2.22	2.64	2.72	2.60	2.66
	IACP	89.56	92.90	93.18	4.28	2.92	3.21	6.16	4.18	3.62
	ICP	91.62	93.16	94.54	3.78	3.02	2.60	4.60	3.82	2.86
	ICCC	90.22	89.12	90.46	2.56	2.52	2.52	2.72	2.84	2.30
	ICCP	82.10	85.42	87.98	5.16	3.70	3.08	6.42	4.74	3.50
1%	BCa	91.90	93.28	94.34	3.94	3.34	2.80	4.16	3.38	2.86
	ICA	97.8	98.54	98.48	0.90	0.60	0.68	1.30	0.86	0.84
	IACC	98.78	98.86	98.80	0.44	0.46	0.56	0.78	0.68	0.64
	IACP	96.24	97.98	98.02	1.52	0.76	0.86	2.24	1.26	1.12
	ICP	96.96	98.10	98.62	1.42	0.72	0.68	1.62	1.18	0.70
	ICCC	98.14	98.56	98.82	0.80	0.60	0.54	1.06	0.84	0.64
	ICCP	95.66	97.38	98.36	1.84	1.02	0.70	2.50	1.60	0.94
	BCa	97.32	98.30	98.78	1.38	0.70	0.70	1.30	1.00	0.52

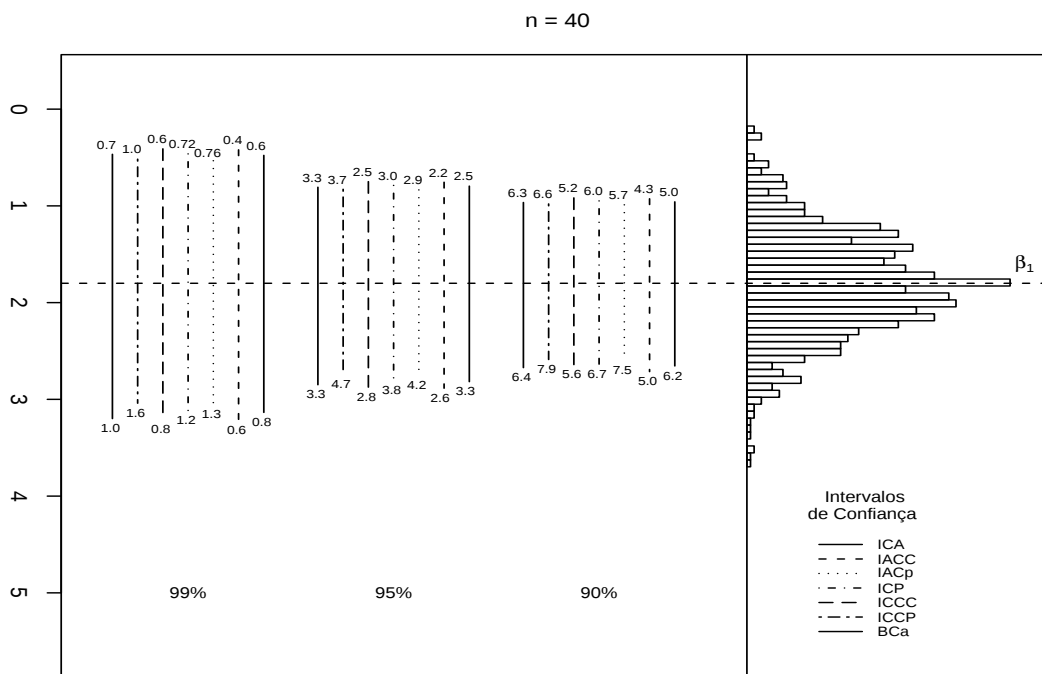


Figura 4.3.2.5: Estimação intervalar para β_1 com $n = 40$.

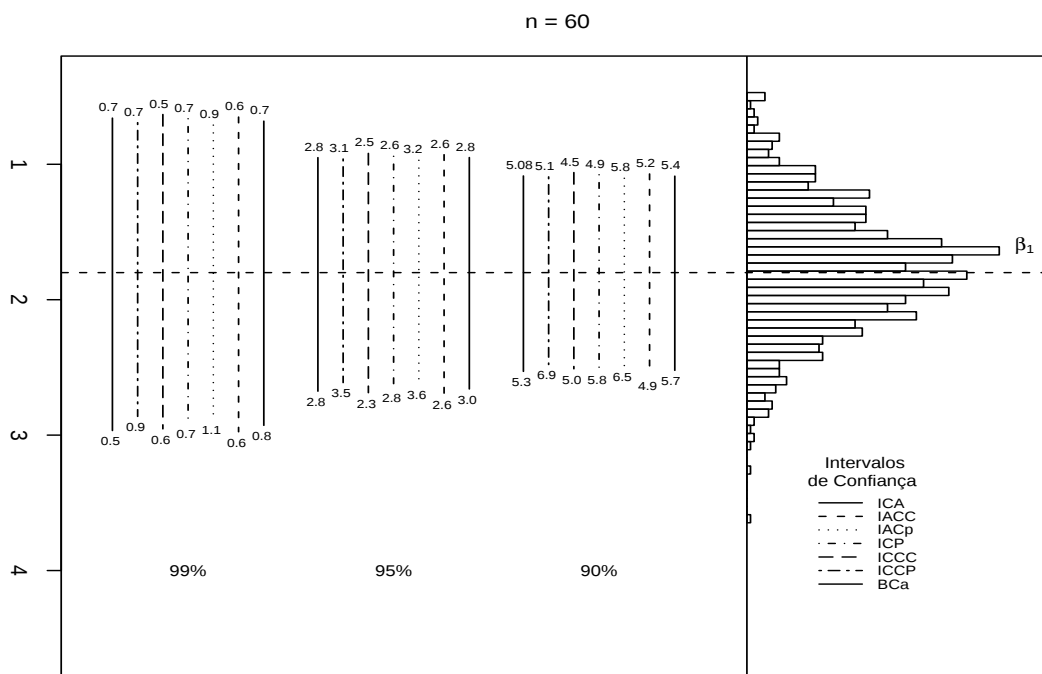


Figura 4.3.2.6: Estimação intervalar para β_1 com $n = 60$.

Ao compararmos os intervalos IACC e IACP nas versões (3.2.1.9) e (3.3.6) com respeito aos outros intervalos, notamos que o intervalo ICCP apresenta o menor comprimento e o IACC a melhor cobertura, para os diferentes níveis e os diferentes tamanhos de amostra considerados. Este resultado é análogo quando analisamos IACC e IACP nas versões (3.2.1.7) e (3.3.4) respectivamente. Novamente, estes intervalos apresentam o mesmo comprimento que o intervalo de confiança IAC, diferindo somente na sua localização para os diferentes tamanhos de amostra considerados. Observamos que, em geral, para os parâmetros de regressão, os intervalos de tipo percentil se comportam melhor em desempenho que os intervalos de tipo BCa para os diferentes níveis de confiança e os diferentes tamanhos de amostra, o que pode acontecer pelo fato do número B de réplicas bootstrap não ser muito grande. Segundo o analisado nas tabelas 4.3.4.a até 4.3.5.b.; a boa cobertura dos intervalos IACC deve-se a que os erros padrões são maiores comparados aos obtidos para os estimadores corrigido preventivo e de bootstrap.

Analisemos detalhadamente a informação de Fisher para os parâmetros da regressão. Por exemplo, para o r -ésimo parâmetro de regressão, β_r podemos deduzir que a informação

$$K(\boldsymbol{\theta})_{rr} = \phi^2 \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} x_{tr} \right)^2,$$

é uma função crescente de ϕ . Desta forma, se a estimativa corrigida preventiva de ϕ for maior, comparada com as estimativas corrigida corretiva e de máxima verossimilhança, sua variância é menor, produzindo assim erros padrões menores que os das outras estimativas, o que implica, que seus respectivos intervalos de confiança assintóticos sejam de menor comprimento.

De forma geral, notamos que as distribuições empírica e assintótica dos estimadores de máxima verossimilhança dos parâmetros da regressão se comportam bem para tamanhos de amostra moderados, fornecendo intervalos de confiança aproximadamente simétricos e balanceados para os diferentes tamanhos de amostra e os diferentes níveis de confiança. Observamos que os intervalos de confiança assintóticos IACC apresentam boas coberturas, já que para os diferentes tamanhos de amostra e os níveis de confiança considerados, as probabilidades observadas de cobertura estão muito próximas dos verdadeiros níveis nominais de cobertura. Adicionalmente, os intervalos de confiança de tipo percentil ICCP apresentam os menores comprimentos para os diferentes níveis de confiança e tamanhos de amostra considerados.

Para finalizar, as tabelas 4.3.6.a e 4.3.6.b apresentam os intervalos de confiança para o parâmetro de precisão $\phi = 120$; os intervalos assintóticos de tipo IACC e IACP para ϕ são da forma (3.2.1.8) e (3.3.5) respectivamente. Observamos na tabela 4.3.6.a que todos os intervalos de confiança analisados não abrangem valores fora do espaço paramétrico. Percebemos que para quase todos os níveis nominais de cobertura e os diferentes tamanhos de amostra, a média dos comprimentos “Tamanho” dos 5000 intervalos de confiança do tipo IACC é menor que as respectivas médias dos outros intervalos, seguida da dos intervalos de

Tabela 4.3.5.c. Comportamento das estimativas intervalares assintóticas do parâmetro β_1 .

Medidas		Inferior			Superior			Tamanho		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	0.2841	0.9453	1.0745	3.2968	2.6358	2.5067	3.0083	1.697	1.4321
	IACC	0.2991	0.9589	1.0832	3.3118	2.6493	2.5154	3.0083	1.697	1.4321
	IACP	0.2735	0.9336	1.0667	3.2862	2.6240	2.4989	3.0083	1.697	1.4321
5%	ICA	-0.0044	0.7834	0.9374	3.5854	2.7977	2.6439	3.5847	2.0221	1.7098
	IACC	0.0106	0.7969	0.9460	3.6004	2.8112	2.6525	3.5847	2.0221	1.7098
	IACP	-0.0150	0.7717	0.9295	3.5748	2.7859	2.6360	3.5847	2.0221	1.7098
1%	ICA	-0.5684	0.4670	0.6692	4.1494	3.1142	2.9120	3.8597	2.1773	1.8409
	IACC	-0.5533	0.4805	0.6779	4.1644	3.1277	2.9207	3.8597	2.1773	1.8409
	IACP	-0.5790	0.4552	0.6614	4.1388	3.1024	2.9042	3.8597	2.1773	1.8409
Medidas		Cobertura			% Esquerdo			% Direito		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	86.48	87.28	89.16	6.12	5.82	4.78	7.40	6.90	6.06
	IACC	86.36	87.14	89.42	6.36	6.20	4.90	7.28	6.66	5.68
	IACP	86.40	87.32	89.10	6.06	5.64	4.60	7.54	7.04	6.30
5%	ICA	92.16	93.50	94.46	3.56	2.70	2.66	4.28	3.80	2.88
	IACC	92.10	93.42	94.50	3.66	2.94	2.76	4.24	3.64	2.74
	IACP	92.06	93.48	94.52	3.54	2.58	2.48	4.40	3.94	3.00
1%	ICA	97.12	98.28	98.72	1.36	0.62	0.60	1.52	1.10	0.68
	IACC	97.14	98.34	98.74	1.38	0.62	0.64	1.48	1.04	0.62
	IACP	97.10	98.26	98.70	1.34	0.60	0.58	1.56	1.14	0.72

Nota: Neste caso os intervalos assintóticos IACC e IACP são construídos com o mesmo erro padrão do ICA.

confiança BCa e ICP; isto acontece de forma análoga para os intervalos de confiança de β_0 e β_1 . Observamos uma grande diferença dentre os valores dos limites inferiores e superiores de cada intervalo de confiança considerado, ao contrário do que sucede com os intervalos de confiança para os parâmetros de regressão. Em termos de cobertura, para um nível de confiança de 99%, o intervalo BCa apresenta os valores mais próximos às verdadeiras probabilidades de cobertura quando $n = 20$ e 40 , seguido em desempenho pelo intervalo assintótico ICA. Quando o tamanho de amostra aumenta, para os níveis de confiança de 95% e 99%, os intervalos de confiança assintóticos apresentam melhores coberturas para os diferentes tamanhos de amostra seguidos em desempenho pelo intervalo IACP. Aumentando o tamanho de amostra os intervalos de tipo percentil melhoram seu desempenho em termos de cobertura, mas os de tipo BCa apresentam menores comprimentos, sendo que eles, não superam o desempenho dos intervalos de tipo assintótico. As Figuras 4.3.2.7, 4.3.2.8 e 4.3.2.9 mostram que todos os intervalos de confiança para ϕ considerando os diferentes níveis nominais de cobertura são bastante assimétricos ao redor de ϕ e apresentam um forte não-balanceamento, dado que as probabilidades observadas “% Esquerdo” são acentuadamente superiores às probabilidades observadas “% Direito”, o que pode estar acontecendo pela positividade do parâmetro ϕ , que faz que a distribuição de $\hat{\phi}$ seja assimétrica.

Analisando a tabela 4.3.6.b.; observamos que, aumentando o tamanho de amostra, a simetria torna-se mais evidente para todos os intervalos de confiança. Porém, não podemos garantir um total balanceamento dos intervalos de confiança, pois os dados revelam que os valores percentuais “% Direito” e “% Esquerdo”, permanecem consideravelmente diferentes. Podemos afirmar observando as Figuras 4.3.2.7, 4.3.2.8 e 4.3.2.9 que os intervalos de tipo BCa são aproximadamente balanceados para os níveis de confiança de 90% e 95%. Já para o nível de confiança de 99%, se perde o balanceamento pois sua precisão depende do número de réplicas bootstrap, que em nosso é de 600 réplicas. Percebemos que o aumento do tamanho de amostra melhora o desempenho de todos os intervalos de confiança, dado que para uma mesma cobertura os comprimentos diminuem. Diferente do comportamento dos intervalos de confiança para β_0 e β_1 , observamos nas tabelas 4.3.6.a e 4.3.6.b que os padrões de comportamento dos intervalos de confiança ICA, ICP e BCa, em termos de comprimento, cobertura e porcentagens “% Esquerdo” e “% Direito” desaparecem, considerando os diferentes níveis de confiança e tamanhos de amostra.

Esta situação se manifesta para os intervalos de confiança IACC com ICCP, e IACP com ICCP. Como se discutiu na seção anterior, as estimativas de máxima verossimilhança são muito viesadas e a correção por viés modifica consideravelmente a locação e variabilidade da distribuição empírica do estimador ϕ^* , como revelaram os dados da tabela 4.3.3. Na medida que os intervalos de confiança de tipo assintótico não abrangem valores fora do espaço paramétrico, estes são os que apresentam os melhores desempenhos em termos de cobertura e comprimento para tamanhos de amostra levemente grandes. A tabela 4.3.6.c apresenta os intervalos de confiança assintóticos para o parâmetro de intercepto $\phi = 120$, sendo que os intervalos assintóticos de tipo IACC e IACP para ϕ são da forma (3.2.1.9) e

Tabela 4.3.6.a. Comportamento das estimativas intervalares do parâmetro ϕ .

Medidas		Inferior			Superior			Tamanho		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	71.5855	84.8059	90.1850	226.3207	183.2602	167.4361	154.7351	98.4543	77.2511
	IACC	57.2681	75.7908	84.1041	180.9720	163.7582	156.1371	123.7038	87.9674	72.0330
	IACP	89.5528	93.6098	96.5734	283.2348	202.3065	179.3081	193.6820	108.6966	82.7346
	ICP	103.1924	100.3979	100.6457	316.3158	214.9100	185.9321	213.1233	114.5121	85.2863
	ICCC	82.5497	90.3284	93.9217	316.3158	214.9100	185.9321	233.7661	124.5816	92.0103
	ICCP	129.0322	111.6143	107.8231	395.7099	238.8936	199.3043	266.6777	127.2792	91.4812
	BCa	70.9730	81.2490	87.7650	215.4940	179.0212	164.8905	144.5209	97.7722	77.1254
5%	ICA	56.7639	75.3753	82.7853	241.1422	192.6909	174.8358	184.3783	117.3156	92.0504
	IACC	45.4189	67.3647	77.2043	192.8212	172.1844	163.0369	147.4022	104.8197	85.8326
	IACP	71.0006	83.1981	88.6485	301.7870	212.7182	187.2330	230.7863	129.5201	98.5844
	ICP	94.4137	94.1073	95.3795	359.6792	233.4071	198.3358	265.2655	139.2997	102.9562
	ICCC	75.5318	84.6862	89.0339	287.8869	210.0542	185.1564	212.3551	125.3680	96.1225
	ICCP	118.0250	104.6301	102.1964	450.2830	259.3473	212.5144	332.2579	154.7171	110.3179
	BCa	70.7937	76.4303	81.1775	235.6825	190.9130	173.9051	164.8888	114.4826	92.7275
1%	ICA	27.7960	56.9436	68.3232	270.1102	211.1225	189.2979	198.5247	126.3165	99.1129
	IACC	22.2604	50.8963	63.7190	215.9798	188.6527	176.5222	193.7194	137.7564	112.8032
	IACP	34.7414	62.8490	73.1598	338.0462	233.0672	202.7217	303.3047	170.2182	129.5618
	ICP	79.3355	82.9306	85.7505	464.3110	273.5393	224.5959	236.9803	131.9794	100.1815
	ICCC	63.5890	74.4965	80.0194	371.7340	246.2073	209.5864	252.7268	140.4135	105.9127
	ICCP	99.2311	92.1101	91.9522	580.2628	303.9615	240.4696	296.4788	146.7835	107.3520
	BCa	70.7937	76.2221	79.7589	278.9758	215.6098	192.4880	245.5221	138.6879	106.1731

Tabela 4.3.6. b. Comportamento das estimativas intervalares do parâmetro ϕ .

α	Medidas	Cobertura			% Esquerdo			% Direito		
		$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	Intervalo									
	ICA	91.10	90.36	89.78	5.40	5.72	5.66	3.50	3.92	4.56
	IACC	84.04	87.12	87.26	1.38	2.16	2.98	14.58	10.72	9.76
	IACP	84.82	87.06	86.94	14.86	11.82	11.12	0.32	1.12	1.94
	ICP	74.14	81.20	84.12	25.84	18.34	14.86	0.02	0.46	1.02
	ICCC	87.66	88.28	89.26	10.72	9.46	8.20	1.62	2.26	2.54
	ICCP	50.14	68.28	74.88	49.86	31.58	24.72	0.00	0.14	0.40
5%	BCa	88.62	89.80	90.36	5.42	5.20	4.62	5.96	5.00	5.02
	ICA	96.72	95.84	94.78	1.32	1.90	2.50	1.96	2.26	2.72
	IACC	89.14	92.30	92.86	0.40	0.74	1.18	10.46	6.96	5.96
	IACP	94.58	94.16	94.28	5.22	5.30	4.88	0.20	0.54	0.84
	ICP	81.46	87.14	90.08	18.54	12.72	9.48	0.00	0.14	0.44
	ICCC	87.66	88.28	89.26	7.10	5.84	4.86	0.46	0.64	1.26
	ICCP	50.14	68.28	74.88	39.76	23.08	17.26	0.00	0.04	0.12
1%	BCa	91.38	94.08	95.46	5.34	3.10	2.04	3.28	2.82	2.50
	ICA	99.36	99.22	99.04	0.02	0.12	0.24	0.62	0.66	0.72
	IACC	94.96	97.36	97.60	0.00	0.02	0.04	5.04	2.62	2.36
	IACP	99.92	99.50	99.30	0.04	0.42	0.54	0.04	0.08	0.16
	ICP	90.92	94.50	96.56	9.08	5.48	3.40	0.00	0.02	0.04
	ICCC	97.16	97.76	98.12	2.84	2.18	1.68	0.00	0.06	0.20
	ICCP	77.42	88.76	92.98	22.58	11.22	6.98	0.00	0.02	0.04
	BCa	94.12	96.48	97.62	5.34	3.00	1.62	0.54	0.52	0.76

(3.3.6) respectivamente. Percebemos que, de forma análoga ao analisado para os intervalos de confiança anteriores, os intervalos de confiança assintóticos para ϕ não abrangem valores fora do espaço paramétrico e seu comportamento é similar aos dos intervalos de confiança IACC e IACP para ϕ da forma (3.2.1.8) e (3.3.5).

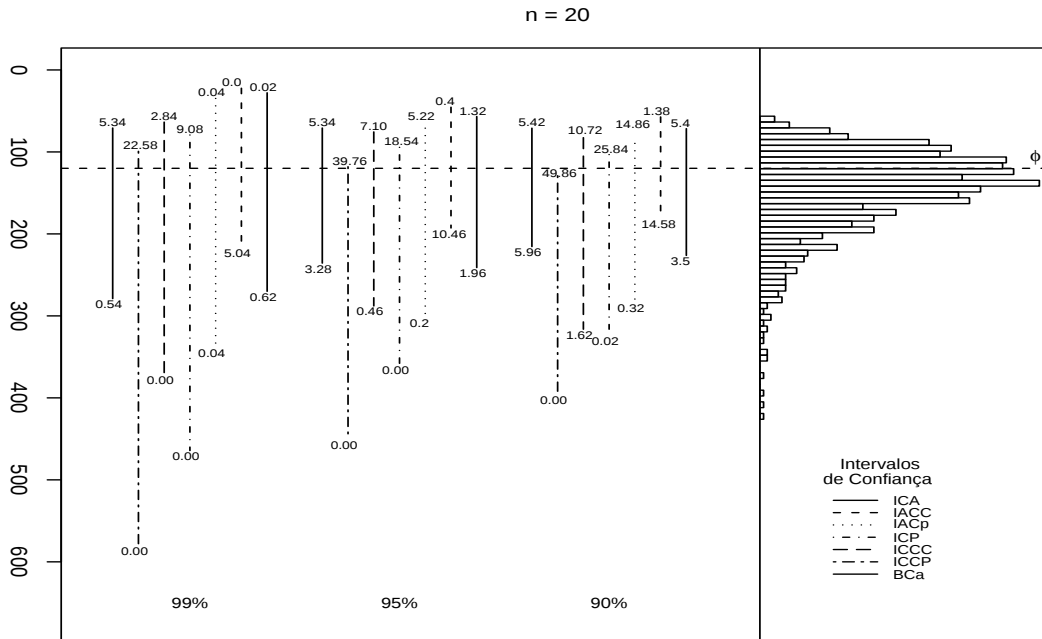


Figura 4.3.2.7: Estimação intervalar para ϕ com $n = 20$.

Ao compararmos os intervalos IACC e IACP nas versões (3.2.1.10) e (3.3.7) com respeito aos outros intervalos, notamos que os intervalos IACC apresentam o menor comprimento e os BCa a melhor cobertura, para os diferentes tamanhos de amostra considerados e um nível de confiança do 90%. Para os níveis de confiança de 95% e 99% as melhores coberturas para os diferentes tamanhos de amostra se apresentam para os intervalos de confiança de tipo ICA e IACP. Este resultado é análogo quando analisamos IACC e IACP nas versões (3.2.1.7) e (3.3.4) respectivamente, dado que eles apresentam os mesmos comprimentos que o intervalo de confiança IAC, diferendo somente na sua localização para os diferentes tamanhos de amostra e níveis de confiança considerados, como era de se esperar, já que em sua construção, eles utilizam os mesmos erros padrões do estimador de máxima verossimilhança. Daí, percebemos que os intervalos de tipo assintótico ICA e IACC são os que apresentam os melhores desempenhos para os diferentes níveis de confiança e tamanhos de amostra; sendo que para um nível de confiança do 99%, os intervalos de tipo BCa têm os menores tamanhos ao considerar tamanhos de amostra pequenos. Desta forma, podemos concluir que a correção analítica e bootstrap fornece boas estimações pontuais e intervalares para os parâmetros do modelo de regressão beta. Apesar de que o intervalo de confiança ICA apresenta um tamanho superior

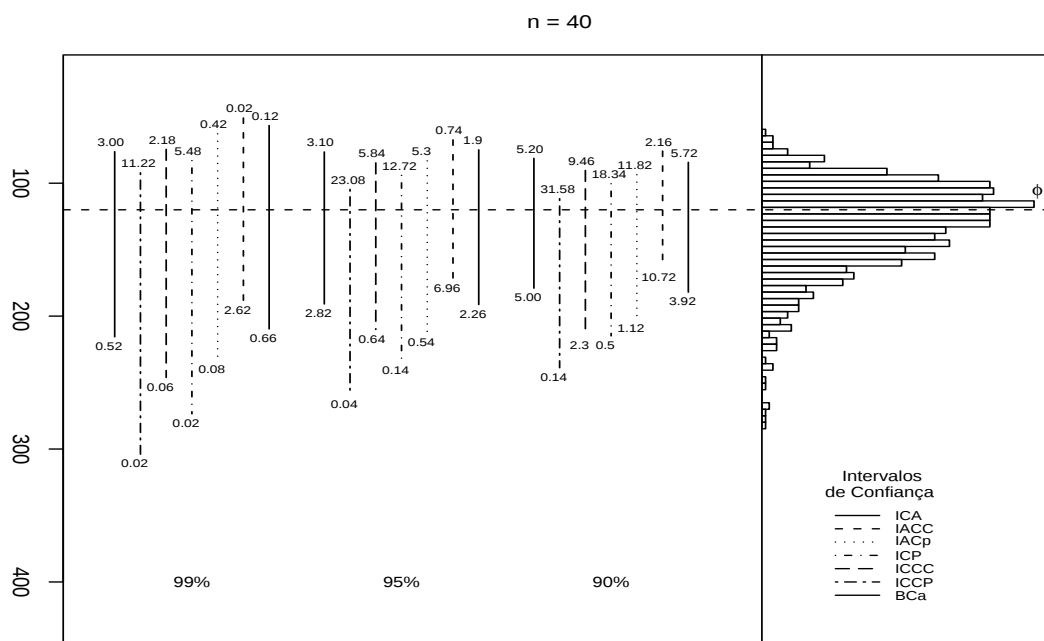


Figura 4.3.2.8: Estimação intervalar para ϕ com $n = 40$.

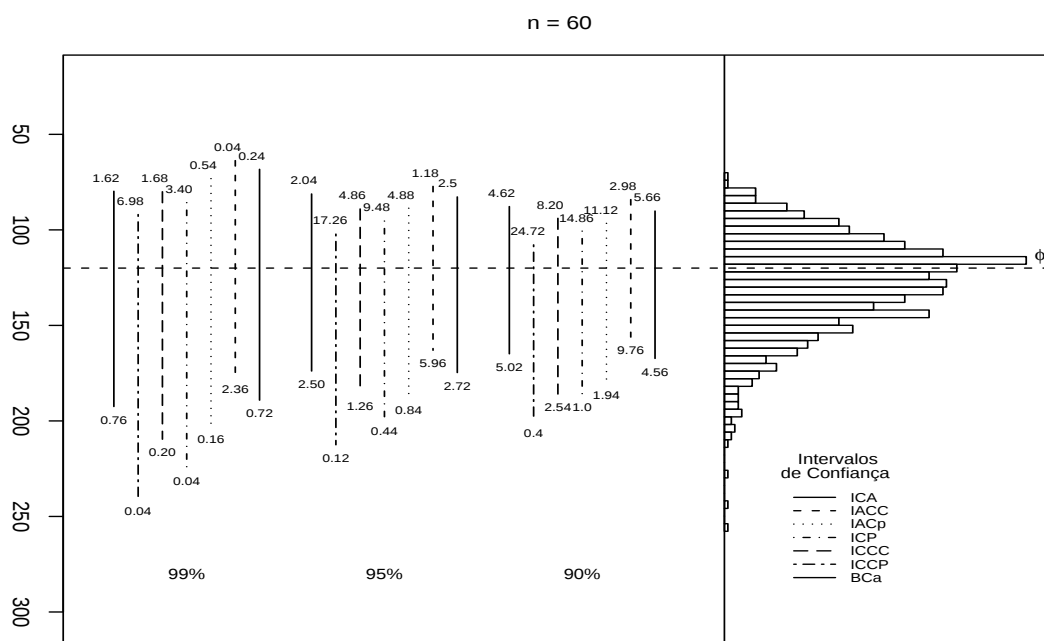


Figura 4.3.2.9: Estimação intervalar para ϕ com $n = 60$.

comparado com o intervalo IACC recomendamos utilizar o intervalo de confiança assintótico ICA pois ele é mais balanceado e sua cobertura é melhor.

Tabela 4.3.6.c. Comportamento das estimativas intervalares assintóticas do parâmetro ϕ .

Medidas		Inferior			Superior			Tamanho		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	71.5855	84.8059	90.1850	226.3207	183.2602	167.4361	154.7351	98.4543	77.2511
	IACC	41.7607	71.3806	81.5833	196.4959	169.8350	158.8344	154.7351	98.4543	77.2511
	IACP	109.0391	99.7595	99.4162	263.7742	198.2139	176.6674	154.7351	98.4543	77.2511
5%	ICA	56.7639	75.3753	82.7853	241.1422	192.6909	174.8358	184.3783	117.3156	92.0504
	IACC	26.9392	61.9500	74.1836	211.3175	179.2656	166.2340	184.3783	117.3156	92.0504
	IACP	94.2175	90.3289	92.0166	278.5958	207.6445	184.0670	184.3783	117.3156	92.0504
1%	ICA	27.7960	56.9436	68.3232	270.1102	211.1225	189.2979	198.5247	126.3165	99.1129
	IACC	0.0287	43.5184	59.7215	240.2854	197.6972	180.6962	242.3142	154.1788	120.9747
	IACP	65.2495	71.8973	77.5544	307.5638	226.0761	198.5292	242.3142	154.1788	120.9747
Medidas		Cobertura			% Esquerdo			% Direito		
α	Intervalo	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$	$n = 20$	$n = 40$	$n = 60$
10%	ICA	90.18	90.24	90.58	5.52	5.98	5.46	4.30	3.78	3.96
	IACC	90.02	90.40	90.34	0.22	1.54	1.92	9.76	8.06	7.74
	IACP	67.86	80.84	84.26	31.22	17.54	13.72	0.92	1.62	2.02
5%	ICA	96.34	95.20	95.38	1.38	2.44	2.32	2.28	2.36	2.30
	IACC	93.38	94.86	95.10	0.00	0.46	0.64	6.62	4.68	4.26
	IACP	81.30	89.74	92.18	18.14	9.56	6.70	0.56	0.70	1.12
1%	ICA	99.30	99.14	98.96	0.00	0.28	0.20	0.70	0.58	0.84
	IACC	97.58	98.34	98.56	0.00	0.00	0.02	2.42	1.66	1.42
	IACP	96.64	98.22	98.50	3.22	1.60	1.08	0.14	0.18	0.42

Nota: Neste caso os intervalos assintóticos IACC e IACP são construídos com o mesmo erro padrão do ICA.

Capítulo 5

Aplicação e conclusões

5.1 Aplicação do modelo de regressão beta

Nesta seção apresentamos uma aplicação do modelo de regressão beta descrito na Seção 2.3. Consideramos os dados de gasolina de Prater descritos anteriormente. Nosso interesse recai na modelagem da proporção de petróleo cru que é convertido em gasolina após a destilação e o fracionamento. Os dados consistem de um conjunto de dez valores de três variáveis explicativas que correspondem aos diferentes tipos de petróleo cru que são submetidos a experimentação sob diferentes condições de destilação. Os dados encontram-se ordenados de forma ascendente de acordo com a covariável que mede a temperatura em que 10% do petróleo cru é vaporizado. Esta variável assume dez valores diferentes que são usados para definir o décimo nível de petróleo cru. A especificação do modelo para a resposta média usa um intercepto ($x_1 = 1$), nove variáveis “dummy” para os primeiros nove níveis do petróleo cru ($x_2, \dots, x_{10}$) e a covariável x_{11} , que mede a temperatura em graus Fahrenheit em que a gasolina é vaporizada. O conjunto de dados está formado por 32 observações na resposta e nas variáveis explicativas.

Para ajustar a resposta média usamos a função de ligação logit. Os parâmetros são estimados através da maximização da função de log-verossimilhança utilizando o método quasi-Newton BFGS com primeiras derivadas analíticas. Fornecemos as respectivas estimativas corrigidas de tipo corretivo, preventivo e de bootstrap dos parâmetros deste modelo. Na Tabela 5.1 apresentamos as estimativas de máxima verossimilhança e suas versões corrigidas com seus respectivos erros-padrão. Observamos que as estimativas pontuais dos parâmetros de regressão β_1 até β_{11} não diferem muito umas das outras, sendo que em geral as estimativas corrigidas preventivas DFirth excedem os valores das estimativas de máxima verossimilhança. Observamos que as estimativas corrigidas corretivamente (ECCS) e por bootstrap (CBC) permanecem quase constantes relativamente às estimativas EMV; por exemplo, para β_6 a EMV é 1.13375 e para ECCS e CBC são 1.13324 e 1.13384, respectivamente.

Desta forma, como foi discutido na Seção 4.3, as correções de Cox & Snell e de bootstrap para os parâmetros da regressão são eficazes, mas o ganho de precisão não é muito alto comparado com as EMV's como se observa a partir dos valores dos erros-padrão. Ao analisar as estimativas do parâmetro de precisão, observamos uma forte mudança entre a estimativa de máxima verossimilhança e as estimativas corrigidas, notando que a correção por bootstrap é mais forte do que a correção de Cox & Snell, com um erro-padrão maior que os das outras estimativas de ϕ .

A correção corretiva não é tão forte quanto comparada com a de bootstrap, mas fornece um erro-padrão menor que as estimativas de máxima verossimilhança, corrigidas preventivamente e por bootstrap. Neste modelo poderíamos considerar como o melhor estimador de ϕ o ECCS.

Tabela 5.1. Parâmetros estimados por máxima verossimilhança e suas versões corrigidas, com erros padrões e usando os dados de gasolina de Prater.

Parâmetro	EMV	ECCS	DFirth	CBC
β_1	-6.15957 (0.18232)	-6.15473 (0.23610)	-6.19875 (0.17176)	-6.13479 (0.18448)
β_2	1.72773 (0.10123)	1.72699 (0.13115)	1.74411 (0.09518)	1.73227 (0.10165)
β_3	1.32260 (0.11790)	1.32149 (0.15270)	1.33326 (0.11087)	1.31756 (0.11851)
β_4	1.57231 (0.11610)	1.57108 (0.15039)	1.57829 (0.10925)	1.56388 (0.11775)
β_5	1.05971 (0.10236)	1.05943 (0.13259)	1.06642 (0.09624)	1.05534 (0.10201)
β_6	1.13375 (0.10352)	1.13324 (0.13412)	1.13551 (0.09736)	1.13384 (0.10181)
β_7	1.04016 (0.10604)	1.03966 (0.13736)	1.04606 (0.09969)	1.03940 (0.10193)
β_8	0.54369 (0.10913)	0.54384 (0.14129)	0.53526 (0.10279)	0.54428 (0.10965)
β_8	0.49590 (0.10893)	0.49582 (0.14107)	0.48789 (0.10255)	0.49891 (0.11456)
β_{10}	0.38579 (0.11859)	0.38508 (0.15363)	0.39056 (0.11143)	0.38778 (0.12342)
β_{11}	0.01097 (0.00041)	0.01096 (0.00053)	0.01107 (0.00039)	0.01090 (0.00039)
ϕ	440.27840 (110.02563)	261.19964 (65.25724)	498.42340 (124.56206)	139.87717 (226.34557)

5.2 Conclusões

No desenvolvimento deste trabalho, foram apresentadas características e propriedades do modelo de regressão beta. Derivamos para este modelo uma expressão que permite remover o viés de segunda ordem das estimativas de máxima verossimilhança dos parâmetros e do vetor de médias do modelo de regressão beta. Mostramos que o viés de segunda or-

dem obtido através da fórmula de Cox & Snell (1968) pode ser escrito em termos de uma regressão de mínimos quadrados generalizados, o que facilita seu cálculo. Além disso, aquela fórmula nos permite construir uma versão modificada da função de máxima verossimilhança, que em teoria nos permite encontrar estimativas de máxima verossimilhança corrigidas preventivamente. Adicionalmente, apresentamos a metodologia de correção dos estimadores de máxima verossimilhança por bootstrap paramétrico. Os resultados de simulação mostram que as correções de viés são eficazes quando os tamanhos de amostra são pequenos (em torno de vinte), sendo que as correções analítica de Cox & Snell (1968) e por bootstrap apresentam os melhores desempenhos em termos de viés, viés relativo, EQM e $\sqrt{\text{EQM}}$, ao contrário do que acontece com a correção preventiva das estimativas dos parâmetros, já que esta fornece estimativas com desempenhos desfavoráveis comparados ao das estimativas de máxima verossimilhança. As estimativas corrigidas dos parâmetros de regressão não apresentam uma forte melhoria no desempenho com relação ao viés e precisão quando comparadas com as estimativas de máxima verossimilhança, indicando que apresentam boas propriedades em amostras finitas. No caso do parâmetro de precisão, observamos que as estimativas de máxima verossimilhança se apresentam muito viesadas e, desta forma, recomendamos fortemente a correção de viés. Para isso, sugerimos a correção analítica de Cox & Snell (1968) ou a correção via bootstrap, onde estas metodologias apresentam desempenhos mais favoráveis ao serem comparadas com a estimação por máxima verossimilhança.

Ao analisarmos o comportamento das estimativas intervalares, observamos que em geral para os parâmetros de regressão os intervalos de tipo percentil ICCP e assintótico IACP são os que apresentam os menores comprimentos e que os intervalos do tipo IACC apresentam as melhores coberturas. Todos os intervalos construídos para os parâmetros de regressão são aproximadamente simétricos. Entretanto, para o parâmetro de precisão os intervalos de confiança construídos são assimétricos e desbalanceados, sendo que os intervalos assintóticos IACC apresentam os menores comprimentos e os IAC as melhores coberturas quando os tamanhos da amostra são relativamente pequenos. Para um nível de cobertura de 90% os intervalos do tipo BCa são os de menor comprimento e os mais balanceados. Em geral, para o parâmetro ϕ os intervalos assintóticos apresentam os melhores desempenhos, sendo que aconselhamos utilizar o intervalo de confiança IAC que se apresenta mais balanceado.

Apêndice

A. Cálculo dos momentos

A.1. Obtenção dos momentos das derivadas da função de log-verossimilhança para os cálculos dos vieses de $\hat{\beta}$ e $\hat{\phi}$

A função de log-verossimilhança para o modelo de regressão beta proposto neste trabalho é da forma

$$\ell(\beta, \phi) = \sum_{t=1}^n \ell_t(\mu_t, \phi),$$

onde

$$\begin{aligned} \ell_t(\mu_t, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t \\ &\quad + \{((1 - \mu_t)\phi - 1)\} \log(1 - y_t). \end{aligned}$$

Seja

$$g(\mu_t) = \sum_{i=1}^k x_{ti} \beta_i = \eta_t,$$

onde $g(\mu) = \log(\frac{\mu}{1-\mu})$ ou $g(\mu) = \Phi^{-1}(\mu)$, $\Phi(\cdot)$ sendo a função de distribuição acumulada da variável normal padrão ou qualquer função diferenciável $g : (0, 1) \rightarrow \mathbb{R}$. Além disso, temos

$$\frac{d\mu_t}{d\eta_t} = \frac{dg^{-1}(\eta_t)}{d\eta_t} = \frac{1}{g'(\mu_t)}.$$

Momentos da função de log-verossimilhança

Sob as condições de regularidade listadas, a função escore é

$$\begin{aligned} U_r &= \frac{\partial \ell(\beta, \phi)}{\partial \beta_r} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \\ &= \sum_{t=1}^n \{-\phi \psi(\mu_t \phi) + \phi \psi((1 - \mu_t)\phi) + \phi \log y_t - \phi \log(1 - y_t)\} \frac{d\mu_t}{d\eta_t} x_{tr} \\ U_r &= \sum_{t=1}^n \phi \left[\log\left(\frac{y_t}{1 - y_t}\right) - \{\psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)\} \right] \frac{d\mu_t}{d\eta_t} x_{tr}. \end{aligned}$$

Definimos $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ e $\mu_t^* = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi)$; das condições de regularidade, temos que $E\left(\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t}\right) = 0$, logo $E(y_t^*) = \mu_t^*$. Além disso,

$$\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t} = \phi[y_t^* - \mu_t^*],$$

logo

$$U_r = \sum_{t=1}^n \phi[y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} x_{tr}. \quad (A.1.1)$$

Adicionalmente,

$$\begin{aligned} U_\phi &= \frac{\partial\ell(\beta, \phi)}{\partial\phi} = \sum_{t=1}^n \frac{\partial\ell_t(\mu_t, \phi)}{\partial\phi} \\ &= \sum_{t=1}^n \{\psi(\phi) - \psi(\mu_t\phi)\mu_t - \psi((1-\mu_t)\phi)(1-\mu_t) + \mu_t \log y_t + (1-\mu_t) \log(1-y_t)\} \\ &= \sum_{t=1}^n \left\{ \psi(\phi) - \psi(\mu_t\phi)\mu_t - \psi((1-\mu_t)\phi) + \mu_t \psi((1-\mu_t)\phi) \right. \\ &\quad \left. + \log(1-y_t) + \mu_t \log\left(\frac{y_t}{1-y_t}\right) \right\} \\ &= \sum_{t=1}^n \left\{ \mu_t \left(\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi) - \psi((1-\mu_t)\phi)] \right) \right. \\ &\quad \left. + \log(1-y_t) + \psi(\phi) - \psi((1-\mu_t)\phi) \right\}. \end{aligned}$$

Assim,

$$U_\phi = \sum_{t=1}^n \{\mu_t[y_t^* - \mu_t^*] + \log(1-y_t) + \psi(\phi) - \psi((1-\mu_t)\phi)\}. \quad (A.1.2)$$

Os momentos de segunda e terceira ordem obtidos pela diferenciação da função de log-verossimilhança são

$$\begin{aligned} U_{rs} &= \frac{\partial^2\ell(\beta, \phi)}{\partial\beta_r\partial\beta_s} = \sum_{t=1}^n \frac{\partial}{\partial\mu_t} \left(\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial\eta_t}{\partial\beta_r} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial\eta_t}{\partial\beta_s} \\ &= \sum_{t=1}^n \left\{ \left(\frac{\partial^2\ell_t(\mu_t, \phi)}{\partial\mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right) + \left(\frac{\partial\ell_t(\mu_t, \phi)}{\partial\mu_t} \left(\frac{\partial}{\partial\mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \right\} x_{ts} x_{tr}, \end{aligned}$$

onde

$$\frac{\partial^2\ell_t(\mu_t, \phi)}{\partial\mu_t^2} = \frac{\partial}{\partial\mu_t} \{\phi[y_t^* - \mu_t^*]\} = -\phi \frac{\partial\mu_t^*}{\partial\mu_t},$$

e

$$-\phi \frac{\partial\mu_t^*}{\partial\mu_t} = -\phi \{\phi[\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)]\}.$$

Definimos $w_t = [\psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi)]$, então $\frac{\partial \mu_t^*}{\partial \mu_t} = \phi w_t$ e $\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} = -\phi^2 w_t$. Assim,

$$U_{rs} = \sum_{t=1}^n \left\{ -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \phi[y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right\} x_{ts} x_{tr}, \quad (A.1.3)$$

$$\begin{aligned} U_{rsu} &= \frac{\partial^3 \ell(\beta, \phi)}{\partial \beta_r \partial \beta_s \partial \beta_u} \\ &= \sum_{t=1}^n \frac{\partial}{\partial \mu_t} \left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \frac{\partial \eta_t}{\partial \beta_s} \frac{\partial \eta_t}{\partial \beta_u}, \\ U_{rsu} &= \sum_{t=1}^n \left\{ \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^3} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + 2 \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right. \\ &\quad + \left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} + \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \\ &\quad \left. + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \right\} \frac{d\mu_t}{d\eta_t} x_{ts} x_{tr} x_{tu}. \end{aligned}$$

Temos que

$$\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^3} = \frac{\partial(-\phi^2 w_t)}{\partial \mu_t} = -\phi^2 \frac{\partial w_t}{\partial \mu_t},$$

com

$$\frac{\partial w_t}{\partial \mu_t} = \phi[\psi''(\mu_t\phi) - \psi''((1 - \mu_t)\phi)].$$

Seja $m_t = [\psi''(\mu_t\phi) - \psi''((1 - \mu_t)\phi)]$, logo $\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^3} = -\phi^3 m_t$. Desta forma,

$$\begin{aligned} U_{rsu} &= \sum_{t=1}^n \left\{ -\phi^3 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 - 2\phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \phi[y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \right. \\ &\quad \left. - \phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \phi[y_t^* - \mu_t^*] \right\} \frac{d\mu_t}{d\eta_t} x_{ts} x_{tr} x_{tu}, \\ U_{rsu} &= \sum_{t=1}^n \left\{ -\phi^3 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 - 3\phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right. \\ &\quad \left. + \phi[y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \right] \right\} x_{ts} x_{tr} x_{tu}, \\ U_{rsu} &= -\phi \sum_{t=1}^n \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + 3\phi w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right. \\ &\quad \left. - [y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \right] \right\} x_{ts} x_{tr} x_{tu}. \end{aligned}$$

Sejam $a_t = 3\left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2$ e $b_t = \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t}\right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)^2 \right]$, portanto

$$U_{rsu} = -\phi \sum_{t=1}^n \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \phi w_t a_t - [y_t^* - \mu_t^*] b_t \right\} x_{ts} x_{tr} x_{tu}. \quad (A.1.4)$$

$$\begin{aligned} U_{\phi\phi} &= \frac{\partial^2 \ell(\beta, \phi)}{\partial \phi^2} = \sum_{t=1}^n \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \phi^2} \\ &= \sum_{t=1}^n \left\{ -\mu_t \frac{\partial \mu_t^*}{\partial \phi} + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t)\phi) \right\}, \end{aligned}$$

onde

$$\begin{aligned} \frac{\partial \mu_t^*}{\partial \phi} &= \mu_t \psi'(\mu_t \phi) - (1 - \mu_t) \psi'((1 - \mu_t)\phi) \\ &= \mu_t \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi) \} - \psi'((1 - \mu_t)\phi) \\ &= \mu_t w_t - \psi'((1 - \mu_t)\phi). \end{aligned}$$

Desta forma,

$$\begin{aligned} U_{\phi\phi} &= \sum_{t=1}^n \left\{ -\mu_t [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t)\phi) \right\} \\ &= \sum_{t=1}^n \left\{ -\mu_t^2 \psi'(\mu_t \phi) + \mu_t (1 - \mu_t) \psi'((1 - \mu_t)\phi) + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t)\phi) \right\} \\ &= \sum_{t=1}^n - \left\{ (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi) \right\}. \end{aligned}$$

Seja $d_t = (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi)$, logo

$$U_{\phi\phi} = \sum_{t=1}^n -d_t. \quad (A.1.5)$$

Adicionalmente

$$\begin{aligned} U_{\phi\phi\phi} &= \frac{\partial^3 \ell(\beta, \phi)}{\partial \phi^3} = \sum_{t=1}^n \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \phi^3} \\ &= - \sum_{t=1}^n \{ (1 - \mu_t)^3 \psi''((1 - \mu_t)\phi) + \mu_t^3 \psi''(\mu_t \phi) - \psi''(\phi) \}. \end{aligned}$$

Seja $s_t = (1 - \mu_t)^3 \psi''((1 - \mu_t)\phi) + \mu_t^3 \psi''(\mu_t \phi) - \psi''(\phi)$, logo $s_t = \frac{\partial d_t}{\partial \phi}$. Assim,

$$U_{\phi\phi\phi} = - \sum_{t=1}^n s_t, \quad (A.1.6)$$

$$\begin{aligned}
U_{r\phi} &= \frac{\partial^2 \ell(\beta, \phi)}{\partial \beta_r \partial \phi} = \sum_{t=1}^n \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \\
&= \sum_{t=1}^n \left\{ \frac{\partial}{\partial \phi} (\phi[y_t^* - \mu_t^*]) \right\} \frac{d\mu_t}{d\eta_t} x_{tr},
\end{aligned}$$

onde

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} = \frac{\partial}{\partial \phi} (\phi[y_t^* - \mu_t^*]) = [y_t^* - \mu_t^*] - \phi \frac{\partial \mu_t^*}{\partial \phi},$$

Seja $c_t = \phi[\mu_t w_t - \psi'((1 - \mu_t)\phi)]$, logo $\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} = [y_t^* - \mu_t^*] - c_t$. Assim,

$$U_{r\phi} = \sum_{t=1}^n -\{c_t - [y_t^* - \mu_t^*]\} \frac{d\mu_t}{d\eta_t} x_{tr}, \quad (A.1.7)$$

$$\begin{aligned}
U_{rs\phi} &= \frac{\partial^3 \ell(\beta, \phi)}{\partial \beta_r \partial \beta_s \partial \phi} \\
&= \sum_{t=1}^n \frac{\partial}{\partial \phi} \left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{\partial \eta_t}{\partial \beta_r} \frac{\partial \eta_t}{\partial \beta_s} \\
&= \sum_{t=1}^n \left\{ \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^2 \partial \phi} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right\} x_{tr} x_{ts}, \\
\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^2 \partial \phi} &= \frac{\partial}{\partial \phi} \{-\phi^2 w_t\} = -2\phi w_t - \phi^2 \frac{\partial w_t}{\partial \phi} \\
&= -2\phi w_t - \phi^2 [\mu_t m_t + \psi''((1 - \mu_t)\phi)] \\
&= -\phi \{2w_t + \phi[\mu_t m_t + \psi''((1 - \mu_t)\phi)]\}.
\end{aligned}$$

Seja $u_t = -\phi\{2w_t + \phi[\mu_t m_t + \psi''((1 - \mu_t)\phi)]\}$, logo

$$U_{rs\phi} = \sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) + [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}, \quad (A.1.8)$$

$$\begin{aligned}
U_{r\phi\phi} &= \frac{\partial^3 \ell(\beta, \phi)}{\partial \beta_r \partial \phi^2} = \sum_{t=1}^n \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r}, \\
\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi^2} &= \frac{\partial}{\partial \phi} \left\{ [y_t^* - \mu_t^*] - \phi \frac{\partial \mu_t^*}{\partial \phi} \right\} = -2 \frac{\partial \mu_t^*}{\partial \phi} - \phi \frac{\partial^2 \mu_t^*}{\partial \phi^2} \\
&= -2 \frac{\partial \mu_t^*}{\partial \phi} - \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \\
&= -2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] - \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)].
\end{aligned}$$

Definimos $r_t = 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi[\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \frac{d\mu_t}{d\eta_t}$, logo

$$U_{r\phi\phi} = \sum_{t=1}^n -r_t x_{tr}. \quad (A.1.9)$$

Cálculo de cumulantes

Tomando o valor esperado dos momentos acima, temos que

$$\begin{aligned}\kappa_{rs} &= E(U_{rs}) = E\left(\sum_{t=1}^n \left\{ -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \phi[y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} x_{ts} x_{tr} \right) \\ &= -\phi^2 \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr},\end{aligned}\tag{A.1.10}$$

$$\begin{aligned}\kappa_{rsu} &= E(U_{rsu}) = E\left(-\phi \sum_{t=1}^n \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \phi w_t a_t - [y_t^* - \mu_t^*] b_t \right\} x_{ts} x_{tr} x_{tu} \right) \\ &= -\phi^2 \sum_{t=1}^n \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + w_t a_t \right\} x_{ts} x_{tr} x_{tu},\end{aligned}\tag{A.1.11}$$

$$\kappa_{\phi\phi} = E(U_{\phi\phi}) = E\left(\sum_{t=1}^n -d_t\right) = -\sum_{t=1}^n d_t,\tag{A.1.12}$$

$$\kappa_{\phi\phi\phi} = E(U_{\phi\phi\phi}) = E\left(\sum_{t=1}^n -s_t\right) = -\sum_{t=1}^n s_t,\tag{A.1.13}$$

$$\begin{aligned}\kappa_{r\phi} &= E(U_{r\phi}) = E\left(\sum_{t=1}^n -\{c_t - [y_t^* - \mu_t^*]\} \frac{d\mu_t}{d\eta_t} x_{tr}\right) \\ &= -\sum_{t=1}^n c_t \frac{d\mu_t}{d\eta_t} x_{tr},\end{aligned}\tag{A.1.14}$$

$$\begin{aligned}\kappa_{rs\phi} &= E(U_{rs\phi}) = E\left(\sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) + [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} \right) \\ &= \sum_{t=1}^n \left\{ -\phi \{2w_t + \phi[\mu_t m_t + \psi''((1-\mu)\phi)]\} \frac{d\mu_t}{d\eta_t} - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}. \\ &= \sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts},\end{aligned}\tag{A.1.15}$$

$$\kappa_{r\phi\phi} = E(U_{r\phi\phi}) = E\left(\sum_{t=1}^n -r_t x_{tr}\right) = -\sum_{t=1}^n r_t x_{tr}.\tag{A.1.16}$$

Derivada de momentos

Derivando os cumulantes de segunda ordem com respeito aos parâmetros de interesse, temos que

$$\begin{aligned}
\kappa_{rs}^{(u)} &= \frac{\partial \kappa_{rs}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(-\phi^2 \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} \\
&= -\phi^2 \sum_{t=1}^n \left(\frac{\partial w_t}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + 2w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} x_{tu} \\
&= -\phi^2 \sum_{t=1}^n \left(\phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 + 2w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} x_{tu} \\
&= -\phi^2 \sum_{t=1}^n \left(\phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{2}{3} w_t a_t \right) x_{tr} x_{ts} x_{tu}, \tag{A.1.17}
\end{aligned}$$

$$\begin{aligned}
\kappa_{rs}^{(\phi)} &= \frac{\partial \kappa_{rs}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(-\phi^2 \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} \right) \\
&= -2\phi \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} - \phi^2 \sum_{t=1}^n \frac{\partial w_t}{\partial \phi} \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} \\
&= -2\phi \sum_{t=1}^n w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} - \phi^2 \sum_{t=1}^n [\mu_t m_t + \psi''((1 - \mu_t)\phi)] \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} \\
&= -\phi \sum_{t=1}^n \{2w_t + \phi[\mu_t m_t + \psi''((1 - \mu_t)\phi)]\} \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ts} x_{tr} \\
&= \sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right\} x_{ts} x_{tr}, \tag{A.1.18}
\end{aligned}$$

$$\begin{aligned}
\kappa_{r\phi}^{(u)} &= \frac{\partial \kappa_{r\phi}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(-\sum_{t=1}^n c_t \frac{d\mu_t}{d\eta_t} x_{tr} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} \\
&= -\sum_{t=1}^n \left\{ \frac{\partial c_t}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}, \\
\frac{\partial c_t}{\partial \mu_t} &= \frac{\partial}{\partial \mu_t} \phi [\mu_t w_t - \psi'((1 - \mu_t)\phi)] \\
&= \phi \left(w_t + \mu_t \frac{\partial w_t}{\partial \mu_t} + \phi \psi''((1 - \mu_t)\phi) \right) \\
&= \phi (w_t + \mu_t \phi m_t + \phi \psi''((1 - \mu_t)\phi)).
\end{aligned}$$

Logo,

$$\kappa_{r\phi}^{(u)} = -\sum_{t=1}^n \left\{ [\phi (w_t + \mu_t \phi m_t + \phi \psi''((1 - \mu_t)\phi))] \frac{d\mu_t}{d\eta_t} + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}, \tag{A.1.19}$$

$$\kappa_{r\phi}^{(\phi)} = \frac{\partial \kappa_{r\phi}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(- \sum_{t=1}^n c_t \frac{d\mu_t}{d\eta_t} x_{tr} \right) = - \sum_{t=1}^n \frac{\partial c_t}{\partial \phi} \frac{d\mu_t}{d\eta_t} x_{tr}.$$

Seja

$$\begin{aligned} \frac{\partial c_t}{\partial \phi} &= \frac{\partial}{\partial \phi} \{ \phi [\mu_t w_t - \psi'((1 - \mu_t)\phi)] \} \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi \left\{ \mu_t \frac{\partial w_t}{\partial \phi} - (1 - \mu_t) \psi''((1 - \mu_t)\phi) \right\} \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi \{ \mu_t [\mu_t m_t + \psi''((1 - \mu_t)\phi)] - (1 - \mu_t) \psi''((1 - \mu_t)\phi) \}. \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]. \end{aligned}$$

Definimos $z_t = [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]$, então

$$\kappa_{r\phi}^{(\phi)} = - \sum_{t=1}^n z_t \frac{d\mu_t}{d\eta_t} x_{tr}, \quad (A.1.20)$$

$$\kappa_{\phi\phi}^{(u)} = \frac{\partial \kappa_{\phi\phi}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(- \sum_{t=1}^n d_t \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} = - \sum_{t=1}^n \frac{\partial d_t}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} x_{tu},$$

$$\begin{aligned} \frac{\partial d_t}{\partial \mu_t} &= \frac{\partial}{\partial \mu_t} \{ (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi) \} \\ &= -2(1 - \mu_t) \psi'((1 - \mu_t)\phi) - \phi (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi) \\ &\quad + 2\mu_t \psi'(\mu_t \phi) + \phi \mu_t^2 \psi''(\mu_t \phi). \\ &= 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) \\ &\quad - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]. \end{aligned}$$

Como $r_t = 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]$, então

$$\kappa_{\phi\phi}^{(u)} = \sum_{t=1}^n -r_t \frac{d\mu_t}{d\eta_t} x_{tu}, \quad (A.1.21)$$

$$\kappa_{\phi\phi}^{(\phi)} = \frac{\partial \kappa_{\phi\phi}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(- \sum_{t=1}^n d_t \right) = - \sum_{t=1}^n \frac{\partial d_t}{\partial \phi} = - \sum_{t=1}^n s_t. \quad (A.1.22)$$

Utilizando os resultados acima, calculamos as quantidades dadas a seguir.

$$\begin{aligned} \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} &= - \phi^2 \sum_{t=1}^n \left(\phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{2}{3} w_t a_t \right) x_{tr} x_{ts} x_{tu} \\ &\quad + \frac{1}{2} \phi^2 \sum_{t=1}^n \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + w_t a_t \right\} x_{ts} x_{tr} x_{tu} \\ &= \sum_{t=1}^n \frac{-\phi^2}{2} \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{1}{3} w_t a_t \right\} x_{ts} x_{tr} x_{tu}. \end{aligned}$$

Seja $W_1 = \text{diag}\left\{\frac{-\phi^2}{2}[\phi m_t \left(\frac{d\mu_t}{d\eta_t}\right)^3 + \frac{1}{3}w_t a_t]\right\}$, então

$$\kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu} = \sum_{t=1}^n w_{1t} x_{tr} x_{ts} x_{tu}, \quad (\text{A.1.23})$$

$$\begin{aligned} \kappa_{rs}^{(\phi)} - \frac{1}{2}\kappa_{rs\phi} &= \sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right\} x_{ts} x_{tr} - \frac{1}{2} \sum_{t=1}^n \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} \\ &= \sum_{t=1}^n \frac{1}{2} \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}. \end{aligned}$$

Definimos $W_2 = \text{diag}\left(\frac{1}{2}\left\{u_t \left(\frac{d\mu_t}{d\eta_t}\right) + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\right\} \frac{d\mu_t}{d\eta_t}\right)$, logo

$$\kappa_{rs}^{(\phi)} - \frac{1}{2}\kappa_{rs\phi} = \sum_{t=1}^n w_{2t} x_{tr} x_{ts}, \quad (\text{A.1.24})$$

$$\begin{aligned} \kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} &= \sum_{t=1}^n \left\{ -\phi \{w_t + \phi[\mu_t m_t + \psi''((1-\mu)\phi)]\} \frac{d\mu_t}{d\eta_t} - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu} \\ &\quad - \frac{1}{2} \sum_{t=1}^n \left\{ -\phi \{2w_t + \phi[\mu_t m_t + \psi''((1-\mu)\phi)]\} \frac{d\mu_t}{d\eta_t} - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu} \\ &= \sum_{t=1}^n \left\{ -\frac{1}{2}\phi^2 [\mu_t m_t + \psi''((1-\mu)\phi)] \frac{d\mu_t}{d\eta_t} - \frac{1}{2} c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}. \end{aligned}$$

Seja $W_3 = \text{diag}\left(-\frac{1}{2}\left\{\phi^2 [\mu_t m_t + \psi''((1-\mu)\phi)] \frac{d\mu_t}{d\eta_t} + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\right\} \frac{d\mu_t}{d\eta_t}\right)$, daí

$$\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} = \sum_{t=1}^n w_{3t} x_{tr} x_{tu}. \quad (\text{A.1.25})$$

De forma análoga,

$$\kappa_{\phi r}^{(u)} - \frac{1}{2}\kappa_{\phi r u} = \kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} = \sum_{t=1}^n w_{3t} x_{tr} x_{tu}. \quad (\text{A.1.26})$$

Agora,

$$\begin{aligned} \kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} &= \sum_{t=1}^n - \left\{ [\mu_t w_t - \psi'((1-\mu_t)\phi)] + \phi[\mu_t^2 \psi''(\mu_t \phi) - (1-\mu_t)^2 \psi''((1-\mu_t)\phi)] \right\} \frac{d\mu_t}{d\eta_t} x_{tr} \\ &\quad - \frac{1}{2} \sum_{t=1}^n - \left\{ 2[\mu_t w_t - \psi'((1-\mu_t)\phi)] + \phi[\mu_t^2 \psi''(\mu_t \phi) - (1-\mu_t)^2 \psi''((1-\mu_t)\phi)] \right\} \frac{d\mu_t}{d\eta_t} x_{tr} \\ &= \sum_{t=1}^n - \frac{1}{2} \phi[\mu_t^2 \psi''(\mu_t \phi) - (1-\mu_t)^2 \psi''((1-\mu_t)\phi)] \frac{d\mu_t}{d\eta_t} x_{tr}. \end{aligned}$$

Definimos $W_4 = \text{diag}\left(\frac{-1}{2}\phi[\mu_t^2\psi''(\mu_t\phi) - (1-\mu_t)^2\psi''((1-\mu_t)\phi)]\frac{d\mu_t}{d\eta_t}\right)$, logo

$$\kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} = \sum_{t=1}^n w_{4t}x_{tr}. \quad (\text{A.1.27})$$

$$\begin{aligned} \kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} &= \sum_{t=1}^n -r_t \frac{d\mu_t}{d\eta_t} x_{tu} - \frac{1}{2} \sum_{t=1}^n -r_t \frac{d\mu_t}{d\eta_t} x_{tu} \\ &= \sum_{t=1}^n -\frac{1}{2} r_t \frac{d\mu_t}{d\eta_t} x_{tu}. \end{aligned}$$

Seja $W_5 = \text{diag}\left(\frac{-1}{2}r_t \frac{d\mu_t}{d\eta_t}\right)$, desta forma

$$\kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} = \sum_{t=1}^n w_{5t}x_{tu}, \quad (\text{A.1.28})$$

$$\kappa_{\phi\phi}^{(\phi)} - \frac{1}{2}\kappa_{\phi\phi\phi} = \sum_{t=1}^n -s_t - \frac{1}{2} \sum_{t=1}^n -s_t = \sum_{t=1}^n -\frac{1}{2}s_t. \quad (\text{A.1.29})$$

Agora calculamos cada parcela da fórmula (3.2.1.1). Seja

$$\begin{aligned} \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu} \right\} &= \sum_{r,s,u} \kappa^{ar} \kappa^{su} \sum_{t=1}^n w_{1t}x_{tr}x_{ts}x_{tu} \\ &= \sum_{t=1}^n w_{1t} \sum_r \kappa^{ar} x_{tr} \sum_{s,u} x_{ts} \kappa^{su} x_{tu}, \end{aligned}$$

definimos $\mathcal{L}_a$ como sendo o a -ésimo vetor coluna da matriz identidade de dimensão $k \times k$. Desta forma,

$$\begin{aligned} \sum_{t=1}^n w_{1t} \sum_r \kappa^{ar} x_{tr} \sum_{s,u} x_{ts} \kappa^{su} x_{tu} &= \sum_{t=1}^n w_{1t} \sum_r \kappa^{ar} x_{tr} (x_t^\top K^{\beta\beta} x_t) \\ &= \sum_{t=1}^n w_{1t} \mathcal{L}_a^\top K^{\beta\beta} x_t (x_t^\top K^{\beta\beta} x_t) \\ &= \mathcal{L}_a^\top K^{\beta\beta} \sum_{t=1}^n w_{1t} x_t (x_t^\top K^{\beta\beta} x_t). \end{aligned}$$

Seja $\delta_{\beta\beta}$ o vetor de dimensão $n \times 1$ obtido da diagonal principal da matriz $XK^{\beta\beta}X^\top$. Assim,

$$\sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu} \right\} = \mathcal{L}_a^\top K^{\beta\beta} X^\top W_1 \delta_{\beta\beta}. \quad (\text{A.1.30})$$

Agora,

$$\begin{aligned}
& \sum_{\phi, s, u} \kappa^{a\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} = \kappa^{a\phi} \sum_{s, u} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\
& = \kappa^{a\phi} \sum_{s, u} \kappa^{su} \sum_{t=1}^n w_{3t} x_{ts} x_{tu} = \kappa^{a\phi} \sum_{t=1}^n w_{3t} \sum_{s, u} x_{ts} \kappa^{su} x_{tu} \\
& = \kappa^{a\phi} \sum_{t=1}^n w_{3t} (x_t^\top K^{\beta\beta} x_t) = \kappa^{a\phi} \text{tr}(W_3 X K^{\beta\beta} X^\top),
\end{aligned}$$

onde $\text{tr}(\cdot)$ é o operador traço de uma matriz quadrada. Logo,

$$\sum_{\phi, s, u} \kappa^{a\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} = \mathcal{L}_a^\top K^{\beta\phi} \text{tr}(W_3 X K^{\beta\beta} X^\top). \quad (A.1.31)$$

Tomemos

$$\begin{aligned}
& \sum_{r, \phi, u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} = \sum_{r, u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} \\
& = \sum_{r, u} \kappa^{ar} \kappa^{\phi u} \sum_{t=1}^n w_{3t} x_{tr} x_{tu} = \sum_{t=1}^n w_{3t} \sum_r \kappa^{ar} x_{tr} \sum_u \kappa^{\phi u} x_{tu} \\
& = \sum_{t=1}^n w_{3t} (\mathcal{L}_a^\top K^{\beta\beta} x_t) (x_t^\top K^{\beta\phi}) = \mathcal{L}_a^\top K^{\beta\beta} \sum_{t=1}^n (x_t w_{3t} x_t^\top) K^{\beta\phi}.
\end{aligned}$$

Porém,

$$\sum_{r, \phi, u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} = \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_3 X) K^{\beta\phi}. \quad (A.1.32)$$

De forma similar

$$\begin{aligned}
& \sum_{r, s, \phi} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} = \sum_{r, s} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\
& = \sum_{r, s} \kappa^{ar} \kappa^{s\phi} \sum_{t=1}^n w_{2t} x_{tr} x_{ts} = \sum_{t=1}^n w_{2t} \sum_r \kappa^{ar} x_{tr} \sum_s \kappa^{s\phi} x_{ts} \\
& = \sum_{t=1}^n w_{2t} (\mathcal{L}_a^\top K^{\beta\beta} x_t) (x_t^\top K^{\beta\phi}) = \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_2 X) K^{\beta\phi}.
\end{aligned}$$

Portanto,

$$\sum_{r, s, \phi} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} = \mathcal{L}_a^\top K^{\beta\beta} (X^\top W_2 X) K^{\beta\phi}. \quad (A.1.33)$$

Calculamos

$$\begin{aligned}
\sum_{\phi, \phi, u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} &= \kappa^{a\phi} \sum_u \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} \\
&= \kappa^{a\phi} \sum_u \kappa^{\phi u} \sum_{t=1}^n w_{5t} x_{tu} = \kappa^{a\phi} \sum_{t=1}^n w_{5t} \sum_u \kappa^{\phi u} x_{tu} \\
&= \kappa^{a\phi} \sum_{t=1}^n w_{5t} (\rho_t^\top X K^{\beta\phi}) = \kappa^{a\phi} \left(\sum_{t=1}^n w_{5t} \rho_t^\top \right) X K^{\beta\phi} \\
&= \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_5) X K^{\beta\phi},
\end{aligned}$$

onde $\text{vdiag}(\cdot)$ é o vetor linha formado pela diagonal principal de uma matriz quadrada e ρ_t é o t -ésimo vetor coluna da matriz identidade de dimensão $n \times n$. Consequentemente,

$$\sum_{\phi, \phi, u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} = \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_5) X K^{\beta\phi}. \quad (A.1.34)$$

Analogamente,

$$\begin{aligned}
\sum_{\phi, s, \phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s \phi} \right\} &= \kappa^{a\phi} \sum_s \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s \phi} \right\} \\
&= \kappa^{a\phi} \sum_s \kappa^{s\phi} \sum_{t=1}^n w_{4t} x_{ts} = \kappa^{a\phi} \sum_{t=1}^n w_{4t} \sum_s \kappa^{s\phi} x_{ts} \\
&= \kappa^{a\phi} \sum_{t=1}^n w_{4t} (\rho_t^\top X K^{\beta\phi}) = \kappa^{a\phi} \left(\sum_{t=1}^n w_{4t} \rho_t^\top \right) X K^{\beta\phi} \\
&= \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_4) X K^{\beta\phi}.
\end{aligned}$$

Logo,

$$\sum_{\phi, s, \phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s \phi} \right\} = \mathcal{L}_a^\top K^{\beta\phi} \text{vdiag}(W_4) X K^{\beta\phi}. \quad (A.1.35)$$

Da mesma forma

$$\begin{aligned}
\sum_{r, \phi, \phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} &= \kappa^{\phi\phi} \sum_r \kappa^{ar} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} \\
&= \kappa^{\phi\phi} \sum_r \kappa^{ar} \sum_{t=1}^n w_{4t} x_{tr} = \kappa^{\phi\phi} \sum_{t=1}^n w_{4t} \sum_r \kappa^{ar} x_{tr} \\
&= \kappa^{\phi\phi} \sum_{t=1}^n w_{4t} (\rho_t^\top X K^{\beta\beta}) \mathcal{L}_a = \kappa^{\phi\phi} \left(\sum_{t=1}^n w_{4t} \rho_t^\top \right) X K^{\beta\beta} \mathcal{L}_a \\
&= \mathcal{L}_a^\top K^{\beta\beta} X^\top \text{vdiag}(W_4)^\top K^{\phi\phi}.
\end{aligned}$$

Assim,

$$\sum_{r, \phi, \phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} = \mathcal{L}_a^\top K^{\beta\beta} X^\top \text{vdiag}(W_4)^\top K^{\phi\phi}. \quad (A.1.36)$$

Por último, temos

$$\begin{aligned}
& \sum_{\phi, \phi, \phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} = \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} \\
& = \kappa^{a\phi} \kappa^{\phi\phi} \sum_{t=1}^n -\frac{1}{2} s_t = \kappa^{a\phi} \kappa^{\phi\phi} \text{tr}(S) \\
& = \mathcal{L}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(S),
\end{aligned}$$

onde S é a matriz $\text{diag}(-\frac{1}{2}s_t)$. Portanto,

$$\sum_{\phi, \phi, \phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} = \mathcal{L}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(S). \quad (A.1.37)$$

B. Programa de simulação

Neste apêndice apresentamos os programas de simulação utilizados neste trabalho; o programa foi desenvolvido na linguagem de programação `Ox`. O programa está dividido em três módulos. No primeiro módulo (Apêndice B.1), encontra-se o programa de simulação de Monte Carlo, o segundo módulo (Apêndice B.2) contém as funções que permitem o cálculo das estimativas corrigidas e a construção dos intervalos de confiança. No terceiro módulo (Apêndice B.3), encontram-se as diretrizes de ligação dos módulos `correction.ox` e `verossim.ox`. O programa de simulação fornece estimativas de máxima verossimilhança com suas versões corrigidas preventiva, corretiva e por bootstrap dos parâmetros do modelo (4.2.1). Ainda são apresentados intervalos de confiança de tipo assintótico e por bootstrap para os níveis nominais de confiança de 90%, 95% e 99%. Além disso, no apêndice B.4 apresentamos, como exemplo, o programa que permite gerar o gráfico do histograma das 5000 estimativas de máxima verossimilhança do parâmetro β_0 com as diferentes versões de intervalos de confiança para os níveis de confiança 90%, 95% e 99% e tamanho de amostra $n = 40$. O programa foi desenvolvido no pacote estatístico R. Para maiores informações sobre esta plataforma, ver <http://www.r-project.org>.

B.1. Programa principal

```
/******  
*  
* PROGRAM: correction.ox  
*  
* USE: Simulates maximum likelihood estimation of the beta distribution  
*       parameters, and also a bias correction scheme by  
*       Cox & Snell, David Firth and Bootstrap parametric method.  
*       The model considered is a beta regression model with a new  
*       scheme of parameters (Ferrari & Cribari-Neto, 2003).  
*  
* AUTHOR: Raydonal Ospina Martinez  
*  
* DATE: June 13, 2003.  
*  
* VERSION: 0.1.5  
*  
* LAST MODIFIED: Novembro 05, 2003. Hora: 11:30  
*  
*****/  
  
/* Arquivos principais */  
#include <oxstd.h>  
#include <oxfloat.h>  
#include <oxprob.h>  
#import <maximize>  
#import <solvernle>  
#include "verosimfinal.h" // no <...> but "..."  
#include "verosimfinal.ox" // also get the code  
  
/* Parametros da simulacao */  
const decl INREP = 5000; //numero de replicas Monte Carlo  
const decl INMIN = 60; //minimo tamanho de amostra  
const decl INMAX = 60; //maximo tamanho de amostra  
const decl B = 600; //numero de replicas Bootstrap  
  
/*corpo principal*/
```

```

main()
{
// Declaracao das variaveis usadas
decl ir1, cfailure, dExecTime, cn, im, vp1, v_eta,v_mu, cs, mmle,v_zEstim,v_eEstim,
v_muXiu, phi_ini, dfunc1, p, q, v_theta, mResMean, vx, vbeta, d_phi,v_z,fpVaEst,
v_betaini,sigmaEst,s_cua,dMucua,d_v_thetaIni, d_denom,d_dem2,m_inv,ci,b_SomaQua,
v_muVee,v_sigma2Est, dfunc,fisherinv, npar,bias,v_vies,gerY,m_inforK, v_Kbetaphi,
d_Khiphi, m_Kbetabeta, mResBias, mResVar, mResMSE, mRelBias, mmlebc, delta_Tilde,
mmDfirth, mBoot, mBootMC, vp2,cfailureboot,bi,v_eta_boot,v_mu_boot,d_phi_boot, bci,
bp, bq, vp3,dfunc3, ir3,d_phi_est,v_eta_est,v_mu_est,v_Bias_theta, gerY1, BootVies,
assim1, assi,AmJK,mBootCox, mBootFirth,ZZZ00,ICBCA,v_varassin,a,b;

// arranque do relodo para o calculo do tempo de execucao
dExecTime = timer();

// turn off warnings
oxwarning( 0 );

// Gerador de numeros uniformes pseudo-aleatorios. "George Marsaglia"
ranseed("GM");

// Parametros verdadeiros do modelo
vbeta = <1.5,1.8>; //b_0 , b_1
d_phi = 120; //phi
v_theta = vbeta | d_phi;
parametro=v_theta;

//Quantis da distribuicao Normal
decl z90 = quann(0.950);
decl z95 = quann(0.975);
decl z99 = quann(0.995);

// inicializacao das matricizes das replicas Monte Carlo
mmle = zeros(3,INREP);
mmDfirth = zeros(3,INREP);
mmlebc = zeros(3,INREP);
mBootMC = zeros(3,INREP);

//Cubrimientos
decl m_cubriNorm90 = zeros(3,INREP);
decl m_cubriCS90 = zeros(3,INREP);
decl m_cubriFirth90 = zeros(3,INREP);
decl m_cubriBoot90 = zeros(3,INREP);
decl m_cubriBoot90Cox = zeros(3,INREP);
decl m_cubriBoot90Firth = zeros(3,INREP);
decl m_cubriBoot90BCA = zeros(3,INREP);

decl m_cubriNorm95 = zeros(3,INREP);
decl m_cubriCS95 = zeros(3,INREP);
decl m_cubriFirth95 = zeros(3,INREP);
decl m_cubriBoot95 = zeros(3,INREP);
decl m_cubriBoot95Cox = zeros(3,INREP);
decl m_cubriBoot95Firth = zeros(3,INREP);
decl m_cubriBoot95BCA = zeros(3,INREP);

decl m_cubriNorm99 = zeros(3,INREP);
decl m_cubriCS99 = zeros(3,INREP);
decl m_cubriFirth99 = zeros(3,INREP);
decl m_cubriBoot99 = zeros(3,INREP);
decl m_cubriBoot99Cox = zeros(3,INREP);
decl m_cubriBoot99Firth = zeros(3,INREP);
decl m_cubriBoot99BCA = zeros(3,INREP);

//limites de confianca ao 90%
decl m_limInfNorm90 = zeros(3,INREP);
decl m_SupInfNorm90 = zeros(3,INREP);
decl m_limInfCS90 = zeros(3,INREP);
decl m_SupInfCS90 = zeros(3,INREP);
decl m_limInfFirth90 = zeros(3,INREP);
decl m_SupInfFirth90 = zeros(3,INREP);
decl m_limInfBoot90 = zeros(3,INREP);
decl m_SupInfBoot90 = zeros(3,INREP);
decl m_limInfBoot90Cox = zeros(3,INREP);
decl m_SupInfBoot90Cox = zeros(3,INREP);
decl m_limInfBoot90Firth = zeros(3,INREP);
decl m_SupInfBoot90Firth = zeros(3,INREP);
decl m_limInfBoot90BCA = zeros(3,INREP);

```

```

decl  m_SupInfBoot90BCA = zeros(3,INREP);

//limites de confianca ao 95%
decl  m_limInfNorm95 = zeros(3,INREP);
decl  m_SupInfNorm95 = zeros(3,INREP);
decl  m_limInfCS95 = zeros(3,INREP);
decl  m_SupInfCS95 = zeros(3,INREP);
decl  m_limInfFirth95 = zeros(3,INREP);
decl  m_SupInfFirth95 = zeros(3,INREP);
decl  m_limInfBoot95 = zeros(3,INREP);
decl  m_SupInfBoot95 = zeros(3,INREP);
decl  m_limInfBoot95Cox = zeros(3,INREP);
decl  m_SupInfBoot95Cox = zeros(3,INREP);
decl  m_limInfBoot95Firth = zeros(3,INREP);
decl  m_SupInfBoot95Firth = zeros(3,INREP);
decl  m_limInfBoot95BCA = zeros(3,INREP);
decl  m_SupInfBoot95BCA = zeros(3,INREP);

//limites de confianca ao 99%
decl  m_limInfNorm99 = zeros(3,INREP);
decl  m_SupInfNorm99 = zeros(3,INREP);
decl  m_limInfCS99 = zeros(3,INREP);
decl  m_SupInfCS99 = zeros(3,INREP);
decl  m_limInfFirth99 = zeros(3,INREP);
decl  m_SupInfFirth99 = zeros(3,INREP);
decl  m_limInfBoot99 = zeros(3,INREP);
decl  m_SupInfBoot99 = zeros(3,INREP);
decl  m_limInfBoot99BCA = zeros(3,INREP);
decl  m_SupInfBoot99BCA = zeros(3,INREP);
decl  m_limInfBoot99Cox = zeros(3,INREP);
decl  m_SupInfBoot99Cox = zeros(3,INREP);
decl  m_limInfBoot99Firth = zeros(3,INREP);
decl  m_SupInfBoot99Firth = zeros(3,INREP);

//vetores das saidas para impressao
mResMean = zeros(4,3);
mResBias = zeros(4,3);
mResVar = zeros(4,3);
mResMSE = zeros(4,3);
mRelBias = zeros(4,3);

//Intervalos de confianca
decl  mResIC90 = zeros(6,3);
decl  mResIC95 = zeros(6,3);
decl  mResIC99 = zeros(6,3);
decl  mResIC90CS = zeros(6,3);
decl  mResIC95CS = zeros(6,3);
decl  mResIC99CS = zeros(6,3);
decl  mResIC90DF = zeros(6,3);
decl  mResIC95DF = zeros(6,3);
decl  mResIC99DF = zeros(6,3);
decl  mResIC90BT = zeros(6,3);
decl  mResIC95BT = zeros(6,3);
decl  mResIC99BT = zeros(6,3);
decl  mResIC90BTCox = zeros(6,3);
decl  mResIC95BTCox = zeros(6,3);
decl  mResIC99BTCox = zeros(6,3);
decl  mResIC90BTFirth = zeros(6,3);
decl  mResIC95BTFirth = zeros(6,3);
decl  mResIC99BTFirth = zeros(6,3);
decl  mResIC90BTBCA = zeros(6,3);
decl  mResIC95BTBCA = zeros(6,3);
decl  mResIC99BTBCA = zeros(6,3);

// impressao de resultados da simulacao
print( "\n\t\t DATA: ", date() );
print( "\n\t\t HORA: ", time() );
print( "\n\t\t PROGRAMA OX: ", oxfilename(0) );
print( "\n\t\t VERSSAO OX: ", oxversion() );
print( "\n\t\t |----- ESTIMACAO E CORRECAO DE VIES -----| " );
print( "\n\t\t |----- NUM MODELO DE REGRESSAO BETA -----| " );
print( "\n\t\t NUM. MIN. DE OBSERVA.: ", INMIN );
print( "\n\t\t NUM. MAX. OBSERVA.: ", INMAX );
print( "\n\t\t NUM. REPLICAS MONTE CARLO : ", INREP);
print( "\n\t\t NUM. REPLICAS BOOTSTRAP: ", B);
print( "\n\t\t PARAMETROS VERDADEIROS: ", "%12.4f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"THETA"}, v_theta );

```

```

// Loop para os diferentes tamanhos de amostra
for (cn = INMIN; cn <= INMAX; cn += 20)
{
    ranseed({1965, 2001});

    // matriz de dados explicativos
    //gerador dos regressores X
    // vx = 2 +(0.5)*rann(cn,1);
    // vx = ranexp(cn,1,3.0);
    // vx = ranu(cn,1);
    // vx = 0.5*rann(cn,1);
    // vx = rann(cn,1);

    //vetor das variaveis regressoras
    s_mX = ones(cn,1)~vx;

    npar=columns(s_mX); // numero de parametros do modelo

    //matriz de dados aumentada
    s_mXAumentada = zeros(cn+1,npar+1);
    s_mXAumentada[0:(cn-1)][0:(npar-1)] = s_mX;
    s_mXAumentada[cn][npar] = 1;

    d_denom = cn-npar;
    d_dem2 = (1/cn);
    v_eta = s_mX*vbeta; //regressores
    v_mu = exp(v_eta) ./ (1.0+exp(v_eta)); //vetor de medias
    m_inv = invtsym(s_mX'*s_mX)*s_mX'; //inversao auxiliar

    // inicializacao do contador de falhas da convergencia
    cfailure = 0;

    ranseed( -1 );

    // Loop de Monte Carlo
    for (cs = 0; cs < INREP; ++cs)
    {
        fgerador_vY(&gerY1,v_mu, d_phi, cn); //gerador da resposta Y
        s_vy=gerY1;

        //logaritmos do vetor Y para entrada da verossimilhanca
        s_vlogy = log(s_vy);
        s_vlogiminusy = log(1.0-s_vy);

        //regressao auxiliar para determinar os pontos iniciais no momento
        //de maximizar a funcao de log-verossimilhanca pelo metodo BFGS
        v_z = log(s_vy ./ (1.0 - s_vy)); //resposta transformada v_z
        v_betaini = m_inv*v_z; //estimacao inicial dos betas
        v_zEstim = s_mX*v_betaini; // vetor v_z estimado
        v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
        b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
        v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim)); //vetor mu transformado
        v_sigma2Est = b_SomaQua/(d_denom); //sigma estimado

        phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
            (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial

        v_thetaIni = v_betaini \ phi_ini; //vetor theta inicial para a maximizacao

        // valores de arranque
        vp1 = v_thetaIni;
        vp2 = v_thetaIni;

        //Maximizacao da funcao de log-verossimilhanca pelo Metodo BFGS

        // parametros da maximizacao por BFGS
        MaxControl( 50, -1 );

        // Estimacao por maxima verossimilhanca
        ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
        if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
        {
            //vetor theta estimado por maxima verossimilhanca
            mmle[cs] = vp1;
        }
    }
}

```

```

//inversa da matriz de informacion
finformacaoFisher(vp1, &dfunc); // informacao de Fisher inversa.
d_phi_est = vp1[npar];

//matriz de covariancias (variancia assintotica)
m_inforK = dfunc;
m_Kbetabeta = dfunc[0:(npar-1)][0:(npar-1)]; //K_bb^-1
v_Kbetaphi = dfunc[0:(npar-1)][npar]; //K_bphi^-1
d_Kphiphi = dfunc[npar][npar]; //K_phiphi^-1

//Calculo do vies por Cox & Snell

v_eta_est = s_mX*vp1[0:(npar-1)]; //regressores estimados
v_mu_est = exp(v_eta_est) ./ (1.0+exp(v_eta_est)); //vetor de medias estimado

//vetor DELTA Xiu
fdeltaXiu( &delta_Tilde, v_mu_est, d_phi_est, v_eta_est,
m_Kbetabeta, v_Kbetaphi, d_Kphiphi, m_inforK);

//Vetor do vies de segunda ordem de Cox & Snell estimado
v_Bias_theta=(m_inforK*s_mXAumentada')*delta_Tilde; //vetor do vies

//vetor do vies para o deslocamento da verossimilhanca (David Firth)
v_vies_desloca = s_mXAumentada'*delta_Tilde;

//solucao do sistema no linear U(theta*)=0
//(Obtencao de estimaticas corrigidas) D. Firth
SolveNLE(fscore, &vp2);

//Estimacao Bootstrap parametrico do vies e da
//assimetria da funcao de log-verossimilhanca
BootVies = funcReBoot(B, vp1, v_thetaIni, cn );

//vetor de reamostragem bootstrap
mBoot=BootVies[0:2] [];

//calculo da assimetria
assim1=moments((BootVies[3:5] [])',3,1);

//estimacao da constante de aceleracao
assi=((1/6)*assim1[3] [])',

// AmJK=funJK(cn, v_thetaIni ); //amostra Jakcnife

//reamostragem Bootstrap para o IC de Estimaticas corrigidas C. & S.
mBootCox=funcReBootCox(B, (vp1-v_Bias_theta), v_thetaIni, cn );

//reamostragem Bootstrap para o IC de Estimaticas corrigidas D. F.
mBootFirth=funcReBootFirth(B, vp2, v_thetaIni, cn );

//estimacao corrigida de Maxima verossimilhanca por Cox & Snell
mmlebc[] [cs] = mmle[] [cs] - v_Bias_theta;

//estimacao corrigida de Maxima verossimilhanca por D. Firth
mmDfirth[] [cs] = vp2;

//estimacao corrigida de Maxima verossimilhanca por Bootstrap
mBootMC[] [cs] = (2*vp1)-meanr(mBoot);

//Fator de correcao Vies do IC BCa
ZZZ00=fZOBCA((vp1)', mBoot', B);

//vetor de variancias assintoticas das EMV
v_varassin=diagonal(m_inforK);

//inversa da matriz de informacion C & S
finformacaoFisher(mmlcbc[] [cs], &a);
decl v_varassin1 = diagonal(a);

//inversa da matriz de informacion D. Firth
finformacaoFisher(mmDfirth[] [cs], &b);
decl v_varassin2 = diagonal(b);

//Intervalo de confianca assintotico Normal para EMV
decl ICCNorm90 = fICCN(v_varassin, z90,vp1);
decl ICCNorm95 = fICCN(v_varassin, z95,vp1);
decl ICCNorm99 = fICCN(v_varassin, z99,vp1);

//Intervalo de confianca assintotico Normal para EMV Corrigida Cox & Snell

```

```

decl ICC_CoxSnell90 = fICCN(v_varassin1, z90,(vp1-v_Bias_theta));
decl ICC_CoxSnell95 = fICCN(v_varassin1, z95,(vp1-v_Bias_theta));
decl ICC_CoxSnell99 = fICCN(v_varassin1, z99,(vp1-v_Bias_theta));

//Intervalo de confianca assintotico Normal para EMV Corrigida D. Firth
decl ICCFirth90 = fICCN(v_varassin2, z90,vp2);
decl ICCFirth95 = fICCN(v_varassin2, z95,vp2);
decl ICCFirth99 = fICCN(v_varassin2, z99,vp2);

//Intervalo de confianca Bootstrap Percentil
decl ICCBoot90 = fICBOOTPer(mBoot', 5.0 );
decl ICCBoot95 = fICBOOTPer(mBoot', 2.5 );
decl ICCBoot99 = fICBOOTPer(mBoot', 0.5 );

//Intervalo de confianca Bootstrap Percentil
decl ICCBoot90Cox = fICBOOTPer(mBootCox', 5.0 );
decl ICCBoot95Cox = fICBOOTPer(mBootCox', 2.5 );
decl ICCBoot99Cox = fICBOOTPer(mBootCox', 0.5 );

//Intervalo de confianca Bootstrap Percentil
decl ICCBoot90Firth = fICBOOTPer(mBootFirth', 5.0 );
decl ICCBoot95Firth = fICBOOTPer(mBootFirth', 2.5 );
decl ICCBoot99Firth = fICBOOTPer(mBootFirth', 0.5 );

//Intervalo de confianca BCa para as EMV
decl ICCBoot90BCA = fICBootAcelerado(mBoot, z90, ZZZ00,assi, B);
decl ICCBoot95BCA = fICBootAcelerado(mBoot, z95, ZZZ00,assi, B);
decl ICCBoot99BCA = fICBootAcelerado(mBoot, z99, ZZZ00,assi, B);

//Saidas ao 90%
//limites
m_limInfNorm90[] [cs] = ICCNorm90[] [0];
m_SupInfNorm90[] [cs] = ICCNorm90[] [1];
m_limInfCS90[] [cs] = ICC_CoxSnell90[] [0];
m_SupInfCS90[] [cs] = ICC_CoxSnell90[] [1];
m_limInfFirth90[] [cs] = ICCFirth90[] [0];
m_SupInfFirth90[] [cs] = ICCFirth90[] [1];
m_limInfBoot90[] [cs] = ICCBoot90[] [0];
m_SupInfBoot90[] [cs] = ICCBoot90[] [1];
m_limInfBoot90Cox[] [cs] = ICCBoot90Cox[] [0];
m_SupInfBoot90Cox[] [cs] = ICCBoot90Cox[] [1];
m_limInfBoot90Firth[] [cs] = ICCBoot90Firth[] [0];
m_SupInfBoot90Firth[] [cs] = ICCBoot90Firth[] [1];
m_limInfBoot90BCA[] [cs] = ICCBoot90BCA[] [0];
m_SupInfBoot90BCA[] [cs] = ICCBoot90BCA[] [1];

//cubrimentos
m_cubriNorm90[] [cs] = fCobertura(ICCNorm90[] [0], ICCNorm90[] [1]);
m_cubriCS90[] [cs] = fCobertura(ICC_CoxSnell90[] [0], ICC_CoxSnell90[] [1]);
m_cubriFirth90[] [cs] = fCobertura(ICCFirth90[] [0], ICCFirth90[] [1]);
m_cubriBoot90[] [cs] = fCobertura(ICCBoot90[] [0], ICCBoot90[] [1]);
m_cubriBoot90Cox[] [cs] = fCobertura(ICCBoot90Cox[] [0], ICCBoot90Cox[] [1]);
m_cubriBoot90Firth[] [cs] = fCobertura(ICCBoot90Firth[] [0], ICCBoot90Firth[] [1]);
m_cubriBoot90BCA[] [cs] = fCobertura(ICCBoot90BCA[] [0], ICCBoot90BCA[] [1]);

//saidas ao 95%
//limites
m_limInfNorm95[] [cs] = ICCNorm95[] [0];
m_SupInfNorm95[] [cs] = ICCNorm95[] [1];
m_limInfCS95[] [cs] = ICC_CoxSnell95[] [0];
m_SupInfCS95[] [cs] = ICC_CoxSnell95[] [1];
m_limInfFirth95[] [cs] = ICCFirth95[] [0];
m_SupInfFirth95[] [cs] = ICCFirth95[] [1];
m_limInfBoot95[] [cs] = ICCBoot95[] [0];
m_SupInfBoot95[] [cs] = ICCBoot95[] [1];
m_limInfBoot95Cox[] [cs] = ICCBoot95Cox[] [0];
m_SupInfBoot95Cox[] [cs] = ICCBoot95Cox[] [1];
m_limInfBoot95Firth[] [cs] = ICCBoot95Firth[] [0];
m_SupInfBoot95Firth[] [cs] = ICCBoot95Firth[] [1];
m_limInfBoot95BCA[] [cs] = ICCBoot95BCA[] [0];
m_SupInfBoot95BCA[] [cs] = ICCBoot95BCA[] [1];

//cubrimentos
m_cubriNorm95[] [cs] = fCobertura(ICCNorm95[] [0], ICCNorm95[] [1]);
m_cubriCS95[] [cs] = fCobertura(ICC_CoxSnell95[] [0], ICC_CoxSnell95[] [1]);
m_cubriFirth95[] [cs] = fCobertura(ICCFirth95[] [0], ICCFirth95[] [1]);
m_cubriBoot95[] [cs] = fCobertura(ICCBoot95[] [0], ICCBoot95[] [1]);
m_cubriBoot95Cox[] [cs] = fCobertura(ICCBoot90Cox[] [0], ICCBoot90Cox[] [1]);

```

```

m_cubriBoot95Firth[] [cs] = fCobertura(ICCBoot90Firth[] [0], ICCBoot90Firth[] [1]);
m_cubriBoot95BCA[] [cs] = fCobertura(ICCBoot95BCA[] [0], ICCBoot95BCA[] [1]);

//Saidas ao 99%
//limites
m_limInfNorm99[] [cs] = ICCNorm99[] [0];
m_SupInfNorm99[] [cs] = ICCNorm99[] [1];
m_limInfCS99[] [cs] = ICC_CoxSnell99[] [0];
m_SupInfCS99[] [cs] = ICC_CoxSnell99[] [1];
m_limInfFirth99[] [cs] = ICCFirth99[] [0];
m_SupInfFirth99[] [cs] = ICCFirth99[] [1];
m_limInfBoot99[] [cs] = ICCBoot99[] [0];
m_SupInfBoot99[] [cs] = ICCBoot99[] [1];
m_limInfBoot99Cox[] [cs] = ICCBoot99Cox[] [0];
m_SupInfBoot99Cox[] [cs] = ICCBoot99Cox[] [1];
m_limInfBoot99Firth[] [cs] = ICCBoot99Firth[] [0];
m_SupInfBoot99Firth[] [cs] = ICCBoot99Firth[] [1];
m_limInfBoot99BCA[] [cs] = ICCBoot99BCA[] [0];
m_SupInfBoot99BCA[] [cs] = ICCBoot99BCA[] [1];

//cubrimentos
m_cubriNorm99[] [cs] = fCobertura(ICCNorm99[] [0], ICCNorm99[] [1]);
m_cubriCS99[] [cs] = fCobertura(ICC_CoxSnell99[] [0], ICC_CoxSnell99[] [1]);
m_cubriFirth99[] [cs] = fCobertura(ICCFirth99[] [0], ICCFirth99[] [1]);
m_cubriBoot99[] [cs] = fCobertura(ICCBoot99[] [0], ICCBoot99[] [1]);
m_cubriBoot99Cox[] [cs] = fCobertura(ICCBoot99Cox[] [0], ICCBoot99Cox[] [1]);
m_cubriBoot99Firth[] [cs] = fCobertura(ICCBoot99Firth[] [0], ICCBoot99Firth[] [1]);
m_cubriBoot99BCA[] [cs] = fCobertura(ICCBoot99BCA[] [0], ICCBoot99BCA[] [1]);

}

else // contador de fallas e avance
{
++cfailure;
--cs;
}

}

// Impressao de resultados
print( "\n\nNUM. OBSERVA. DA AMOSTRA: ", cn );
print( "\n\nNUM. DE REPLICAS COM FALHAS ", cfailure );
print( "\n\nVALORES DE ARRANQUE BFGS: ", vpl );

//medias estimadas
mResMean[0] [] = (meanr(mmle))';
mResMean[1] [] = (meanr(mmlebc))';
mResMean[2] [] = (meanr(mmDfirth))';
mResMean[3] [] = (meanr(mBootMC))';

//vieses estimados
mResBias[0] [] = mResMean[0] [] -v_theta';
mResBias[1] [] = mResMean[1] [] -v_theta';
mResBias[2] [] = mResMean[2] [] -v_theta';
mResBias[3] [] = mResMean[3] [] -v_theta';

//variancias estimadas
mResVar[0] [] = (varr(mmle))';
mResVar[1] [] = (varr(mmlebc))';
mResVar[2] [] = (varr(mmDfirth))';
mResVar[3] [] = (varr(mBootMC))';

//Erro quadratico meio
mResMSE[0] [] = (meanr( (mmle-v_theta) .^ 2 ))';
mResMSE[1] [] = (meanr( (mmlebc-v_theta) .^ 2 ))';
mResMSE[2] [] = (meanr( (mmDfirth-v_theta) .^ 2 ))';
mResMSE[3] [] = (meanr( (mBootMC-v_theta) .^ 2 ))';

//vies relativo
mRelBias[0] [] = mResBias[0] [] ./ (v_theta)';
mRelBias[1] [] = mResBias[1] [] ./ (v_theta)';
mRelBias[2] [] = mResBias[2] [] ./ (v_theta)';
mRelBias[3] [] = mResBias[3] [] ./ (v_theta)';

//Saidas para IC Normal 90%
mResIC90[0] [] = (meanr(m_limInfNorm90))';
mResIC90[1] [] = (meanr(m_SupInfNorm90))';
mResIC90[2] [] = mResIC90[1] [] -mResIC90[0] [];
mResIC90[3] [] = 100*(meanr(m_cubriNorm90))';
mResIC90[4:5] [] = funTamLateral(m_limInfNorm90, m_SupInfNorm90, parametro, INREP)';

```

```

//Saidas para IC Normal 95%
mResIC95[0][] = (meanr(m_limInfNorm95))';
mResIC95[1][] = (meanr(m_SupInfNorm95))';
mResIC95[2][] = mResIC95[1][]-mResIC95[0][];
mResIC95[3][] = 100*(meanr(m_cubriNorm95))';
mResIC95[4:5][] = funTamLateral(m_limInfNorm95, m_SupInfNorm95, parametro, INREP)';

//Saidas para IC Normal 99%
mResIC99[0][] = (meanr(m_limInfNorm99))';
mResIC99[1][] = (meanr(m_SupInfNorm99))';
mResIC99[2][] = mResIC99[1][]-mResIC99[0][];
mResIC99[3][] = 100*(meanr(m_cubriNorm99))';
mResIC99[4:5][] = funTamLateral(m_limInfNorm99, m_SupInfNorm99, parametro, INREP)';

//Saidas para IC CS 90%
mResIC90CS[0][] = (meanr(m_limInfCS90))';
mResIC90CS[1][] = (meanr(m_SupInfCS90))';
mResIC90CS[2][] = mResIC90CS[1][]-mResIC90CS[0][];
mResIC90CS[3][] = 100*(meanr(m_cubriCS90))';
mResIC90CS[4:5][] = funTamLateral(m_limInfCS90, m_SupInfCS90, parametro, INREP)';

//Saidas para IC CS 95%
mResIC95CS[0][] = (meanr(m_limInfCS95))';
mResIC95CS[1][] = (meanr(m_SupInfCS95))';
mResIC95CS[2][] = mResIC95CS[1][]-mResIC95CS[0][];
mResIC95CS[3][] = 100*(meanr(m_cubriCS95))';
mResIC95CS[4:5][] = funTamLateral(m_limInfCS95, m_SupInfCS95, parametro, INREP)';

//Saidas para IC CS 99%
mResIC99CS[0][] = (meanr(m_limInfCS99))';
mResIC99CS[1][] = (meanr(m_SupInfCS99))';
mResIC99CS[2][] = mResIC99CS[1][]-mResIC99CS[0][];
mResIC99CS[3][] = 100*(meanr(m_cubriCS99))';
mResIC99CS[4:5][] = funTamLateral(m_limInfCS99, m_SupInfCS99, parametro, INREP)';

//Saidas para IC DF 90%
mResIC90DF[0][] = (meanr(m_limInfFirth90))';
mResIC90DF[1][] = (meanr(m_SupInfFirth90))';
mResIC90DF[2][] = mResIC90DF[1][]-mResIC90DF[0][];
mResIC90DF[3][] = 100*(meanr(m_cubriFirth90))';
mResIC90DF[4:5][] = funTamLateral(m_limInfFirth90, m_SupInfFirth90, parametro, INREP)';

//Saidas para IC DF 95%
mResIC95DF[0][] = (meanr(m_limInfFirth95))';
mResIC95DF[1][] = (meanr(m_SupInfFirth95))';
mResIC95DF[2][] = mResIC95DF[1][]-mResIC95DF[0][];
mResIC95DF[3][] = 100*(meanr(m_cubriFirth95))';
mResIC95DF[4:5][] = funTamLateral(m_limInfFirth95, m_SupInfFirth95, parametro, INREP)';

//Saidas para IC DF 99%
mResIC99DF[0][] = (meanr(m_limInfFirth99))';
mResIC99DF[1][] = (meanr(m_SupInfFirth99))';
mResIC99DF[2][] = mResIC99DF[1][]-mResIC99DF[0][];
mResIC99DF[3][] = 100*(meanr(m_cubriFirth99))';
mResIC99DF[4:5][] = funTamLateral(m_limInfFirth99, m_SupInfFirth99, parametro, INREP)';

//Saidas para IC Boot 90%
mResIC90BT[0][] = (meanr(m_limInfBoot90))';
mResIC90BT[1][] = (meanr(m_SupInfBoot90))';
mResIC90BT[2][] = mResIC90BT[1][]-mResIC90BT[0][];
mResIC90BT[3][] = 100*(meanr(m_cubriBoot90))';
mResIC90BT[4:5][] = funTamLateral(m_limInfBoot90, m_SupInfBoot90, parametro, INREP)';

//Saidas para IC Boot 95%
mResIC95BT[0][] = (meanr(m_limInfBoot95))';
mResIC95BT[1][] = (meanr(m_SupInfBoot95))';
mResIC95BT[2][] = mResIC95BT[1][]-mResIC95BT[0][];
mResIC95BT[3][] = 100*(meanr(m_cubriBoot95))';
mResIC95BT[4:5][] = funTamLateral(m_limInfBoot95, m_SupInfBoot95, parametro, INREP)';

//Saidas para IC Boot 99%
mResIC99BT[0][] = (meanr(m_limInfBoot99))';
mResIC99BT[1][] = (meanr(m_SupInfBoot99))';
mResIC99BT[2][] = mResIC99BT[1][]-mResIC99BT[0][];
mResIC99BT[3][] = 100*(meanr(m_cubriBoot99))';
mResIC99BT[4:5][] = funTamLateral(m_limInfBoot99, m_SupInfBoot99, parametro, INREP)';

//Saidas para IC Boot Cox 90%
mResIC90BTCox[0][] = (meanr(m_limInfBoot90Cox))';

```

```

mResIC90BTCox[1] [] = (meanr(m_SupInfBoot90))';
mResIC90BTCox[2] [] = mResIC90BTCox[1] [] - mResIC90BTCox[0] [];
mResIC90BTCox[3] [] = 100*(meanr(m_cubriBoot90Cox))';
mResIC90BTCox[4:5] [] = funTamLateral(m_limInfBoot90Cox, m_SupInfBoot90Cox, parametro, INREP)';

//Saidas para IC Boot Cox 95%
mResIC95BTCox[0] [] = (meanr(m_limInfBoot95Cox))';
mResIC95BTCox[1] [] = (meanr(m_SupInfBoot95Cox))';
mResIC95BTCox[2] [] = mResIC95BTCox[1] [] - mResIC95BTCox[0] [];
mResIC95BTCox[3] [] = 100*(meanr(m_cubriBoot95Cox))';
mResIC95BTCox[4:5] [] = funTamLateral(m_limInfBoot95Cox, m_SupInfBoot95Cox, parametro, INREP)';

//Saidas para IC Boot Cox 99%
mResIC99BTCox[0] [] = (meanr(m_limInfBoot99Cox))';
mResIC99BTCox[1] [] = (meanr(m_SupInfBoot99Cox))';
mResIC99BTCox[2] [] = mResIC99BTCox[1] [] - mResIC99BTCox[0] [];
mResIC99BTCox[3] [] = 100*(meanr(m_cubriBoot99Cox))';
mResIC99BTCox[4:5] [] = funTamLateral(m_limInfBoot99Cox, m_SupInfBoot99Cox, parametro, INREP)';

//Saidas para IC Boot Firth 90%
mResIC90BTFirth[0] [] = (meanr(m_limInfBoot90Firth))';
mResIC90BTFirth[1] [] = (meanr(m_SupInfBoot90Firth))';
mResIC90BTFirth[2] [] = mResIC90BTFirth[1] [] - mResIC90BTFirth[0] [];
mResIC90BTFirth[3] [] = 100*(meanr(m_cubriBoot90Firth))';
mResIC90BTFirth[4:5] [] = funTamLateral(m_limInfBoot90Firth, m_SupInfBoot90Firth, parametro, INREP)';

//Saidas para IC Boot Firth 95%
mResIC95BTFirth[0] [] = (meanr(m_limInfBoot95Firth))';
mResIC95BTFirth[1] [] = (meanr(m_SupInfBoot95Firth))';
mResIC95BTFirth[2] [] = mResIC95BTFirth[1] [] - mResIC95BTFirth[0] [];
mResIC95BTFirth[3] [] = 100*(meanr(m_cubriBoot95Firth))';
mResIC95BTFirth[4:5] [] = funTamLateral(m_limInfBoot95Firth, m_SupInfBoot95Firth, parametro, INREP)';

//Saidas para IC Boot Firth 99%
mResIC99BTFirth[0] [] = (meanr(m_limInfBoot99Firth))';
mResIC99BTFirth[1] [] = (meanr(m_SupInfBoot99Firth))';
mResIC99BTFirth[2] [] = mResIC99BTFirth[1] [] - mResIC99BTFirth[0] [];
mResIC99BTFirth[3] [] = 100*(meanr(m_cubriBoot99Firth))';
mResIC99BTFirth[4:5] [] = funTamLateral(m_limInfBoot99Firth, m_SupInfBoot99Firth, parametro, INREP)';

//Saidas para IC Boot BCA 90%
mResIC90BTBCA[0] [] = (meanr(m_limInfBoot90BCA))';
mResIC90BTBCA[1] [] = (meanr(m_SupInfBoot90BCA))';
mResIC90BTBCA[2] [] = mResIC90BTBCA[1] [] - mResIC90BTBCA[0] [];
mResIC90BTBCA[3] [] = 100*(meanr(m_cubriBoot90BCA))';
mResIC90BTBCA[4:5] [] = funTamLateral(m_limInfBoot90BCA, m_SupInfBoot90BCA, parametro, INREP)';

//Saidas para IC Boot BCA 95%
mResIC95BTBCA[0] [] = (meanr(m_limInfBoot95BCA))';
mResIC95BTBCA[1] [] = (meanr(m_SupInfBoot95BCA))';
mResIC95BTBCA[2] [] = mResIC95BTBCA[1] [] - mResIC95BTBCA[0] [];
mResIC95BTBCA[3] [] = 100*(meanr(m_cubriBoot95BCA))';
mResIC95BTBCA[4:5] [] = funTamLateral(m_limInfBoot95BCA, m_SupInfBoot95BCA, parametro, INREP)';

//Saidas para IC Boot BCA 99%
mResIC99BTBCA[0] [] = (meanr(m_limInfBoot99BCA))';
mResIC99BTBCA[1] [] = (meanr(m_SupInfBoot99BCA))';
mResIC99BTBCA[2] [] = mResIC99BTBCA[1] [] - mResIC99BTBCA[0] [];
mResIC99BTBCA[3] [] = 100*(meanr(m_cubriBoot99BCA))';
mResIC99BTBCA[4:5] [] = funTamLateral(m_limInfBoot99BCA, m_SupInfBoot99BCA, parametro, INREP)';

//impresao dos resultados
print("\n |----Resultados respeito aos parametros estimados----|\n");

print( "\nMEAN ESTIMATES: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, mResMean );
print( "\nESTIMATED BIASES: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, mResBias );
print( "\nRELATIVE BIASES: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, mRelBias );
print( "\nESTIMATED VARIANCES: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, mResVar );
print( "\nESTIMATED MSE'S: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, mResMSE );
print( "\nESTIMATED RMSE'S: ", "%12.6f", "%c", {"beta0", "beta1",
"phi"}, "%r", {"MLE", "Cox&Snell", "DFirth", "Boot"}, sqrt(mResMSE) );

print("\n |----Resultados respeito aos Intervalos de confianca----|\n");

```


B.2. Biblioteca de funções

```

/*****
* PROGRAM: verossim.ox
* AUTHOR: Raydonal Ospina Martinez
* DATE: June 13, 2003.
* VERSION: 0.1.5
* LAST MODIFIED: Novembro 05, 2003. Hora: 11:30
*****/

/* header files */
#include <oxstd.h>
#include <oxfloat.h>
#include <oxprob.h>
#import <maximize>
#import <solvenle>

/* global variables */
static decl s_vy;
static decl s_mX;
static decl s_mXX;
static decl s_mXAumentada ;
static decl s_vlogy;
static decl s_vlog1minusy;
static decl v_vies_desloca;
static decl v_vies_boot;
static decl s_vy_boot;
static decl infoBoot;
static decl parametro;
static decl g_mY;

//funcao de log-verossimilhanca do modelo de regressao Beta
floglik(const vP, const adFunc, const avScore, const amHess)
{
    decl k = rows(vP) - 1;
    decl eta = s_mX*vP[0:(k-1)];
    decl mu = exp(eta) ./ (1.0+exp(eta));
    decl phi = vP[k];
    decl ynew = log( s_vy ./ (1.0-s_vy) );
    decl munew = polygamma(mu*phi, 0) - polygamma((1.0-mu)*phi, 0);
    decl T = diag( exp(eta) ./ (1.0+exp(eta)) .^2 );

    adFunc[0] = double( sumc( loggamma(phi) - loggamma(mu*phi)
        - loggamma((1-mu)*phi) + (mu*phi-1) .* log(s_vy)
        + ( (1-mu)*phi-1 ) .* log(1-s_vy) ));

    //Derivadas analiticas da funcao de log-verossimilhanca
    if(avScore)
    {
        (avScore[0])[0:(k-1)] = phi*s_mX'*T*(ynew-munew);
        (avScore[0])[k] = double(sumc( polygamma(phi, 0) - mu .*
            polygamma(mu*phi, 0) - (1.0-mu) .* polygamma( (1.0-mu)*phi, 0) +
            mu .* log(s_vy) + (1.0-mu) .* log(1-s_vy) ));
    }

    if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
    return 0;
    else
    return 1; // 1 indica sucesso
}

//Funcao de log-verossimilhanca para o reamostragem Jackknife
floglikJ(const vP, const adFunc, const avScore, const amHess)
{
    decl k = rows(vP) - 1;
    decl eta = s_mXX*vP[0:(k-1)];
    decl mu = exp(eta) ./ (1.0+exp(eta));
    decl phi = vP[k];
    decl ynew = log( g_mY ./ (1.0-g_mY) );
    decl munew = polygamma(mu*phi, 0) - polygamma((1.0-mu)*phi, 0);
    decl T = diag( exp(eta) ./ (1.0+exp(eta)) .^2 );

    adFunc[0] = double( sumc( loggamma(phi) - loggamma(mu*phi)
        - loggamma((1-mu)*phi) + (mu*phi-1) .* log(g_mY)
        + ( (1-mu)*phi-1 ) .* log(1-g_mY) ));

    //Derivadas analiticas da funcao de log-verossimilhanca

```

```

    if(avScore)
    {
        (avScore[0])[0:(k-1)] = phi*s_mXX'*T*(ynew-munew);
        (avScore[0])[k] = double(sumc( polygamma(phi, 0) - mu .*
        polygamma(mu*phi, 0) - (1.0-mu) .* polygamma( (1.0-mu)*phi, 0) +
        mu .* log(g_mY) + (1.0-mu) .* log(1.0-g_mY) ));
    }

    if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
    return 0;
    else
    return 1; // 1 indica sucesso
}

//Calculo da matriz de informacao e sua inversa (variancia assintotica)
finformacaoFisher(const vpi, const adFunc)
{ //notacao arquivo de Ferrari & Cribari--Neto (2003).

    decl p = columns(s_mX);
    decl etahat = s_mX*vpi[0:(p-1)];
    decl muhat = exp(etahat) ./ (1+exp(etahat));
    decl phihat = vpi[p];
    decl psi1 = polygamma(muhat*phihat, 1);
    decl psi2 = polygamma((1.0-muhat)*phihat, 1);
    decl T = diag( exp(etahat) ./ (1.0+exp(etahat)) .^2 );
    decl W = diag(phihat*(psi1+psi2)) * T .^2;
    decl vc = phihat*(psi1.*muhat-psi2.*(1-muhat));
    decl D = diag(psi1.*(muhat.^2)+psi2.*(1.0-muhat).^2-polygamma(phihat,1));
    decl tempinv = invsym(s_mX'*W*s_mX);
    decl g = trace(D)-(1.0/phihat)*vc'*T'*s_mX*tempinv*s_mX'*T*vc;
    decl K1 = tempinv*(g*unit(p)+(1/phihat)*s_mX'*T*vc*vc'*T'*s_mX*tempinv);
    decl K2 = -tempinv*s_mX'*T*vc;
    decl fisherinv = (1/(phihat*g))*((K1~K2)|(-vc'*T'*s_mX*tempinv~phihat));

    adFunc[0]= fisherinv;

    return 1;
}

//funcao para calcular o vetor delta Xiu que serve para calcular o vies.
fdeltaXiu( const adFunc, const v_mu_est, const d_phi_est,
const v_eta_est, const m_Kbetabeta, const v_Kbetaphi,
const d_Khiphi, const m_inforK)
{
    decl wt = polygamma(v_mu_est*d_phi_est, 1)+polygamma((1-v_mu_est)*d_phi_est, 1);
    decl mt = polygamma(v_mu_est*d_phi_est, 2)-polygamma((1-v_mu_est)*d_phi_est, 2);
    decl d_mu_d_eta = exp(v_eta_est) ./((1+exp(v_eta_est)).^2);
    decl D_d_mu_d_eta = (exp(v_eta_est).*(1-exp(v_eta_est)))./((1+exp(v_eta_est)).^3);
    decl at = 3*((D_d_mu_d_eta ).*(d_mu_d_eta .^2));
    decl m_W1 = diag((-0.5*(d_phi_est^2))*(d_phi_est*(mt .*(d_mu_d_eta .^3))+
    ((1.0/3.0)*(wt .*at ))));
    decl ut = (-d_phi_est)*(2*wt+d_phi_est*((v_mu_est .* mt)
    +polygamma((1-v_mu_est)*d_phi_est, 2)));
    decl ct = d_phi_est*((v_mu_est .* wt)-polygamma((1-v_mu_est)*d_phi_est, 1));
    decl m_W2 = diag( 0.5*((ut .* d_mu_d_eta)+(ct .* D_d_mu_d_eta)).* d_mu_d_eta);
    decl m_W3 = diag((-0.5)*((d_phi_est^2)*((v_mu_est .*mt)
    +polygamma((1-v_mu_est)*d_phi_est, 2)) .* d_mu_d_eta)
    +(ct .* D_d_mu_d_eta )).* d_mu_d_eta);
    decl m_W4 = diag((-0.5*d_phi_est)*((v_mu_est .^2) .* polygamma((v_mu_est)*d_phi_est, 2)
    -((1-v_mu_est) .^2) .* polygamma((1-v_mu_est)*d_phi_est, 2)) .* d_mu_d_eta );
    decl parcel1 = 2*((v_mu_est .* wt)-polygamma((1-v_mu_est)*d_phi_est, 1));
    decl parcel2 = d_phi_est*((v_mu_est .^2) .* polygamma((v_mu_est)*d_phi_est, 2)
    -((1-v_mu_est) .^2) .* polygamma((1-v_mu_est)*d_phi_est, 2)));
    decl rt = parcel1+parcel2;
    decl m_W5 = diag( (-0.5)*(rt.*d_mu_d_eta));
    decl st = (((1-v_mu_est) .^3) .*polygamma((1-v_mu_est)*d_phi_est, 2)
    +((v_mu_est) .^3) .*polygamma((v_mu_est)*d_phi_est, 2) )-polygamma(d_phi_est, 2);
    decl m_S = diag( (-0.5)*st);
    decl v_deltabetabeta = diagonal(s_mX*m_Kbetabeta*s_mX');
    decl termo1_delta = (m_W1*v_deltabetabeta')+(m_W2+m_W3)*s_mX*v_Kbetaphi
    +(d_Khiphi*diagonal(m_W4)');

    decl termo2_delta = trace(m_W3*s_mX*m_Kbetabeta*s_mX')+(d_Khiphi*trace(m_S))+
    diagonal(m_W4+m_W5)*s_mX*v_Kbetaphi;
    decl delta_Tilde = termo1_delta|termo2_delta; //widetilde/delta

```

```

        adFunc[0] = delta_Tilde;
return 1; // 1 indica sucesso
}

//funcao escore modificada de David Firth
fscore(const avScore, const vP)
{
    //funco de ligaco utilizada logit g(x)=log(x/1-x)
    decl i_Nobs = rows(s_mX); //numero de observacoes
    decl i_Npar = (rows(vP) - 1); //numero de parametros
    decl v_eta = s_mX*vP[0:(i_Npar-1)]; //regressores
    decl v_mu = exp(v_eta) ./ (1.0+exp(v_eta)); //vetor de medias
    decl d_phi = vP[i_Npar]; //parametro de dispersao
    decl v_yast = log(s_vy ./ (1.0-s_vy)); // yast=log(y/1-y)
    decl v_muast = polygamma(v_mu*d_phi, 0)-polygamma((1-v_mu)*d_phi, 0); //media modificada
    decl m_T = diag( exp(v_eta) ./ (1.0+exp(v_eta)) .^2 ); //matriz diagonal primeiras derivadas de eta
    decl v_one = ones(1,rows(s_mX));
    decl v_Ubeta = d_phi*s_mX'*m_T*(v_yast-v_muast); // vetor escore de beta U(beta)
    decl v_h = log(1-s_vy)-polygamma((1-v_mu)*d_phi, 0); //vetor auxiliar
    decl d_diGphi = i_Nobs*polygamma(d_phi, 0); // digama de phi*N
    decl d_Uphi = (((v_yast-v_muast)')*diag(v_mu)+v_h')*ones(i_Nobs,1)+d_diGphi; // U(phi)
    decl v_U = v_Ubeta| d_Uphi;

    avScore[0] = v_U+v_vies_desloca;

    return 1; // 1 indica sucesso
}

//funcao de log-verossimilhanca do modelo de regressao beta para o
// calculo Bootstrap
floglikBoot(const vP, const adFunc, const avScore, const amHess)
{
    decl k = rows(vP) - 1;
    decl eta = s_mX*vP[0:(k-1)];
    decl mu = exp(eta) ./ (1.0+exp(eta));
    decl phi = vP[k];
    decl ynew = log( s_vy_boot ./ (1.0-s_vy_boot) );
    decl mnew = polygamma(mu*phi, 0) - polygamma((1.0-mu)*phi, 0);
    decl T = diag( exp(eta) ./ (1.0+exp(eta)) .^2 );

    adFunc[0] = double( sumc( loggamma(phi) - loggamma(mu*phi)
        - loggamma((1-mu)*phi) + (mu*phi-1) .* log(s_vy_boot)
        + ( (1-mu)*phi-1 ) .* log(1-s_vy_boot) ));

    //Derivadas analiticas
    if(avScore)
    {
        (avScore[0])[0:(k-1)] = phi*s_mX'*T*(ynew-mnew);
        (avScore[0])[k] = double(sumc( polygamma(phi, 0) - mu .*
polygamma(mu*phi, 0) - (1.0-mu) .* polygamma( (1.0-mu)*phi, 0) +
mu .* log(s_vy_boot) + (1.0-mu) .* log(1.0-s_vy_boot) ));

    }

    if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
return 0;
    else
return 1; // 1 indica sucesso
}

//construcao do intervalo de confianca assintotico Normal
fICCN(const diagFisher, const qantil, const vp)
{
    decl k=rows(vp);
    decl m_ICC = zeros(k,2);
    decl liminf= vp'-(qantil*sqrt(diagFisher));
    decl limsup= vp'+(qantil*sqrt(diagFisher));

    m_ICC[] [0]=liminf';
    m_ICC[] [1]=limsup';

    return m_ICC;
}

```

```

//calculo da cobertura do intervalo
fCobertura(const limInf, const limSup)
{
    decl cobert= (limInf.<= parametro .&& parametro .<=limSup);
    return cobert;
}

//construcao do intervalo de confianca por Bootstrap percentil
fICBOOTPer(const vreamostra, const nivel )
{
    decl dims=rows(vreamostra);
    decl vlimICB=zeros(3,2);
    decl vordem=sortc(vreamostra);
    decl posica=dims*nivel;

    if(imod(posica,100)==0)
    {
        decl perc=posica/100;

        vlimICB[0][0] = vordem[perc-1][0]; //lim. inf b0
        vlimICB[1][0] = vordem[perc-1][1]; //lim. inf b1
        vlimICB[2][0] = vordem[perc-1][2]; //lim. inf phi
        vlimICB[0][1] = vordem[dims-perc-1][0]; //lim. sup b0
        vlimICB[1][1] = vordem[dims-perc-1][1]; //lim. sup b1
        vlimICB[2][1] = vordem[dims-perc-1][2]; //lim. sup phi

        return vlimICB;
    }

    else
    {
        decl perc = floor((posica+nivel)/100);
        vlimICB[0][0] = vordem[perc-1][0]; //lim. inf b0
        vlimICB[1][0] = vordem[perc-1][1]; //lim. inf b1
        vlimICB[2][0] = vordem[perc-1][2]; //lim. inf phi
        vlimICB[0][1] = vordem[dims-perc-1][0]; //lim. sup b0
        vlimICB[1][1] = vordem[dims-perc-1][1]; //lim. sup b1
        vlimICB[2][1] = vordem[dims-perc-1][2]; //lim. sup phi
        return vlimICB;
    }
}

//funcao escore modificado de David Firth para o calculo do vies Bootstrap
fScoreDFBT(const avScore, const vP)
{
    //funco de ligaco utilizada logit g(x)=log(x/1-x)
    decl i_Nobs = rows(s_mX); //numero de observacoes
    decl i_Npar = (rows(vP) - 1); //numero de parametros
    decl v_eta = s_mX*vP[0:(i_Npar-1)]; //regressores
    decl v_mu = exp(v_eta) ./ (1.0+exp(v_eta)); //vetor de medias
    decl d_phi = vP[i_Npar]; //parametro de dispersao
    decl v_yast = log(s_vy) ./ (1.0-s_vy); // yast=log(y/1-y)
    decl v_muast = polygamma(v_mu*d_phi, 0)-polygamma((1-v_mu)*d_phi, 0); //media modificada
    decl m_T = diag( exp(v_eta) ./ (1.0+exp(v_eta)) .^2 ); //matriz diagonal primeiras derivadas de eta
    decl v_one = ones(1,rows(s_mX));
    decl v_Ubeta = d_phi*s_mX'*m_T*(v_yast-v_muast); // vetor score de beta U(beta)
    decl v_h = log(1-s_vy)-polygamma((1-v_mu)*d_phi, 0); //vetor auxiliar
    decl d_diGphi = i_Nobs*polygamma(d_phi, 0); // digama de phi*N
    decl d_Uphi = (((v_yast-v_muast)')*diag(v_mu))+v_h')*ones(i_Nobs,1)+d_diGphi; // U(phi)
    decl v_U = v_Ubeta| d_Uphi;

    avScore[0] = v_U+(infoBoot*v_vies_boot);

    return 1; // 1 indica sucesso
}

//funcao geradora de Y
fgerador_vY(const func, const v_mu, const d_phi, const cn)
{
    decl ci;
    decl vecY=zeros(cn,1);
    for ( ci = 0; ci < cn; ci++)
    {
        decl p = (v_mu[ci]*d_phi);
        decl q = ((1-v_mu[ci])*d_phi);
    }
}

```

```

        vecY[ci] = ranbeta(1,1,p,q);
    }
    func[0]=vecY;
    return 1;
}

//ciclo bootstrap Cox & Snell
funcReBootCox(const B, const vp1, const v_thetaIni, const cn )
{
    decl bi,gerY,dfunc3,vp3,ir3;
    decl cfailureboot=0;
    decl mBoot = zeros(3,B);

    for(bi=0; bi < B; ++bi)
    {
        decl v_eta_boot = s_mX*vp1[0:1]; //regressores
        decl v_mu_boot = exp(v_eta_boot) ./ (1.0+exp(v_eta_boot)); //vetor de medias
        decl d_phi_boot = vp1[2]; //parametro de precisao

        // Geracao do vetor Y de reamostragem
        fgerador_vY(&gerY,v_mu_boot, d_phi_boot, cn);
        s_vy_boot = gerY;

        vp3 = v_thetaIni; //valores iniciais dos parametros Bootstrap

        // estimacao por maxima verossimilhanca para o metodo Bootstrap
        ir3 = MaxBFGS(floglikBoot, &vp3, &dfunc3, 0, FALSE);

        // control da convergencia na estimacao Bootstrap
        if( (ir3 == MAX_CONV) || (ir3 == MAX_WEAK_CONV) )
        {
            mBoot[][bi] = vp3;
        }
        else
        {
            ++cfailureboot;
            --bi;
        }
    }

    return mBoot;
}

//ciclo bootstrap para D. Firth
funcReBootFirth(const B, const vp1, const v_thetaIni, const cn )
{
    decl bi,gerY,dfunc3,vp3,ir3;
    decl cfailureboot=0;
    decl mBoot = zeros(3,B);

    for(bi=0; bi < B; ++bi)
    {
        decl v_eta_boot = s_mX*vp1[0:1]; //regressores
        decl v_mu_boot = exp(v_eta_boot) ./ (1.0+exp(v_eta_boot)); //vetor de medias
        decl d_phi_boot = vp1[2]; //parametro de precisao

        // Geracao do vetor Y de reamostragem
        fgerador_vY(&gerY,v_mu_boot, d_phi_boot, cn);
        s_vy_boot = gerY;

        vp3 = v_thetaIni; //valores iniciais dos parametros Bootstrap

        // estimacao por maxima verossimilhanca para o Bootstrap
        ir3 = MaxBFGS(floglikBoot, &vp3, &dfunc3, 0, FALSE);

        // control da convergencia na estimacao Bootstrap
        if( (ir3 == MAX_CONV) || (ir3 == MAX_WEAK_CONV) )
        {
            mBoot[][bi] = vp3;
        }
        else
        {
            ++cfailureboot;
            --bi;
        }
    }
}

```

```

    return mBoot;
}

//intervalo de confianca de tipo BCa
fICBootAcelerado(const vreamostra, const d_quantil, const vZ, const vA, const B)
{
    decl cc,cr,ct,saida;
    ct=0;

    //inicializacao da matriz de percentis da distribuicao bootstrap dos parametros
    decl mPer=zeros(2,3);

    //ordenacao das observacoes
    decl m0=sortc(vreamostra');

    //inicializacao do vetor dos limites do intervalo de confianca dos parametros
    decl vlmt=zeros(1,6);

    //percentil inferior b0
    mPer[0][0]=probn( vZ[0][0]+(( vZ[0][0]-d_quantil)/(1-vA[0][0]*(vZ[0][0]-d_quantil))));

    //percentil inferior b0
    mPer[1][0]=probn( vZ[0][0]+(( vZ[0][0]+d_quantil)/(1-vA[0][0]*(vZ[0][0]+d_quantil))));

    //percentil inferior b1
    mPer[0][1]=probn( vZ[1][0]+(( vZ[1][0]-d_quantil)/(1-vA[1][0]*(vZ[1][0]-d_quantil))));

    //percentil inferior b1
    mPer[1][1]=probn( vZ[1][0]+(( vZ[1][0]+d_quantil)/(1-vA[1][0]*(vZ[1][0]+d_quantil))));

    //percentil inferior phi
    mPer[0][2]=probn( vZ[2][0]+(( vZ[2][0]-d_quantil)/(1-vA[2][0]*(vZ[2][0]-d_quantil))));

    //percentil inferior phi
    mPer[1][2]=probn( vZ[2][0]+(( vZ[2][0]+d_quantil)/(1-vA[2][0]*(vZ[2][0]+d_quantil))));

    //obtencao dos limites dos intervalos de confianca dos parametros
    for(cc=0;cc<3;++cc)
    {
    for(cr=0;cr<2;++cr)
    {
    decl ipr = B*mPer[cr][cc];
    decl id = floor(ipr);
    if( (ipr-id)==0 )
    {
    if(ipr>=1)
    vlmt[0][ct] = m0[ipr-1][cc];
    else
    vlmt[0][ct] = m0[ipr][cc];
    }
    else
    {
    if(ipr>=1)
    vlmt[0][ct]=m0[ipr-1+cr][cc];
    else
    vlmt[0][ct]=m0[ipr+cr][cc];
    }
    ++ct;
    }
    }

    saida =(vlmt[0][0] | vlmt[0][2] | vlmt[0][4])~(vlmt[0][1] | vlmt[0][3] | vlmt[0][5]);

    return saida;
}

//funcao reamostragem Jackknife
funJK(const cn, const v_thetaIni )
{
    decl cl,dfuncjk,mEjMV,vro,irj,vIdx;
    mEjMV=zeros(cn,3);

```

```

    for (cl=0;cl<cn;++cl)
    {
        vIdx=ones(cn,1);
        vIdx[cl][0]=0;
        s_mXX=selectifr(s_mX,vIdx);
        g_mY=selectifr(s_vy,vIdx);

        vro=v_thetaIni;

        // parametros da maximizacao por BFGS
        MaxControl( 50, -1 );

        irj= MaxBFGS(floglikJ, &vro, &dfuncjk, 0, FALSE);
        mEjMV[cl][]=vro';
    }

    return mEjMV; //estimativas para as replicas jackknife
}

//calculo da constante de aceleracao
faICBCA(const mEJ, const cn)
{
    decl vNum, vDev, vAj;

    vNum=vDev=vAj=zeros(3,1);

    vNum[0][0]=sumc((meanc(mEJ[][0])*ones(cn,1)-mEJ[][0]) .^3);
    vNum[1][0]=sumc((meanc(mEJ[][1])*ones(cn,1)-mEJ[][1]) .^3);
    vNum[2][0]=sumc((meanc(mEJ[][2])*ones(cn,1)-mEJ[][2]) .^3);

    vDev[0][0]=6*(sumc((meanc(mEJ[][0])*ones(cn,1)-mEJ[][0]) .^2))^(3/2);
    vDev[1][0]=6*(sumc((meanc(mEJ[][1])*ones(cn,1)-mEJ[][1]) .^2))^(3/2);
    vDev[2][0]=6*(sumc((meanc(mEJ[][2])*ones(cn,1)-mEJ[][2]) .^2))^(3/2);

    vAj[0][0]=vNum[0][0]/vDev[0][0];
    vAj[1][0]=vNum[1][0]/vDev[1][0];
    vAj[2][0]=vNum[2][0]/vDev[2][0];

    return vAj;
}

//Calculo de la constante Zo de correcao para O IC BCa
fZOBCA(const vE, const mEB, const B)
{
    decl vRz, vZ;
    decl mCMP=zeros(B,3);
    vRz=vZ=zeros(3,1);

    mCMP[][0]=ones(B,1)*vE[][0];
    mCMP[][1]=ones(B,1)*vE[][1];
    mCMP[][2]=ones(B,1)*vE[][2];

    vRz[0][0]=rows(selectifr(mEB[][0],mEB[][0].<mCMP[][0]))/B;
    vRz[1][0]=rows(selectifr(mEB[][1],mEB[][1].<mCMP[][1]))/B;
    vRz[2][0]=rows(selectifr(mEB[][2],mEB[][2].<mCMP[][2]))/B;

    vZ[0][0]=(quann(vRz[0][0]));
    vZ[1][0]=(quann(vRz[1][0]));
    vZ[2][0]=(quann(vRz[2][0]));

    return vZ;
}

//Derivadas da funcao de log-verossimilhanca
funDerloglik(const vP)
{
    decl cosa;
    decl k = rows(vP) - 1;
    decl avScore;
    decl eta = s_mX*vP[0:(k-1)];
    decl mu = exp(eta) ./ (1.0+exp(eta));

```

```

    decl phi = vP[k];
    decl ynew = log( s_vy_boot ./ (1.0-s_vy_boot) );
    decl munew = polygamma(mu*phi, 0) - polygamma((1.0-mu)*phi, 0);
    decl T = diag( exp(eta) ./ (1.0+exp(eta)) .^2 );
    decl c1=phi*s_mX'*T*(ynew-munew);
    decl c2=double(sumc( polygamma(phi, 0) - mu .*
polygamma(mu*phi, 0) - (1.0-mu) .* polygamma( (1.0-mu)*phi, 0) +
mu .* log(s_vy_boot) + (1.0-mu) .* log(1.0-s_vy_boot) ));

    return avScore=(c1|c2);

}

//Loop bootstrap (reamostragem)
funcReBoot(const B, const vp1, const v_thetaIni, const cn )
{
    decl bi,gerY,dfunc3,vp3,ir3;
    decl cfailureboot=0;
    decl mBoot = zeros(3,B);
    decl mSKed = zeros(3,B);
    for(bi=0; bi < B; ++bi)
    {
        decl v_eta_boot = s_mX*vp1[0:1]; //regressores
        decl v_mu_boot = exp(v_eta_boot) ./ (1.0+exp(v_eta_boot)); //vetor de medias
        decl d_phi_boot = vp1[2]; //parametro de precisao

        // Geracao do vetor Y de reamostragem
        fgerador_vY(&gerY,v_mu_boot, d_phi_boot, cn);
        s_vy_boot = gerY;

        mSKed[bi] = funDerloglik(vp1);

        vp3 = v_thetaIni; //valores iniciais dos parametros Bootstrap

        // estimacao por maxima verossimilhanca para o metodo de Bootstrap
        ir3 = MaxBFGS(floglikBoot, &vp3, &dfunc3, 0, FALSE);

        // control da convergencia na estimacao Bootstrap
        if( (ir3 == MAX_CONV) || (ir3 == MAX_WEAK_CONV) )
        {
            mBoot[bi] = vp3;
        }
        else
        {
            ++cfailureboot;
            --bi;
        }
    }

    return mBoot|mSKed;
}

//funcao para calcular as coberturas laterais
funTamLateral(const limif, const limsup, const parametro, const rep)
{
    decl vNum=zeros(3,2);

    vNum[0][0] = 100*sumr( parametro[0][0] .< limif[0][ ] )/rep;
    vNum[1][0] = 100*sumr( parametro[1][0] .< limif[1][ ] )/rep;
    vNum[2][0] = 100*sumr( parametro[2][0] .< limif[2][ ] )/rep;
    vNum[0][1] = 100*sumr( parametro[0][0] .> limsup[0][ ] )/rep;
    vNum[1][1] = 100*sumr( parametro[1][0] .> limsup[1][ ] )/rep;
    vNum[2][1] = 100*sumr( parametro[2][0] .> limsup[2][ ] )/rep;

    return vNum;
}

```

B.3. Programa de ligação

```

/*****
* PROGRAM: verossim.h
* AUTHOR: Raydonal Ospina Martinez
* DATE: June 13, 2003.
* VERSION: 0.1.5
* LAST MODIFIED: Novembro 05, 2003. Hora: 11:30
*****/

//Funcoes criadas para a simulacao

//funcao de log-verossimilhanca
floglik(const vP, const adFunc, const avScore, const amHess);

//calculo da matriz de informacao de Fisher
finformacaoFisher(const vp1, const adFunc);

//funcao geradora do vetor Y
fgerador_vY(const func, const v_mu, const d_phi, const cn);

//funcao  escore
fscore(const avScore, const vP);

//funcao  para o calculo do vies
fdeltaXiu(const adFunc, const v_mu, const d_phi, const v_eta, const m_Kbetabeta,
           const v_Kbetaphi, const d_Kphiphi, const m_inforK);

//funcao de log-verossimilhana do modelo de regressao beta para o
// calculo Bootstrap
floglikBoot(const vP, const adFunc, const avScore, const amHess);

//construcao de intervalo de confianca assintotico Normal
fICCN(const diagFisher, const qantil, const vp);

//calculo da cobertura do intervalo de confianca
fCobertura(const limInf, const limSup);

//construcao do intervalo de confianca por Bootstrap percentil
fICBOOTPer(const vreamostra, const nivel );

//funcao escore modificada de David Firth para o calculo do vies Bootstrap
fscoreDFBT(const avScore, const vP);

//ciclo bootstrap
funcReBoot(const B, const vp1, const v_thetaIni, const cn );

//ciclo bootstrap Cox & Snell
funcReBootCox(const B, const vp1, const v_thetaIni, const cn );

//ciclo bootstrap D. Firth
funcReBootFirth(const B, const vp1, const v_thetaIni, const cn );

//intervalo de confianca de tipo BCa
fICBootAcelerado(const vreamostra, const d_quantil, const vZ, const vA, const
B);

//funcao geradora da amostra Jackknife
funJK(const cn, const v_thetaIni );

//verossimilhanca jackknife
floglikJ(const vP, const adFunc, const avScore, const amHess);

//calculo da constante de aceleracion
faICBCA(const mEJ, const cn);

//calculo da constante Zo para O IC BCa
fZOBCA(const vE, const mEB, const B);

//derivadas da funcao  de Log-verossimilhanca
funDerloglik(const vP);

//funcao para calcular as coberturas laterais
funTamLateral(const limif, const limsup, const parametro, const rep);
```

B.4. Programa de geração do histograma das 5000 estimativas de máxima verossimilhança com as diferentes versões de intervalos de confiança, para os níveis de confiança de 90%, 95% e 99%.

```
#Grafico do histograma das 5000 estimativas de
#maxima verossimilhança com suas estimativas intervalares

##livrarias principais
library(MASS)
library(KernSmooth)

##letura dos dados
vetor1<-matrix(scan("c:/Alunos/raydonal/graficos/arquivo.txt"),ncol=6,byrow=T)

##Tamanho 40 para phi
postscript("c:/Alunos/raydonal/saida40.eps",horizontal=FALSE,height=7.5,width=5.5, pointsize=10)

#criacao do histograma
x <- vetor1[,1]
h <- dph(x)/2
xli<-min(x)
xls<-max(x)
bins <- seq(xli-0.001, xls+0.001+h, by=h)
hist(x, breaks=bins, ylim=c(-180,70), xlim=c(xli-0.01,xls+0.15),
type="n",xlab="",yaxt='n',ylab="",main="")
abline(h=0)
abline(v=1.5, lty=2)

##Intervalos de confiança 90%
segments( 1.4091, -20, 1.5934, -20, lty=1)
segments( 1.4080, -27, 1.5923, -27, lty=2)
segments( 1.4104, -34, 1.5947, -34, lty=3)
segments( 1.4097, -41, 1.5942, -41, lty=4)
segments( 1.4037, -48, 1.5942, -48, lty=5)
segments( 1.4154, -55, 1.5907, -55, lty=6)
segments( 1.4081, -62, 1.5939, -62, lty=7)

##Intervalos de confiança 95%
segments( 1.3915, -77, 1.6111, -77, lty=1)
segments( 1.3903, -84, 1.6099, -84, lty=2)
segments( 1.3927, -91, 1.6123, -91, lty=3)
segments( 1.3919, -98, 1.6116, -98, lty=4)
segments( 1.3849, -105, 1.6165, -105, lty=5)
segments( 1.3985, -112, 1.6073, -112, lty=6)
segments( 1.3900, -119, 1.6120, -119, lty=7)

##Intervalos de confiança 99%
segments( 1.3570, -134, 1.6456, -134, lty=1)
segments( 1.3558, -141, 1.6444, -141, lty=2)
segments( 1.3582, -148, 1.6468, -148, lty=3)
segments( 1.3551, -155, 1.6441, -155, lty=4)
segments( 1.3461, -162, 1.6504, -162, lty=5)
segments( 1.3633, -169, 1.6381, -169, lty=6)
segments( 1.3499, -176, 1.6479, -176, lty=7)

##legenda limite inferior dos intervalos de confiança
text( 1.4091 - 0.02, -20, " 6.44 ", cex = .5, srt=90)
text( 1.4080 - 0.02, -27, " 6.10 ", cex = .5, srt=90)
text( 1.4104 - 0.02, -34, " 6.66 ", cex = .5, srt=90)
text( 1.4097 - 0.02, -41, " 6.56 ", cex = .5, srt=90)
text( 1.4037 - 0.02, -48, " 5.68 ", cex = .5, srt=90)
text( 1.4154 - 0.02, -55, " 8.06 ", cex = .5, srt=90)
text( 1.4081 - 0.02, -62, " 6.26 ", cex = .5, srt=90)

text( 1.3915 - 0.02, -77, " 3.56 ", cex = .5, srt=90)
text( 1.3903 - 0.02, -84, " 3.26 ", cex = .5, srt=90)
text( 1.3927 - 0.02, -91, " 3.74 ", cex = .5, srt=90)
text( 1.3919 - 0.02, -98, " 3.62 ", cex = .5, srt=90)
text( 1.3849 - 0.02, -105, " 2.76 ", cex = .5, srt=90)
text( 1.3985 - 0.02, -112, " 4.54 ", cex = .5, srt=90)
text( 1.3900 - 0.02, -119, " 3.36 ", cex = .5, srt=90)

text( 1.3570 - 0.02, -134, " 0.9 ", cex = .5, srt=90)
text( 1.3558 - 0.02, -141, " 0.9 ", cex = .5, srt=90)
text( 1.3582 - 0.02, -148, " 0.96 ", cex = .5, srt=90)
text( 1.3551 - 0.02, -155, " 0.96 ", cex = .5, srt=90)
text( 1.3461 - 0.02, -162, " 0.64 ", cex = .5, srt=90)
text( 1.3633 - 0.02, -169, " 1.12 ", cex = .5, srt=90)
```

```

text( 1.3499 - 0.02, -176, " 0.84 ", cex = .5, srt=90)

#legenda limite superior dos intervalos de confianca

text( 1.5934 + 0.02, -20, " 5.88 ", cex = .5, srt=90)
text( 1.5923 + 0.02, -27, " 6.10 ", cex = .5, srt=90)
text( 1.5947 + 0.02, -34, " 5.58 ", cex = .5, srt=90)
text( 1.5942 + 0.02, -41, " 5.76 ", cex = .5, srt=90)
text( 1.5942 + 0.02, -48, " 5.20 ", cex = .5, srt=90)
text( 1.5907 + 0.02, -55, " 6.54 ", cex = .5, srt=90)
text( 1.5939 + 0.02, -62, " 5.88 ", cex = .5, srt=90)

text( 1.6111 + 0.02, -77, " 3.38 ", cex = .5, srt=90)
text( 1.6099 + 0.02, -84, " 3.50 ", cex = .5, srt=90)
text( 1.6123 + 0.02, -91, " 3.24 ", cex = .5, srt=90)
text( 1.6116 + 0.02, -98, " 3.42 ", cex = .5, srt=90)
text( 1.6165 + 0.02, -105, " 2.74 ", cex = .5, srt=90)
text( 1.6073 + 0.02, -112, " 3.92 ", cex = .5, srt=90)
text( 1.6120 + 0.02, -119, " 3.28 ", cex = .5, srt=90)

text( 1.6456 + 0.02, -134, " 0.74 ", cex = .5, srt=90)
text( 1.6444 + 0.02, -141, " 0.78 ", cex = .5, srt=90)
text( 1.6468 + 0.02, -148, " 0.74 ", cex = .5, srt=90)
text( 1.6441 + 0.02, -155, " 1.00 ", cex = .5, srt=90)
text( 1.6504 + 0.02, -162, " 0.74 ", cex = .5, srt=90)
text( 1.6381 + 0.02, -169, " 1.26 ", cex = .5, srt=90)
text( 1.6479 + 0.02, -176, " 0.92 ", cex = .5, srt=90)

#legenda dos niveis de confianca
text( 1.8, -41, " 90% ", cex = .8, srt=90)
text( 1.8, -98, " 95% ", cex = .8, srt=90)
text( 1.8, -155, " 99% ", cex = .8, srt=90)

#legenda do parametro verdadeiro
text(1.48, 74, expression(beta[0]),cex = .9,srt=90)

#segmentos do diferentes tipos de intervalos de confianca
segments( 1.7500, 10+10, 1.7500, 22+10, lty=1)
segments( 1.7650, 10+10, 1.7650, 22+10, lty=2)
segments( 1.7800, 10+10, 1.7800, 22+10, lty=3)
segments( 1.7950, 10+10, 1.7950, 22+10, lty=4)
segments( 1.8100, 10+10, 1.8100, 22+10, lty=5)
segments( 1.8250, 10+10, 1.8250, 22+10, lty=6)
segments( 1.8400, 10+10, 1.8400, 22+10, lty=7)

#legenda do diferentes tipos de intervalos de confianca
text( 1.71, 22+10, " Intervalos", cex = .8, srt=90)
text( 1.73, 22+10, " de Confiana ", cex = .8, srt=90)

text( 1.7500, 30+5, " ICA ", cex = .7, srt=90, adj=0)
text( 1.7650, 30+5, " IACC ", cex = .7, srt=90, adj=0)
text( 1.7800, 30+5, " IACp ", cex = .7, srt=90, adj=0)
text( 1.7950, 30+5, " ICP ", cex = .7, srt=90, adj=0)
text( 1.8100, 30+5, " ICCc ", cex = .7, srt=90, adj=0)
text( 1.8250, 30+5, " ICCP ", cex = .7, srt=90, adj=0)
text( 1.8400, 30+5, " BCa ", cex = .7, srt=90, adj=0)

mtext("n = 40",side=2,line=1)

box() #fecha o grafico
dev.off()#fim

```

Referências

- [1] ABRAMOWITZ, M. L. & STEGUN, I. A. (1965). *Handbook of Mathematical Functions*. New York: Dover.
- [2] AMEMIYA, T. (1985). *Advanced Econometrics*. Great Britain: Basil Blackwell.
- [3] AUBIN, E. C. Q. & CORDEIRO, G. M. (1997). Bias in linear regression models with unknown covariance matrix. *Communications in Statistics, Simulation and Computation*, **26**, 813–828.
- [4] ATKINSON, A. C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.
- [5] ANDERSON, J. A. & RICHARDSON, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, **22**, 621–627.
- [6] BARNDORFF-NIELSEN, O. E. & COX, D. R. (1989). *Asymptotics Techniques for Use in Statistics*. London: Chapman and Hall.
- [7] BARNDORFF-NIELSEN, O. E. & COX, D. R. (1994). *Inference and Asymptotics*. London: Chapman and Hall.
- [8] BARTLETT, M. S. (1953). Aproximate confidence intervals. *Biometrika*, **40**, 12–19.
- [9] BERAN, R. (1986). Discussion of “Jackknife, bootstrap and other resampling methods in regression analysis” by C.F.J. Wu. *Annals of Statistics*, **14**, 1295–1298.
- [10] BERKOWITZ, J. & KILIAN, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, **19**, 1–48.
- [11] BICKEL, P. J. & DOKSUM, K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. 2^a ed. New Jersey: Prentice Hall.
- [12] BICKEL, P. J. & FREEDMAN, D. A. (1980). On Edgeworth expansions for the bootstrap. Thechnical Report . Departament os Statistics. University of California. Berkeley.
- [13] BILLINGSLEY, P. (1995). *Probability and Measure*, 2^a ed. New York: John Wiley and Sons.
- [14] BOTTER, D. A. & CORDEIRO, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Communications in Statistics, Simulation and Computation*, **59**, 1–14.
- [15] BOWMAN, K. & SHENTON, L. R. (1965). Biases and covariances of maximum likelihood estimators. *Report K-1633, Union Carbide Corporation, Oak Ridge*.
- [16] BOX, M. (1971). Bias in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society B*, **33**, 171–201.
- [17] BURY, K. (1999). *Statistical Distributions in Engineering*. New York: Cambridge University Press.
- [18] COOK, R., TSAI, C. & WEI, B. (1986). Bias in nonlinear regression. *Biometrika*, **73**, 615–623.
- [19] COPAS, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society B*, **50**, 225–265.
- [20] CORDEIRO, G. M. & MCCULLAGH, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B*, **53**, 629–643.

- [21] CORDEIRO, G. M. & CRIBARI-NETO, F. (1993). On Bartlett corrections, bias reduction and a new class of transformations. *Brazilian Journal of Probability and Statistics*, **7**, 179–200.
- [22] CORDEIRO, G. M. (1993). Barlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, **22**, 169–188.
- [23] CORDEIRO, G. M. & KLEIN, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, **19**, 169–176.
- [24] CORDEIRO, G. M., ROCHA, E. C., ROCHA, J. G. C. & CRIBARI-NETO, F. (1997). Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, **58**, 21–35.
- [25] 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.
- [26] CORDEIRO, G. M., & CRIBARI-NETO, F. (1998). On bias reduction in exponential and non-exponential family regression models. *Communications in Statistics, Simulation and Computation*, **27**, 485–500.
- [27] CORDEIRO, G. M., VASCONCELLOS, K. L. P. & SANTOS, M. L. (1998). On the second order bias of parameter estimates in nonlinear regressions models with Student t erros. *Journal of Statistical Computation and Simulation*, **60**, 363–378.
- [28] 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.
- [29] COX, D. & SNELL, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, **30**, 248–275.
- [30] COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [31] CRIBARI-NETO, F., BOTTER, D. A., CORDEIRO, G. M. & FERRARI, S. L. P. (1998). Bias reduction in one-parameter exponential familiy models. *Communications in Statistics, Simulation and Computation*, **27**, 761–782.
- [32] 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.
- [33] CRIBARI-NETO, F. & ZARKOS, S. G. (2003). Econometric and statistical computing using Ox. *Computational Economics*, **21**, 277–295.
- [34] DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. New York: Cambridge University Press.
- [35] DICICCIO, T. J. & TIBSHIRANI, R. (1987). Bootstrap confidence intervals and bootstrap aproximations. *Journal of the American Statistical Association*, **82**, 163–170.
- [36] DISHON, M. & WEISS, G. H. (1980). Small sample comparison of estimation methods for the beta distribution. *Journal of Statistical Computation and Simulation*, **11**, 1–11.
- [37] DOORNIK, J. A. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4th ed. London: Timberlake Consultants & Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik>.
- [38] EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

- [39] EFRON, B. & TIBSHIRANI, R. J. (1986). Bootstrap methods for standard errors: confidence intervals and other measures of statistical accuracy. *Statistical Science*, **1**, 54–77.
- [40] EFRON, B. & TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- [41] 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.
- [42] FERRARI, S. L. P. & CRIBARI-NETO, F. (1998). On bootstrap and analytical bias corrections. *Economics Letters*, **58**, 7–15.
- [43] FERRARI, S. L. P. & CRIBARI-NETO, F. (2003). Beta regression for modeling rates and proportions. *RT-MAE-2003-04*.
- [44] FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.
- [45] FISHER, R. A. (1925). Theory of statistical estimation. *Proceedings of Cambridge Philosophical Society*, **22**, 700–725.
- [46] FRYER, J. G. & ROBERTSON, C. A. (1972). A comparison of some methods for estimating mixed normal distributions. *Biometrika*, **59**, 639–648.
- [47] GRAHAM, V. A. & HOLLANDS, K. G. T. (1990). Method to generate synthetic hourly solar radiation globally. *Solar Energy*, **44**, 333–341.
- [48] GREENE, W. H. (2000). *Econometric Analysis*, 3^a ed. New York: Prentice Hall.
- [49] HALDANE, J. B. S. (1953). The estimation of two parameters from a sample. *Sankhyā*, **12**, 313–320.
- [50] HALDANE, J. B. S. & SMITH, S. M. (1956). The sampling distribution of a maximum likelihood estimate. *Biometrika*, **43**, 96–103.
- [51] HOROWITZ, J. L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, **61**, 395–411.
- [52] JANARDAN, K. G. & PADMANABHAN, G. (1986). Double bounded beta distribution for hydrologic variables. *Proc. 17th Annual Pittsburgh Conference* (parte 3), **17**, 1107–1111.
- [53] JEONG, J. & MADDALA, G. S. (1993). A perspective on application of bootstrap methods in econometrics. Em Maddala, G.S., Rao, C.R. & Vinod, H.D., eds., *Handbook of Statistics: Econometrics*, **11**, 573–610. Amsterdam: North-Holland.
- [54] JOHNSON, N., KOTZ, S. & BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*. 2^a ed. New York: John Wiley and Sons.
- [55] KNUTH, D. (1986). *The T_EXbook*. New York: Addison–Wesley.
- [56] LAWLEY, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.
- [57] LEHMANN, E. L. & CASELLA, E. (1998). *Theory of Point Estimation*, 2^a ed. New York: Springer-Verlag.
- [58] MAFFET, A. L. & WACKERMAN, C. C. (1991). The modified beta density function as a model for synthetic aperture radar clutter statistics. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 277–283.

- [59] MACKINNON, J. G. & SMITH, JR. A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, **85**, 205–230.
- [60] MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.
- [61] MCCULLAGH, P. & NELDER, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- [62] MCLACHLAN, G. J. (1980). A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics*, **21**, 71–78.
- [63] McNALLY, R. J. (1990). Maximum likelihood estimation of the parameters of the prior distributions of three variables that strongly influence reproductive performance in cows. *Biometrics*, **446**, 501–514.
- [64] MILYUTIN, E. R. & YAROMENKO, YU. I. (1991). Statistical characteristics of atmospheric transparency index over tilted routes. *Meteorologiya i Gidrologiya*, **12**, 72–76.
- [65] PATIL, G. P. & SHOROCK, R. (1965). On certain properties of the exponential-type distribution. *Journal of the Royal Statistical Society B*, **27**, 94–99.
- [66] PAULA, G. A. (1992). Bias correction for exponential family nonlinear models. *Journal of Statistical Computation and Simulation*, **40**, 43–54.
- [67] PAULA, G. A. & CORDEIRO, G. M. (1995). Bias correction and improved residuals for non-exponential family nonlinear models. *Communications in Statistics, Theory and Methods*, **24**, 1193–1210.
- [68] PRATER, N. H. (1956). Estimate gasoline yields from crude. *Petroleum Refiner*, **35**, 236–238.
- [69] PRESS, W. H., TEULOSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2^a ed. London: Prentice Hall.
- [70] RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2^a ed. New York: Wiley.
- [71] ROBERTSON, C. A. & FRYER, J. G. (1970). The bias and accuracy of moment estimators. *Biometrika*, **57**, 57–65.
- [72] SEVERINI, T. A. (2000). *Likelihood Methods in Statistics*. New York: Oxford University Press.
- [73] SHENTON, L. R. & WALLINGTON, P. A. (1962). The bias of moment estimators with application to the negative binomial distribution. *Biometrika*, **49**, 193–204.
- [74] SHENTON, L. R. & BOWMAN, K. (1963). Higher moments of a maximum likelihood estimate. *Journal of the Royal Statistical Society B*, **25**, 305–317.
- [75] SHENTON, L. R. & BOWMAN, K. (1977). *Maximum Likelihood Estimation in Small Samples*. Londres: Charles Griffin.
- [76] SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap *Annals of Statistics*, **9**, 1187–1195.
- [77] SOWDEN, R. R. (1972). On the first-order bias of parameter estimates in a quantal response model under alternative estimation procedures. *Biometrika*, **59**, 573–579.
- [78] VASCONCELLOS, K. L. P. & CORDEIRO, G. M. (1997a). Approximate bias for multivariate nonlinear heteroscedastic regressions. *Brazilian Journal of Probability and Statistics*, **11**, 141–159.

- [79] VASCONCELLOS, K. L. P. & CORDEIRO, G. M. (1997b). Estimadores corrigidos para modelos SUR não lineares. *Revista de Econometria*, **17**, 45–65.
- [80] VASCONCELLOS, K. L. P. & CORDEIRO, G. M. (1999). Bias corrected estimates in multivariate student t regression models. *Communications in Statistics, Simulation and Computation*, **29**, 797–822.
- [81] VASCONCELLOS, K. L. P. & CRIBARI-NETO, F. (2003). Improved maximum likelihood estimation in a new class of beta regression models. *Artigo submetido para publicação*.
- [82] VINOD, H. D. (1993). Bootstrap methods: applications in econometrics. Em Maddala, G.S., Rao, C.R. & Vinod, H.D., eds., *Handbook of Statistics: Econometrics*, **11**, 629–661. Amsterdam: North-Holland.
- [83] WILEY, J. A., HERSCHOKORU, S. J. & PADIAU, N. S. (1989). Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse. *Statistics in Medicine*, **8**, 93–102.
- [84] WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, **14**, 1261–1295.
- [85] YOUNG, D. & BAKIR, S. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183–191.