**Redução de Vício de Estimadores
de Máxima Verossimilhança
na Família Exponencial
Uniparamétrica**

Aldy Fernandes da Silva

Dissertação apresentada ao Instituto de
Matemática e Estatística da Universidade de São Paulo
para Obtenção do Grau de Mestre em Estatística

Área de Concentração: Estatística
Orientadora: **Prof^a Silvia Lopes de Paula Ferrari**

Durante este trabalho o autor recebeu apoio financeiro do CNPq.

- São Paulo, dezembro de 1996 -

Redução de Vício de Estimadores de Máxima Verossimilhança na Família Exponencial Uniparamétrica

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Aldy Fernandes da Silva
e aprovada pela comissão julgadora.

São Paulo, 09 de dezembro de 1996.

Banca Examinadora:

- Prof^a Dr^a Silvia Lopes de Paula Ferrari (orientadora) - DE-IME-USP
- Prof. Dr. Heleno Bolfarine - DE-IME-USP
- Prof. Dr. Enrico Antônio Colosimo - EST-UFMG

Resumo

O objetivo principal desta tese é a comparação de alguns métodos de redução de vício de estimadores de máxima verossimilhança em modelos da família exponencial uniparamétrica. Duas abordagens são apresentadas. Na abordagem analítica, são apresentados os métodos de correção de vícios desenvolvidos por Ferrari e outros (1996) e Firth (1993). Neste caso, expressões para vícios e estimadores corrigidos são apresentadas até ordens n^{-1} e n^{-2} . Na abordagem de reamostragem, são discutidos os métodos de redução de vício através das metodologias Jackknife e Bootstrap.

Ilustramos o uso destas técnicas, apresentando alguns resultados de simulação, para algumas distribuições clássicas, mostrando a importância prática dos resultados desenvolvidos. Finalmente, apresentamos algumas conclusões e sugerimos alguns tópicos para pesquisa.

Abstract

In this dissertation we compare some methods of bias reduction of maximum likelihood estimators in one-parameter exponential family. Two boardings are presented. In the analitic boarding, we present some methods of bias correction developed by Ferrari et al (1996) and Firth (1993). In this case, bias and corrected estimators expressions are presented to order n^{-1} and n^{-2} . In the resampling boarding, we discuss methods of bias reduction through the Jackknife e Bootstrap boarding methodologies.

We illustrate the practical importance of these techniques presenting some results of simulations, involving some classic distributions. Finally, we give some conclusions and we suggest some topics for future research.

À minha Família
Em especial,
aos meus pais
Nívia e Manoel
Aos meus amigos

Agradecimentos

Esta dissertação é o resultado de um ideal profissional. Um ideal que busco a cada dia, com muita vontade e dedicação. E é por este ideal, que estou me aproximando, que venho agradecer a algumas pessoas, muito importantes, que me ajudaram a continuar o meu caminho.

- Primeiro, gostaria de expressar agradecimento à minha orientadora, **Profª Silvia Lopes de Paula Ferrari**, que com enorme competência me guiou nesta minha caminhada, dando-me força através de suas palavras sempre entusiastas de carinho e apoio incondicional,

- ao **Prof. Francisco Cribari-Neto** pela sua enorme colaboração e paciência, que me ajudaram na elaboração dos programas computacionais de simulação,

- à **Profª Denise Aparecida Botter**, obrigado pelos momentos de escuta, de discussão e pela oportunidade de compartilhar de sua alegria,

- aos meus amigos **Roberto Quinino**, **Deusélio Fioresi** e **Erika Fukunaga**, pessoas que sempre estiveram presentes nos momentos críticos da minha vida, ajudando-me a superar imensos obstáculos,

- à **Sueli Mingoti**, ex-professora e amiga, que sempre incentivou e acreditou em meu trabalho.

Sumário

1	Introdução	2
2	Métodos Analíticos de Redução de Vício na Família Exponencial Uniparamétrica	11
2.1	Introdução	11
2.2	A Família Exponencial Uniparamétrica	15
2.3	Casos Especiais	18
2.4	Verificação das Fórmulas dos Vícios em Alguns Casos Especiais	30
2.5	Expansões Assintóticas	31
2.6	Análise Gráfica	35
3	Estudos de Simulação	44
3.1	Introdução	44
3.2	A Distribuição Binomial Negativa	45
3.3	A Distribuição Gama	50
3.4	A Distribuição de Rayleigh	54
3.5	A Distribuição de Pareto	56
3.6	A Distribuição Lognormal	58
	Conclusão	61
	Apêndice 1	63
	Apêndice 2	65
	Referências Bibliográficas	69

Capítulo 1

Introdução

Estimadores de máxima verossimilhança (E.M.V.) são bastante utilizados na prática pois possuem boas propriedades assintóticas como consistência, eficiência e suficiência. Entretanto, é bem conhecido que os E.M.V.'s podem ser viciados quando o tamanho da amostra n ou a informação total de Fisher é pequeno. Em alguns casos, os vícios são muito relevantes e podem levar a conclusões práticas muito deficientes (Picard e Prum, 1992). Surge, então, a importância de um processo que elimine parcialmente, até uma certa ordem de magnitude, ou totalmente o vício deste estimador.

Nos últimos anos, grande número de métodos tem sido desenvolvido para melhorar o E.M.V. Dos vários métodos propostos destacam-se duas abordagens, a analítica e a de reamostragem. Neste capítulo, apresentamos uma série de trabalhos que, ao longo dos últimos anos, contribuíram para o desenvolvimento dos vários métodos de redução de vício em E.M.V.'s. Além disso, serão apresentados os métodos de reamostragem, Jackknife e Bootstrap, para redução de vícios.

Originalmente, Bartlett (1953) obteve uma expansão para o vício até ordem n^{-1} do E.M.V. em modelos uniparamétricos para o caso de observações independentes e identicamente distribuídas (*i.i.d.*). Mais tarde, Shenton e Bowman (1963) desenvolveram, ainda no caso uniparamétrico, os quatro primeiros momentos amostrais do estimador de máxima verossimilhança, respectivamente, até ordens n^{-2} , n^{-3} , n^{-3} e n^{-4} . Shenton e Wallington (1962) encontraram expressões gerais para o vício até ordem n^{-1} de estimadores de momentos e de máxima verossimilhança para o caso de estimação simultânea de vários parâmetros. Cox e Snell (1968) também apresentaram fórmulas gerais para o vício até ordem n^{-1} dos E.M.V.'s

em modelos multiparamétricos.

Em um modelo regular (Cox e Hinkley, 1974) com um parâmetro unidimensional θ , a expansão para o vício do E.M.V. $\hat{\theta}$ pode ser escrita como

$$B(\theta) = \frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2} + \dots, \quad (1.1)$$

onde n é usualmente o número de observações. A expansão em (1.1) também é válida no caso multidimensional e, nesta situação, uma expansão análoga a (1.1) é obtida para cada componente do vetor paramétrico. Vários métodos analíticos e de reamostragem para redução de vício têm sido extensivamente estudados nos últimos anos. Ambos os métodos têm êxito na remoção do termo $B_1(\theta)/n$ da expansão de $B(\theta)$. Os métodos de reamostragem têm a vantagem de não requerer o cálculo de $B_1(\theta)$. Por outro lado, necessitam de um intensivo uso computacional. Os métodos analíticos são precisos, porém, em alguns casos são de difícil solução. Um método analítico padrão envolve a obtenção de $B_1(\theta)$ e a substituição do parâmetro desconhecido θ por $\hat{\theta}$. O estimador corrigido por vício é então calculado como

$$\hat{\theta}_c = \hat{\theta} - \frac{B_1(\hat{\theta})}{n}. \quad (1.2)$$

Neste mesmo contexto analítico, Withers (1987) apresenta um método de redução de vício baseado em expansão de Taylor (Método Delta) para estimar uma função $t(\theta)$. Ele mostra como construir, para uma vasta classe de estimadores θ e para cada k , um estimador de $t(\theta)$ com vício aproximadamente n^{-k} . Doss e Sethuraman (1989) mostraram que quando não existe estimador não viciado de um parâmetro θ , reduzindo o seu vício para zero através de uma seqüência de estimadores $\hat{\theta}, \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots$, onde seus vícios convergem para zero, necessariamente, a variância dos estimadores é forçada a tender para infinito. Segundo eles, isto dá um suporte qualitativo para se dizer que a redução de vício é uma propriedade não desejável, mas torna-se desejável somente se estiver acompanhada por uma redução no erro quadrático médio. Cribari-Neto, Botter, Cordeiro e Ferrari (1995) obtiveram expressões em forma fechada para vícios de ordens n^{-1} e n^{-2} dos estimadores de máxima verossimilhança para distribuições na família exponencial uniparamétrica e realizaram uma análise gráfica mostrando como cada vício e erro quadrático médio dos E.M.V.'s e suas versões corrigidas variam de acordo com o parâmetro que indexa algumas distribuições.

A redução de vícios de E.M.V.'s tem sido bastante utilizada em várias áreas da Estatística como Modelos Normais de Regressão, Modelos Lineares Generalizados, Análise Discriminante, etc. E.M.V.'s de parâmetros em modelos de regressão não-linear normal

também são geralmente viciados. O mesmo é verdade para outras quantidades que são úteis em análises de modelos de regressão não-linear, tais como os resíduos ordinários que não têm esperança zero. Box (1971) apresentou um exemplo simples da dificuldade do cálculo do vício dos parâmetros deste modelo e obteve uma aproximação para este vício em uma versão multivariada. Ele ainda considerou que a estimação de máxima verossimilhança pode ser vista como um caso especial de estimação Bayesiana onde uma distribuição a priori uniforme é assumida, e obteve uma expressão para o vício quando a estimação é Bayesiana, com uma distribuição a priori não-uniforme. Finalmente, seu enfoque foi aplicado para determinar os vícios dos E.M.V.'s de parâmetros da distribuição gama. Cook e outros (1986) descrevem um método relativamente simples para investigar os vícios de resíduos e E.M.V.'s de parâmetros de modelos de regressão não-linear normal. Cordeiro e McCullagh (1991) derivaram uma fórmula geral para os vícios de ordem n^{-1} dos E.M.V.'s de modelos lineares generalizados. Esta fórmula pode ser implementada no programa GLIM facilitando o cálculo das estimativas de máxima verossimilhança corrigidas por vício até ordem n^{-1} . Cordeiro (1993) também apresentou, em notação matricial, fórmulas gerais para o vício de ordem n^{-1} dos E.M.V.'s de dois modelos de regressão heterocedásticos. Paula (1991) discute o cálculo do vício até ordem n^{-1} em uma classe geral de modelos de regressão não-linear e o expressa em uma fórmula matricial, generalizando trabalhos anteriores encontrados na literatura estatística (Box, 1971; Cordeiro e McCullagh, 1991). Young e Bakir (1987) apresentaram aproximações do vício de E.M.V.'s dos parâmetros de um modelo de regressão no qual a variável resposta tem distribuição log-gama generalizada. Um estimador corrigido para o parâmetro de escala foi proposto e este tem menor erro quadrático médio do que o estimador sem correção, desde que exista no mínimo uma variável regressora no modelo. Os resultados encontrados por eles sugerem que a correção do vício para o estimador do parâmetro de escala é importante em pequenas amostras para todas as densidades da família log-gama generalizada.

Anderson e Richardson (1979), utilizando a fórmula geral para o vício até ordem n^{-1} dada por Cox e Snell (1968), encontraram os vícios de E.M.V.'s em problemas de discriminação logística. Além disso, mostraram através de um estudo de simulação que esta correção é efetiva. Eles ainda notaram que os estimadores corrigidos produzem uma redução no vício sem um correspondente aumento na variabilidade. McLachlan (1980) também apresentou correções de vícios para E.M.V.'s dos parâmetros de um problema de discriminação logística sobre esquemas de amostragem misto e separado.

Num contexto analítico um pouco diferente do que foi apresentado acima, Schäbe (1990)

propôs reduzir o vício de E.M.V.'s usando estimadores "kernel". Schäbe modificou a função de verossimilhança, substituindo a distribuição empírica por um estimador "kernel" da distribuição. A redução do vício foi então realizada introduzindo um vício adicional, o vício do estimador "kernel". Resultados sobre estimadores "kernel" são encontrados em Bean e Tsokos (1980). O estimador que foi obtido maximizando a verossimilhança modificada tem um vício de ordem n^{-2} sob certas condições, e tem a mesma distribuição assintótica do E.M.V. usual. Como aplicação de seus resultados, Schäbe propôs um E.M.V. modificado do parâmetro da distribuição de Weibull no caso de amostras censuradas e não-censuradas. Firth (1993) mostrou que o termo de ordem n^{-1} do vício assintótico do E.M.V. é removido por uma modificação conveniente na função escore. Sua idéia é que o vício do estimador $\hat{\theta}$ pode ser reduzido introduzindo um pequeno vício na função escore. Ele chamou esse tipo de correção de preventiva. Seus resultados foram aplicados para o caso da família exponencial com parametrização canônica.

Uma outra abordagem muito utilizada nos últimos anos é a reamostragem. Alguns autores criticam o uso desta abordagem, uma vez que seu desenvolvimento demanda uma intensiva utilização de recursos computacionais além de não ser um método exato. Por outro lado, seus defensores argumentam que com o desenvolvimento computacional esta desvantagem é mínima e ainda evita-se a necessidade de se obter analiticamente uma expansão para o vício do estimador. Nesta abordagem, dois métodos são frequentemente utilizados: o Jackknife (Quenouille, 1949, 1956), também chamado método de Quenouille e o método Bootstrap (Efron, 1979). Quenouille (1949) apresentou um método para correção de coeficientes de correlação parcial em séries temporais e posteriormente estudou suas propriedades (Quenouille, 1956). Ele sugeriu o uso deste método adequado para correção de vício de qualquer estimador. Posteriormente, o uso geral desta técnica foi proposto para reduzir o vício de estimadores e então denominada de Jackknife. O procedimento Jackknife pode ser descrito da seguinte forma: Seja θ um parâmetro desconhecido e $X_1, X_2, \dots, X_n$ uma amostra de n observações *i.i.d.* com função distribuição conjunta F_θ , que depende de θ . Suponha que um método de estimação de θ razoavelmente bom (mas viciado) seja usado. Denote por $\hat{\theta}_{(i)}$, $i = 1, \dots, n$, o estimador de θ obtido eliminando a i -ésima observação, isto é, $\hat{\theta}_{(i)} = \hat{\theta}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$. Seja $\hat{\theta}$ o estimador de θ baseado em todas as n observações. Defina novos estimadores, chamados por Tukey de pseudo-valores, como

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)} = \hat{\theta} - (n-1)(\hat{\theta}_{(i)} - \hat{\theta}).$$

O estimador jackknife de θ corrigido por vício até ordem n^{-1} é então a média dos $\tilde{\theta}_i$,

$i = 1, \dots, n$, isto é,

$$\tilde{\theta}_J^1 = n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)} = \hat{\theta} - (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}), \quad (1.3)$$

onde $(n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$ é o estimador jackknife do vício e $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(i)}/n$.

O estimador jackknife $\tilde{\theta}_J^1$ tem a propriedade de eliminar o termo de ordem n^{-1} de qualquer vício da forma (1.1) (Miller, 1964, 1974).

Para eliminar o vício de ordem n^{-2} , Quenouille (1956) obteve um estimador jackknife dado por

$$\tilde{\theta}_J^2 = \frac{n^2\tilde{\theta}_J^1 - (n-1)^2 \sum_{j=1}^n \tilde{\theta}_{(j)}^1/n}{n^2 - (n-1)^2}, \quad (1.4)$$

onde $\tilde{\theta}_{(j)}^1$ é o estimador $\tilde{\theta}_J^1$ aplicado na amostra de tamanho $n-1$ com a j -ésima observação removida. O cálculo do estimador jackknife corrigido até ordem n^{-2} necessita então de um total de $n(n-1)$ amostras.

Em termos do estimador original $\hat{\theta}$, o estimador jackknife de ordem n^{-2} pode ser escrito como

$$\tilde{\theta}_J^2 = (n-1)^{-1} \left\{ n^3\hat{\theta} - (2n^2 - 2n + 1)(n-1)\hat{\theta}_{(\cdot)} + (n-1)^2(n-2)\frac{2}{n(n-1)} \sum_{i < j} \hat{\theta}_{(ij)} \right\}, \quad (1.5)$$

onde $\hat{\theta}_{(ij)}$ denota o estimador original aplicado à amostra de tamanho $n-2$ com a i -ésima e a j -ésima observações removidas.

Uma generalização da técnica Jackknife pode ser feita quando a amostra de tamanho n é dividida em g grupos de tamanho h , $n = gh$. Resultados sobre este procedimento podem ser encontrados com maiores detalhes em Quenouille (1956), Miller (1974), Efron (1982) e Efron e Tibshirani (1993, Capítulo 11). Miller (1974) apresentou uma importante revisão baseada no trabalho original de Quenouille e tentou incluir uma bibliografia de todos os trabalhos publicados com a metodologia Jackknife até aquela época.

Outro método de reamostragem muito usado para redução de vícios é o bootstrap. Esta técnica é, sem dúvida, de grande relevância e considera dois aspectos de estimação: a paramétrica e a não-paramétrica. O procedimento bootstrap pode ser descrito da seguinte forma: Considere o princípio “plug-in”, um método simples de estimação de parâmetros. O estimador “plug-in” de um parâmetro $\theta = t(F)$ é definido por $\hat{\theta} = t(\hat{F})$. Em outras palavras, estimamos a função $\theta = t(F)$ da distribuição de probabilidade F pela mesma função distribuição empírica $\hat{F}$, $\hat{\theta} = t(\hat{F})$. O procedimento bootstrap baseia-se simplesmente no princípio “plug-in”. Suponha que queremos estimar o vício do estimador de um parâmetro $\theta = t(F)$ desconhecido. A distribuição F é desconhecida, mas uma amostra aleatória $\mathbf{x} = (x_1, \dots, x_n)$

com função distribuição empírica $\hat{F}$ pode ser obtida. Considere então um estimador $\hat{\theta}$ de θ . O vício de $\hat{\theta}$ é definido como

$$\text{vicio}_F = E_F(\hat{\theta}) - \theta = E_F(\hat{\theta}) - t(F).$$

Retire uma amostra bootstrap $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ com distribuição de probabilidade empírica $\hat{F}$. Uma amostra bootstrap $x_1^*, \dots, x_n^*$ é definida como uma amostra aleatória de tamanho n retirada com reposição do conjunto de dados original $(x_1, \dots, x_n)$. O bootstrap assim definido é comumente chamado bootstrap não-paramétrico. Desta forma, o estimador bootstrap do vício de $\hat{\theta}$ é definido como

$$\text{vicio}_{\hat{F}} = E_{\hat{F}}(\hat{\theta}^*) - t(\hat{F}),$$

onde $E_{\hat{F}}(\hat{\theta}^*)$ é a esperança de $\hat{\theta}$ sobre a distribuição de probabilidade empírica da amostra bootstrap e $t(\hat{F})$ é o estimador “plug-in” de θ .

O estimador bootstrap ideal do vício deve ser aproximado por um procedimento de simulação Monte Carlo. Selecione B amostras bootstrap independentes $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$ da distribuição empírica $\hat{F}$. Calcule as réplicas bootstrap $\hat{\theta}_b^*$ e aproxime a esperança $E_{\hat{F}}(\hat{\theta}^*)$ pela média $\hat{\theta}_{(\cdot)}^* = \sum_{b=1}^B \hat{\theta}_b^* / B$. O estimador bootstrap do vício de $\hat{\theta}$, de ordem n^{-1} , baseado nas B réplicas é então dado por

$$\text{vicio}_B = \sum_{b=1}^B \frac{\hat{\theta}_b^*}{B} - \hat{\theta}.$$

Como nosso objetivo é corrigir o vício do estimador $\hat{\theta}$ de um parâmetro θ , então o estimador bootstrap corrigido por vício até ordem n^{-1} é

$$\hat{\theta}_B^1 = 2\hat{\theta} - \hat{\theta}_{(\cdot)}^*. \quad (1.6)$$

Se o interesse é reduzir o vício, até ordem n^{-2} , do estimador $\hat{\theta}$, selecione B amostras bootstrap de cada amostra $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, totalizando, portanto, B^2 amostras denotadas por $\mathbf{x}^{**1}, \dots, \mathbf{x}^{**B^2}$. Calcule as réplicas $\hat{\theta}_b^{**}$, e aproxime a esperança $E_{\hat{F}}(\hat{\theta}^{**})$ pela média $\hat{\theta}_{(\cdot)}^{**} = \sum_{b=1}^{B^2} \hat{\theta}_b^{**} / B^2$. Da mesma forma de (1.6), pode-se mostrar que o estimador bootstrap corrigido por vício até ordem n^{-2} é dado por

$$\hat{\theta}_B^2 = 3\hat{\theta} - 3\hat{\theta}_{(\cdot)}^* + \hat{\theta}_{(\cdot)}^{**}. \quad (1.7)$$

Hall (1992) mostra que, a cada iteração bootstrap a ordem de magnitude do vício é reduzida pelo fator n^{-1} .

Assim como o bootstrap não-paramétrico, podemos calcular o estimador bootstrap paramétrico corrigido por vício. O bootstrap paramétrico assume um modelo paramétrico F para os dados e é útil em problemas onde existe algum conhecimento sobre a forma da distribuição da população. Suponha que F dependa de um parâmetro θ . Aproximamos o estimador bootstrap ideal por uma amostragem bootstrap, mas de uma forma diferente da não-paramétrica. Em vez de amostragem com reposição dos dados, geramos B amostras de tamanho n de uma população com distribuição F porém, de parâmetro $\hat{\theta}$. Para cada amostra bootstrap gerada, podemos obter as réplicas bootstrap $\hat{\theta}_b^*$, $b = 1, \dots, B$. Geradas B amostras bootstrap independentes desta distribuição através de um procedimento de simulação Monte Carlo, podemos utilizar (1.6) para calcular a respectiva quantidade bootstrap paramétrica. Para reduzir o vício até ordem n^{-2} , é necessário gerar B amostras de tamanho n de populações com distribuição F , de parâmetros $\hat{\theta}_b^*$, respectivamente. O procedimento requer então a geração de B^2 amostras bootstrap e (1.7) pode ser usada para calcular o estimador bootstrap paramétrico corrigido por vício até ordem n^{-2} . Hall (1992, Capítulo 2) apresenta os métodos bootstrap paramétrico e não-paramétrico e dá uma fórmula geral para calcular o estimador corrigido por vício depois de j iterações bootstrap.

Os procedimentos de reamostragem são muito estudados e discutidos em vários trabalhos. Por serem métodos tipicamente computacionais, estes procedimentos tendem a se aprimorarem cada vez mais rápido com o crescente desenvolvimento da computação. Nos últimos anos, vários autores estudaram o comportamento destes métodos em situações práticas e os tomaram como base para o estudo de novos métodos de redução de vícios. Schucany, Gray e Owen (1971) apresentaram um procedimento geral para reduzir o vício de estimadores pontuais, que inclui o Jackknife como um caso especial. Considere duas funções t_1 e t_2 definidas sobre n observações $x_1, \dots, x_n$, *i.i.d.*, como dois estimadores diferentes de um parâmetro δ . Suponha que t_1 e t_2 são estimadores viciados tais que

$$E[t_k(X_1, \dots, X_n)] - \delta = b_k(n, \delta), k = 1, 2,$$

onde $b_k(n, \delta) \neq 0$ para algum δ no espaço paramétrico. Definindo o estimador

$$\hat{\delta} = \frac{t_1 - R t_2}{1 - R}, R \neq 1,$$

onde $R = b_1(n, \delta)/b_2(n, \delta)$, eles mostraram que quando R é conhecido (existem muitos casos onde R é conhecido e pode ser escrito como $R = f_1(n)/f_2(n)$), a quantidade $\hat{\delta}$ é um estimador não viciado de δ . Foi também mostrado que o estimador considerado $\hat{\delta}$ tem as mesmas propriedades assintóticas do Jackknife. Por fim, eles apresentaram três exemplos de aplicação desta técnica e sugeriram, como uma área de pesquisa, o estudo da seleção da quantidade R .

Uma variedade de trabalhos também tem sido realizada com ênfase no estudo da efetividade dos métodos de reamostragem. Adams, Gray e Watkins (1971) investigaram a efetividade dos estimadores jackknife de ordens n^{-1} e n^{-2} , como definidos em Schucany, Gray e Owen (1971). Eles deram caracterizações dos valores esperados dos estimadores jackknife de ordens n^{-1} e n^{-2} em termos do vício do estimador original e determinaram sobre quais condições podemos comparar cada uma dessas estatísticas com o estimador original com o intuito de calcular suas efetividades na redução do vício. Relações entre os vícios destes estimadores também foram investigadas.

Young e Daniels (1990) estudaram dois problemas de estimação de probabilidades caudais através do bootstrap e notaram que para tamanhos de amostra pequenos, o bootstrap é sistematicamente viciado. Segundo eles, o procedimento bootstrap envolve reamostragem de uma distribuição $\hat{F}$ que é de suporte finito, enquanto que F pode não ser. Então, pode-se esperar que em certas circunstâncias, tal como a estimação de pequenas probabilidades caudais, isto leve a um notável vício.

Outra área de aplicação dos métodos de reamostragem é na redução de vícios em estimadores razão. Rao (1965) encontrou a variância do estimador razão da forma $r = \bar{y}/\bar{x}$, assumindo um modelo de regressão onde a relação de y em x é linear e x é normalmente distribuído. Ele considerou o estimador jackknife r_Q obtido de uma amostra dividida em g grupos aleatórios de tamanho m , $n = mg$, e mostrou que para g geral, o vício de ordem n^{-2} e a variância de r_Q são funções decrescentes de g , tal que $g = n$ seja uma escolha ótima para o cálculo do vício. Rao e Webster (1966) apresentaram novos resultados considerando o modelo onde a regressão de y em x é linear mas, x/n tem distribuição gama($h, 1$). Eles obtiveram a variância e o vício de r_Q para g geral assumindo este modelo e mostraram numericamente que r_Q tem menor variância que r para $g > 2$.

Picard e Prum (1992) apresentaram e compararam vários métodos para correção do vício do E.M.V. na família exponencial curvada. Neste artigo, foi investigado o comportamento desses diferentes métodos quando a suposição de regularidade da matriz de informação de Fisher não é satisfeita. Os resultados foram apresentados em um estudo com modelos genéticos onde vícios muito importantes foram observados e que levaram a conclusões

práticas deficientes. Eles mostraram que tanto os métodos de correção de vício analítico e os de reamostragem (Jackknife e Bootstrap) não melhoraram o estimador de máxima verossimilhança. Entretanto, com algumas modificações cuidadosas nos métodos Jackknife e Bootstrap, podem-se obter resultados mais satisfatórios do que com os métodos originais comuns.

Para finalizar, uma questão de grande importância não discutida até o momento é a comparação dos vários métodos de redução de vício. Em relação aos métodos de reamostragem, Efron e Tibshirani (1993) argumentaram que quando se olha somente para as n amostras jackknife, o jackknife usa uma informação limitada sobre a estimativa $\hat{\theta}$, e então, pode-se esperar que ele seja menos eficiente que o bootstrap. De fato, eles mostram no Capítulo 20, através de um argumento geométrico, que o jackknife pode ser visto como uma aproximação do bootstrap. Ferrari e Cribari-Neto (1996) exploraram a relação entre correção de vício através do bootstrap e de expansões analíticas de séries em modelos uniparamétricos. Eles mostraram que existe uma estreita conexão entre a correção de vício baseada no bootstrap e a correção analítica derivada por Ferrari e outros (1996).

No Capítulo 2, discutimos os métodos de redução de vício propostos por Ferrari e outros (1996) e Firth (1993). O desenvolvimento destes métodos é apresentado e então aplicados a modelos da família exponencial uniparamétrica. Expressões para o vício de ordem n^{-1} e n^{-2} são também apresentadas para cada modelo considerado. No Capítulo 3, realizamos alguns estudos de simulação dos modelos descritos no Capítulo 2. Uma comparação entre os estimadores corrigidos obtidos através dos métodos propostos por Ferrari e outros (1996), Firth (1993) e os métodos de reamostragem Jackknife e Bootstrap é apresentada em termos de seus vícios e E.Q.M.'s. Finalmente, apresentamos nossas conclusões e alguns comentários a respeito do trabalho desenvolvido nesta tese.

Capítulo 2

Métodos Analíticos de Redução de Vício na Família Exponencial Uniparamétrica

2.1 Introdução

No Capítulo 1, vários métodos de redução de vícios (analíticos e de reamostragem) foram brevemente apresentados. Neste capítulo, apresentamos os métodos analíticos de redução de vícios propostos por Ferrari e outros (1996) e Firth (1993). Estes métodos serão aplicados a modelos da família exponencial uniparamétrica para obtenção de estimadores corrigidos por vício e seus resultados discutidos posteriormente.

Considere a função densidade ou de probabilidade

$$\pi(y; \theta) = \exp\{t(y; \theta)\}, \quad (2.1)$$

onde θ é um parâmetro escalar, e seja $\hat{\theta}$ o E.M.V. de θ . As expansões para o vício e variância de $\hat{\theta}$ até ordem n^{-2} , podem ser escritas, respectivamente, como

$$B(\theta) = \frac{B_1(\theta)}{n} + \frac{B_2(\theta)}{n^2}$$

e

$$V(\theta) = \frac{V_1(\theta)}{n} + \frac{V_2(\theta)}{n^2}.$$

Ferrari e outros (1996) propõem três estimadores corrigidos por vício definidos por

$$\hat{\theta}_1 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n}, \quad (2.2)$$

$$\hat{\theta}_2 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2(\hat{\theta})}{n^2} \quad (2.3)$$

e

$$\hat{\theta}_3 = \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2^*(\hat{\theta})}{n^2}, \quad (2.4)$$

onde $B_2^*(\theta) = B_2(\theta) - B_1(\theta)B'_2(\theta) - B''_1(\theta)V_1(\theta)/2$.

O estimador $\hat{\theta}_3$ foi obtido como uma modificação do estimador $\hat{\theta}_2$. Ferrari e outros (1996) mostram que $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}_3$ são não viciados até ordem n^{-1} , mas somente $\hat{\theta}_3$ é não viciado até ordem n^{-2} . Para apresentar expressões para os coeficientes $B_1(\theta)$, $B_2(\theta)$, $V_1(\theta)$ e $V_2(\theta)$ é necessário introduzir alguma notação para as derivadas com respeito a θ do logaritmo da função de verossimilhança para uma observação. Sejam

$$\begin{aligned} \kappa_{\theta\theta} &= E\{t''(y; \theta)\}, \kappa_{\theta,\theta} = -\kappa_{\theta\theta} = E[\{t'(y; \theta)\}^2], \kappa_{\theta\theta\theta} = E\{t'''(y; \theta)\}, \\ \kappa_{\theta\theta\theta\theta} &= E\{t^{iv}(y; \theta)\}, \kappa_{\theta\theta\theta\theta\theta} = E\{t^v(y; \theta)\}, \kappa_{\theta,\theta,\theta} = E[\{t'(y; \theta)\}^3], \\ \kappa_{\theta\theta,\theta} &= E\{t''(y; \theta)t'(y; \theta)\}, \kappa_{\theta\theta,\theta\theta} = E[\{t''(y; \theta)\}^2] - \kappa_{\theta\theta}^2, \kappa_{\theta,\theta\theta\theta} = E\{t'(y; \theta)t'''(y; \theta)\}, \\ \kappa_{\theta,\theta\theta\theta\theta} &= E\{t'(y; \theta)t^{iv}(y; \theta)\}, \kappa_{\theta,\theta\theta,\theta\theta} = [\{t'(y; \theta)\}\{t''(y; \theta)\}^2] - 2\kappa_{\theta,\theta\theta}\kappa_{\theta\theta}, \\ \kappa_{\theta\theta,\theta\theta\theta} &= E\{t''(y; \theta)t'''(y; \theta)\} - \kappa_{\theta\theta}\kappa_{\theta\theta\theta}, \kappa_{\theta\theta}^{(\theta)} = \frac{d\kappa_{\theta\theta}}{d\theta}, \kappa_{\theta\theta}^{(\theta\theta)} = \frac{d^2\kappa_{\theta\theta}}{d\theta^2}, \text{etc,} \end{aligned}$$

onde as linhas indicam derivadas com respeito a θ .

Usando resultados de Shenton e Bowman (1977, p. 44-47) com a notação para cumulantes introduzida acima e as relações de regularidade (Lawley, 1956) $\kappa_{\theta,\theta} = -\kappa_{\theta\theta}$, $\kappa_{\theta,\theta,\theta} = 2\kappa_{\theta\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)}$, $\kappa_{\theta,\theta\theta} = \kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}$, $\kappa_{\theta,\theta,\theta\theta} = \kappa_{\theta\theta\theta\theta} - 2\kappa_{\theta\theta\theta}^{(\theta)} + \kappa_{\theta\theta}^{(\theta\theta)} - \kappa_{\theta\theta,\theta\theta}$, pode ser mostrado depois de alguma álgebra que

$$\begin{aligned} B_1(\theta) &= -\frac{1}{2\kappa_{\theta\theta}^2}(\kappa_{\theta\theta\theta} - 2\kappa_{\theta\theta}^{(\theta)}), \\ B_2(\theta) &= -\frac{1}{8\kappa_{\theta\theta}^3}(\kappa_{\theta\theta\theta\theta} + 12\kappa_{\theta\theta,\theta\theta\theta} + 4\kappa_{\theta,\theta\theta\theta\theta} + 8\kappa_{\theta,\theta\theta,\theta\theta} + 4\kappa_{\theta,\theta,\theta\theta\theta}) + \frac{1}{12\kappa_{\theta\theta}^4}(13\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta\theta} \end{aligned} \quad (2.5)$$

$$\begin{aligned}
& -18\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta\theta} - 18\kappa_{\theta\theta\theta\theta}\kappa_{\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta,\theta\theta}\kappa_{\theta\theta}^{(\theta)} - 36\kappa_{\theta\theta\theta}\kappa_{\theta\theta\theta}^{(\theta)} + 36\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta}^{(\theta)} \\
& - 18\kappa_{\theta\theta\theta}\kappa_{\theta\theta}^{(\theta\theta)}) - \frac{1}{8\kappa_{\theta\theta}^5}(\kappa_{\theta\theta\theta}^3 - 48\kappa_{\theta\theta}^2\kappa_{\theta\theta}^{(\theta)} + 48\kappa_{\theta\theta\theta}(\kappa_{\theta\theta}^{(\theta)})^2), \tag{2.6}
\end{aligned}$$

$$V_1(\theta) = -\frac{1}{\kappa_{\theta\theta}}, \tag{2.7}$$

$$V_2(\theta) = -\frac{1}{\kappa_{\theta\theta}^3}(\kappa_{\theta\theta,\theta\theta} - \kappa_{\theta\theta\theta}^{(\theta)} + 2\kappa_{\theta\theta}^{(\theta\theta)}) + \frac{1}{2\kappa_{\theta\theta}^4}(10(\kappa_{\theta\theta}^{(\theta)})^2 - \kappa_{\theta\theta\theta}^2 - 4\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta}). \tag{2.8}$$

Utilizando expansões em série de Taylor, pode-se obter o vício, a variância e o erro quadrático médio dos estimadores $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}_3$ dados em (2.2) a (2.4). Ferrari e outros (1996) também mostram que os erros quadráticos médios (E.Q.M.'s) dos três estimadores corrigidos são iguais até ordem n^{-2} e dados por

$$\text{EQM}(\hat{\theta}_1) = \text{EQM}(\hat{\theta}_2) = \text{EQM}(\hat{\theta}_3) = \frac{V_1(\theta)}{n} + \frac{1}{n^2}\{V_2(\theta) - 2V_1(\theta)B_1'(\theta)\}.$$

Assim, entre $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}_3$, o estimador $\hat{\theta}_3$ deve ser preferido, pois possui menor vício. Para se optar entre $\hat{\theta}$ e $\hat{\theta}_3$ é necessário comparar os E.Q.M.'s destes estimadores. O E.Q.M. do estimador $\hat{\theta}$ é dado por

$$\text{EQM}(\hat{\theta}) = \frac{V_1(\theta)}{n} + \frac{1}{n^2}\{V_2(\theta) + B_1(\theta)^2\} + O(n^{-3}).$$

A diferença entre os E.Q.M.'s de $\hat{\theta}$ e $\hat{\theta}_3$ pode então ser escrita como

$$\text{EQM}(\hat{\theta}) - \text{EQM}(\hat{\theta}_3) = \frac{1}{n^2}\{B_1(\theta)^2 + 2V_1(\theta)B_1'(\theta)\} + O(n^{-3}).$$

A escolha entre $\hat{\theta}$ e $\hat{\theta}_3$ deve ser baseada no sinal de $\Delta = \Delta(\theta) = B_1(\theta)^2 + 2V_1(\theta)B_1'(\theta)$. Valores negativos de $\Delta(\theta)$ indicam que $\hat{\theta}$ deve ser o escolhido, enquanto que valores positivos indicam a escolha de $\hat{\theta}_3$. Esta estratégia de escolha do “melhor” estimador é adequada quando Δ não depende de θ , mas quando Δ varia com θ é aconselhável basear o critério na análise do sinal de $\Delta(\hat{\theta})$. Isto conduz à definição de um novo estimador

$$\bar{\theta} = \hat{\theta}1_{\{\Delta(\hat{\theta}) \leq 0\}} + \hat{\theta}_31_{\{\Delta(\hat{\theta}) > 0\}}, \tag{2.9}$$

onde $1_{\{\cdot\}}$ é a função indicadora. Um tópico de pesquisa interessante seria o desenvolvimento do E.Q.M. deste estimador. Fórmulas simples para $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$, $V_1(\theta)$, $V_2(\theta)$ e $\Delta(\theta)$ na família exponencial uniparamétrica são apresentadas na Seção 2.2.

Discutiremos agora o método de remoção de vício de ordem n^{-1} proposto por Firth (1993). Este método baseia-se em uma modificação conveniente da função escore para remover o termo de ordem n^{-1} da expansão do vício $B(\theta)$, dada em (1.1), do E.M.V. Em problemas regulares, o E.M.V. é definido como uma solução da equação de verossimilhança

$$l'(\theta) = U(\theta) = 0,$$

onde $l(\theta) = \log L(\theta)$ é o logaritmo da função de verossimilhança.

A base do trabalho de Firth está na idéia de que o vício de $\hat{\theta}$ pode ser reduzido introduzindo um pequeno vício na função escore. A modificação apropriada para $U(\theta)$ é dada por uma simples geometria de triângulo, ilustrada na Figura 2.1.

Figura 2.1: Modificação da função escore não-viciada.

Se $\hat{\theta}$ está sujeito a um vício positivo $B(\theta)$, a função escore é deslocada para baixo para cada ponto θ por uma quantidade $i(\theta)B(\theta)$, onde $-i(\theta) = U'(\theta)$. Isto define a função escore modificada dada por

$$U^*(\theta) = U(\theta) - i(\theta)B(\theta) \quad (2.10)$$

e um estimador modificado θ^* dado como solução de $U^*(\theta) = 0$. No caso de um vetor de parâmetros, $U^*(\theta)$ deve ser lido como uma equação vetorial, onde $i(\theta)$ é a matriz de informação de Fisher.

O vício $B(\theta)$ que aparece em (2.10) é, em geral, desconhecido, mas como vimos anteriormente é muitas vezes possível obter o termo de ordem n^{-1} , $B_1(\theta)/n$, deste vício. Empregando

notação e método de McCullagh (1987, §7.3), Firth (1993) mostra que se θ^* é solução de

$$U(\theta) - i(\theta) \frac{B_1(\theta)}{n} = 0, \quad (2.11)$$

onde $i(\theta)$ pode tanto representar a informação total observada quanto a esperada, então o termo de ordem n^{-1} é eliminado do vício do estimador. A expressão que Firth dá para $B_1(\theta)$ é

$$B_1(\theta) = -\frac{1}{2\kappa_{\theta,\theta}^2}(\kappa_{\theta\theta,\theta} + \kappa_{\theta,\theta,\theta}).$$

Utilizando o fato de que $\kappa_{\theta,\theta} = -\kappa_{\theta\theta}$, $\kappa_{\theta,\theta,\theta} = 2\kappa_{\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)}$ e $\kappa_{\theta\theta,\theta} = \kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}$, pode ser mostrado, depois de alguma álgebra, que

$$-\frac{1}{2\kappa_{\theta,\theta}^2}(\kappa_{\theta\theta,\theta} + \kappa_{\theta,\theta,\theta}) = -\frac{1}{2\kappa_{\theta\theta}^2}(\kappa_{\theta\theta\theta} - 2\kappa_{\theta\theta}^{(\theta)}),$$

como dado em (2.5).

Desta forma, o estimador obtido como solução de (2.11) é equivalente, até ordem n^{-1} , aos estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ dados em (2.2) e (2.3) bem como aos estimadores jackknife e bootstrap obtidos com uma única iteração e dados em (1.3) e (1.6), respectivamente. Em amostras grandes, as estimativas do parâmetro θ obtidas pelos diversos métodos serão aproximadamente iguais mas, para amostras pequenas, diferenças consideráveis poderão ser notadas. Consequentemente, os vícios dos diversos estimadores poderão diferir bastante se o tamanho da amostra for pequeno. Estudos de simulação apresentados no Capítulo 3 mostrarão este fato.

Um tópico interessante para pesquisa futura seria obter uma modificação na função escore que remove não só o termo de ordem n^{-1} do vício do estimador mas também o vício de ordem n^{-2} . Este estimador seria então equivalente, até ordem n^{-2} , ao estimador $\hat{\theta}_3$ dado em (2.4) e aos estimadores jackknife e bootstrap obtidos com duas iterações e dados em (1.4) e (1.7), respectivamente.

2.2 A Família Exponencial Uniparamétrica

Nesta seção, nos restringimos ao desenvolvimento dos métodos apresentados na Seção 2.1 quando aplicados à família exponencial uniparamétrica. Sejam $y_1, \dots, y_n$ n variáveis aleatórias

i.i.d., com função densidade ou de probabilidade na família exponencial uniparamétrica, isto é,

$$\pi(y; \theta) = \frac{1}{\xi(\theta)} \exp\{-\alpha(\theta)d(y) + v(y)\}, \quad (2.12)$$

onde θ é um parâmetro escalar e $\xi(\cdot)$, $\alpha(\cdot)$, $d(\cdot)$ e $v(\cdot)$ são funções conhecidas. Suponha que $\alpha = \alpha(\theta)$ e $\xi = \xi(\theta)$ têm as cinco primeiras derivadas contínuas com respeito a θ , $\xi(\theta)$ é positivo e $d\alpha(\theta)/d\theta$ e $d\beta(\theta)/d\theta$ são diferentes de zero para todos valores de θ no espaço paramétrico, onde $\beta = \beta(\theta)$ é definido como

$$\beta = \frac{d\xi(\theta)}{d\theta} \left(\xi(\theta) \frac{d\alpha(\theta)}{d\theta} \right)^{-1}. \quad (2.13)$$

A função densidade (ou de probabilidade) dada em (2.12) é um caso especial daquela definida em (2.1) com $t(y; \theta) = -\log \xi(\theta) - \alpha(\theta)d(y) + v(y)$. O logaritmo da função de verossimilhança total de θ pode ser escrito como

$$l(\theta) = -n \log \xi(\theta) - \alpha(\theta) \sum d(y_i) + \sum v(y_i).$$

Depois de alguma álgebra, pode-se mostrar que a função escore $U(\theta)$ é dada por

$$U(\theta) = -n\alpha'(\beta + \bar{d}),$$

com β dado em (2.13) e $\bar{d} = \sum d(y_i)/n$. Temos então que $t'(y; \theta) = -\alpha'\{\beta + d(y)\}$ e como $E\{t'(y; \theta)\} = 0$ segue-se que $\mu = E\{d(y)\} = -\beta$. A estimativa $\hat{\theta}$ de máxima verossimilhança é então obtida da equação

$$n^{-1} \sum d(y_i) = -\beta(\hat{\theta}). \quad (2.14)$$

Pode ser também mostrado que

$$\begin{aligned} t''(y; \theta) &= -\alpha''\{\beta + d(y)\} - \alpha'\beta', \\ t'''(y; \theta) &= -\alpha'''\{\beta + d(y)\} - 2\alpha''\beta' - \alpha'\beta'', \\ t^{iv}(y; \theta) &= -\alpha^{iv}\{\beta + d(y)\} - 3(\alpha'''\beta' + \alpha''\beta'') - \alpha'\beta''', \\ t^v(y; \theta) &= -\alpha^v\{\beta + d(y)\} - 4\alpha^{iv}\beta' - \alpha''\beta^{iv} - \alpha'''\beta''' - 3(\alpha''\beta''' + \alpha'''\beta''). \end{aligned}$$

Usando os resultados acima, os termos $B_1(\theta)$, $B_2(\theta)$, $V_1(\theta)$ e $V_2(\theta)$ dados em (2.5) a (2.8) e $B_2^*(\theta)$ foram obtidos por Cribari-Neto e outros (1995) e são dados por

$$B_1(\theta) = \frac{-\beta''}{2\alpha'\beta'^2}, \quad (2.15)$$

$$B_2(\theta) = \frac{1}{24\alpha'^3\beta'^5}(-12\beta'\alpha''\beta'^2 - 33\alpha'\beta'^3 + 4\beta'2\alpha''\beta'' + 26\alpha'\beta'\beta''\beta'' - 3\alpha'\beta'^2\beta'^4), \quad (2.16)$$

$$B_2^*(\theta) = \frac{1}{24\alpha'^3\beta'^5}(12\beta'^2\alpha''^2\beta'' + 18\alpha'\beta'\alpha''\beta''^2 + 15\alpha'^2\beta'^3 - 6\alpha'\beta'^2\beta''\alpha'' - 8\alpha'\beta'^2\alpha''\beta''^3 - 16\alpha'^2\beta'\beta''\beta''^3 + 3\alpha'^2\beta'^2\beta'^4), \quad (2.17)$$

$$V_1(\theta) = \frac{1}{\alpha'\beta'}, \quad (2.18)$$

$$V_2(\theta) = \frac{1}{2\alpha'^3\beta'^4}(2\beta'\alpha''\beta'' + 5\alpha'\beta'^2 - 2\alpha'\beta'\beta''^3). \quad (2.19)$$

Essas expressões têm a vantagem de requererem somente o conhecimento de α, β e de suas primeiras derivadas com respeito a θ . Como apresentado na Seção 2.1, os E.Q.M.'s de $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\theta}_3$ até ordem n^{-2} são iguais e a diferença dos E.Q.M.'s de $\hat{\theta}$ e $\hat{\theta}_3$ é dada por $\text{EQM}(\hat{\theta}) - \text{EQM}(\hat{\theta}_3) = \Delta(\theta)/n^2 + O(n^{-3})$, onde, na família exponencial uniparamétrica,

$$\Delta(\theta) = \frac{4\beta'\alpha''\beta'' + 9\alpha'\beta''^2 - 4\alpha'\beta'\beta''^3}{4\alpha''^3\beta'^4}. \quad (2.20)$$

Da Seção 2.1. vimos que Firth modifica a função escore da seguinte forma

$$U^*(\theta) = U(\theta) - i(\theta)\frac{B_1(\theta)}{n}, \quad (2.21)$$

onde $i(\theta)$ é a informação (esperada ou observada) de Fisher, e o estimador corrigido é obtido da solução da equação $U^*(\theta) = 0$. Tomando $i(\theta) = E[-U'(\theta)] = -n\alpha'\beta'$, a informação esperada de Fisher, temos que $U^*(\theta) = 0$ se, e somente se,

$$\bar{d} = -\beta + \gamma/n, \quad (2.22)$$

onde $\gamma = \gamma(\theta) = \beta''/(2\alpha'\beta')$. Por outro lado, substituindo a informação esperada pela informação observada, $o(\theta) = -dU(\theta)/d\theta$, temos que $U^*(\theta) = 0$ se, e somente se,

$$\bar{d} = -\beta + \delta/n, \quad (2.23)$$

onde $\delta = \alpha'\beta'\beta''/\{2\alpha'^2\beta'^2 - \alpha''\beta''/n\}$. É interessante notar que

$$\delta = \beta''/\{2\alpha'\beta'\} + O(n^{-1}),$$

ou seja, $\delta = \gamma + O(n^{-1})$. Assim, a menos que n seja pequeno, a diferença entre as estimativas de máxima verossimilhança obtidas por (2.22) ou (2.23) devem ser bem próximas.

Os estimadores corrigidos do vício até ordem n^{-1} , utilizando as informações esperada e observada, são então obtidos através das equações (2.22) e (2.23), para cada caso especial da família exponencial uniparamétrica.

2.3 Casos Especiais

Nesta seção, utilizaremos as expressões para $\beta(\theta)$, $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$, $V_1(\theta)$, $V_2(\theta)$, $\gamma(\theta)$ e $\delta(\theta)$ dadas em (2.13), (2.15), (2.16), (2.17), (2.18), (2.19), (2.22) e (2.23), respectivamente, para obter fórmulas em forma fechada para E.M.V.'s corrigidos por vício para um grande número de distribuições da família exponencial uniparamétrica. Em alguns casos, os E.M.V.'s e os estimadores corrigidos por vício não apresentam forma fechada devendo então ser obtidos através de métodos numéricos tais como Newton Raphson ou "Scoring de Fisher". As expressões para $\Delta(\theta)$ dado em (2.20) são também apresentadas para cada um desses casos especiais da família exponencial. Todas estas quantidades foram obtidas com a ajuda do programa MAPLE V (Abell e Braselton, 1994). O programa desenvolvido, de fácil utilização, é apresentado no Apêndice 1. As fórmulas para $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ também foram obtidas por Cribari-Neto e outros (1995).

Muitas das distribuições consideradas aqui são bastante conhecidas e têm um vasto campo de aplicação prática em diversas áreas, tais como, engenharia, biologia, zoologia, economia e medicina, entre outros. As distribuições apresentadas nos casos (i) a (viii) envolvem variáveis aleatórias discretas enquanto nos casos (ix) a (xxiv) são consideradas variáveis aleatórias contínuas. Maiores detalhes e propriedades de todas as distribuições apresentadas aqui podem ser obtidas em Aubin (1995) ou, alternativamente em, "Discrete Distributions" e "Continuous Univariate Distributions" volumes I e II, publicados pela Houghton Mifflin Company, 1970, de autoria de N. L. Johnson e S. Kotz.

Os seguintes casos particulares são considerados (em diversos casos, usa-se a notação: $\bar{y} = \sum_{i=1}^n y_i/n$):

(i) **Binomial** ($0 \leq \theta \leq 1$, $m \in N$, m conhecido, $y = 0, 1, 2, \dots, m$): $\alpha(\theta) = -\log\{\theta/(1-\theta)\}$, $\xi(\theta) = (1-\theta)^{-m}$, $d(y) = y$, $v(y) = \log \binom{m}{y}$;

$$\beta(\theta) = -m\theta, B_1(\theta) = B_2(\theta) = B_2^*(\theta) = 0,$$

$$\begin{aligned} V_1(\theta) &= -\theta(\theta-1)/m, V_2(\theta) = \Delta(\theta) = \gamma(\theta) = \delta(\theta) = 0, \\ \hat{\theta} &= \hat{\theta}_1 = \tilde{\theta}_2 = \bar{y}/m. \end{aligned}$$

onde $\hat{\theta}_1$ e $\tilde{\theta}_2$ são os estimadores corrigidos por vício obtidos como solução de (2.22) e (2.23).

(ii) **Binomial Negativa** ($0 \leq \theta < 1, r > 0, r$ conhecido, $y = 0, 1, 2, \dots$): $\alpha(\theta) = -\log(\theta)$,
 $\xi(\theta) = (1-\theta)^{-r}, d(y) = y, v(y) = \log \binom{r+y-1}{y}$;

$$\begin{aligned} \beta(\theta) &= \frac{r\theta}{\theta-1}, B_1(\theta) = -\frac{\theta(1-\theta)}{r}, B_2(\theta) = \frac{\theta(1-\theta)(1-2\theta)}{r^2}, B_2^*(\theta) = -\frac{\theta(1-\theta)^2}{r^2}, \\ V_1(\theta) &= \frac{\theta(1-\theta)^2}{r}, V_2(\theta) = \frac{2\theta(3\theta-1)(1-\theta)^2}{r^2}, \\ \Delta(\theta) &= \frac{\theta(5\theta-2)(1-\theta)^2}{r^2}, \gamma(\theta) = \frac{\theta}{\theta-1}, \delta(\theta) = -\frac{nr\theta}{(\theta-1)(-nr-1+\theta)}, \\ \hat{\theta} &= \frac{\bar{y}}{r+\bar{y}}, \hat{\theta}_1 = \frac{\bar{y}}{r+\bar{y}-(1/n)}, \\ \tilde{\theta}_2 &= \frac{(nr\bar{y}+2\bar{y}+nr^2) - (n^2r^2\bar{y}^2+2n^2r^3\bar{y}+n^2r^4-4r\bar{y})^{1/2}}{2(\bar{y}+r)}. \end{aligned}$$

(iii) **Poisson** ($\theta > 0, y = 0, 1, 2, \dots$): $\alpha(\theta) = -\log(\theta), \xi(\theta) = e^\theta, d(y) = y, v(y) = -\log(y!)$;

$$\begin{aligned} \beta(\theta) &= -\theta, B_1(\theta) = B_2(\theta) = B_2^*(\theta) = 0, \\ V_1(\theta) &= \theta, V_2(\theta) = \Delta(\theta) = \gamma(\theta) = \delta(\theta) = 0, \\ \hat{\theta} &= \tilde{\theta}_1 = \tilde{\theta}_2 = \bar{y}. \end{aligned}$$

(iv) **Poisson Truncada** ($\theta > 0, y = 1, 2, \dots$): $\alpha(\theta) = -\log(\theta), \xi(\theta) = e^\theta(1-e^{-\theta}), d(y) = y, v(y) = -\log(y!)$;

$$\begin{aligned} \beta(\theta) &= \frac{\theta}{e^{-\theta}-1}, \\ B_1(\theta) &= \frac{\theta(e^{-\theta}-1)\{\theta(e^{-2\theta}+e^{-\theta})+2e^{-2\theta}-2e^{-\theta}\}}{2(\theta e^{-\theta}+e^{-\theta}-1)^2}, \\ B_2(\theta) &= -\{24(\theta e^{-\theta}+e^{-\theta}-1)^5\}^{-1}\theta(e^{-\theta}-1)\{\theta^4(10e^{-6\theta}+2e^{-5\theta}+2e^{-4\theta}+10e^{-3\theta}) \\ &\quad +\theta^3(52e^{-6\theta}-44e^{-5\theta}-28e^{-3\theta}+20e^{-2\theta}+\theta^2(97e^{-6\theta}-217e^{-5\theta}+66e^{-4\theta} \\ &\quad +134e^{-3\theta}-83e^{-2\theta}+3e^{-\theta})+\theta(52e^{-6\theta}-272e^{-5\theta}+520e^{-4\theta}-448e^{-3\theta} \\ &\quad +164e^{-2\theta}-16e^{-\theta})-36e^{-6\theta}+156e^{-5\theta}-264e^{-4\theta}+216e^{-3\theta}-84e^{-2\theta} \end{aligned}$$

$$+12\epsilon^{-\theta}\},$$

$$\begin{aligned} B_2^*(\theta) = & -\{24(\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1)^5\}^{-1}\theta(\epsilon^{-\theta} - 1)\{\theta^4(-2\epsilon^{-6\theta} + 2\epsilon^{-5\theta} + 2\epsilon^{-4\theta} - 2\epsilon^{-3\theta}) \\ & + \theta^3(-2\epsilon^{-6\theta} + 16\epsilon^{-5\theta} - 12\epsilon^{-4\theta} + 8\epsilon^{-3\theta} - 10\epsilon^{-2\theta}) + \theta^2(19\epsilon^{-6\theta} - 7\epsilon^{-5\theta} - 18\epsilon^{-4\theta} \\ & - 22\epsilon^{-3\theta} - 31\epsilon^{-2\theta} - 3\epsilon^{-\theta}) + \theta(52\epsilon^{-6\theta} - 116\epsilon^{-5\theta} + 16\epsilon^{-4\theta} + 128\epsilon^{-3\theta} \\ & - 100\epsilon^{-2\theta} + 20\epsilon^{-\theta}) + 48\epsilon^{-6\theta} - 216\epsilon^{-5\theta} + 384\epsilon^{-4\theta} - 336\epsilon^{-3\theta} + 144\epsilon^{-2\theta} \\ & - 24\epsilon^{-\theta}\}, \end{aligned}$$

$$V_1(\theta) = -\frac{\theta(\epsilon^{-\theta} - 1)^2}{\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1},$$

$$\begin{aligned} V_2(\theta) = & \{2(\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1)^4\}^{-1}\theta(\epsilon^{-\theta} - 1)^2\{\theta^3(3\epsilon^{-4\theta} + 2\epsilon^{-3\theta} + 3\epsilon^{-2\theta}) \\ & + \theta^2(10\epsilon^{-4\theta} - 6\epsilon^{-3\theta} - 6\epsilon^{-2\theta} + 2\epsilon^{-\theta}) + \theta(-4\epsilon^{-4\theta} - 24\epsilon^{-3\theta} + 24\epsilon^{-2\theta} - 8\epsilon^{-\theta}) \\ & + 12\epsilon^{-3\theta} - 12\epsilon^{-2\theta} + 4\epsilon^{-\theta}\}, \end{aligned}$$

$$\begin{aligned} \Delta(\theta) = & \{4(\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1)^4\}^{-1}\theta(\epsilon^{-\theta} - 1)^2\{\theta^3(5\epsilon^{-4\theta} + 2\epsilon^{-3\theta} + 5\epsilon^{-2\theta}) \\ & + \theta^2(16\epsilon^{-4\theta} - 12\epsilon^{-3\theta} - 8\epsilon^{-2\theta} + 4\epsilon^{-\theta}) + \theta(12\epsilon^{-4\theta} - 40\epsilon^{-3\theta} + 44\epsilon^{-2\theta} - 16\epsilon^{-\theta}) \\ & - 8\epsilon^{-4\theta} + 24\epsilon^{-3\theta} - 24\epsilon^{-2\theta} + 8\epsilon^{-\theta}\}, \end{aligned}$$

$$\gamma(\theta) = -\frac{\theta\{\theta(\epsilon^{-2\theta} + \epsilon^{-\theta}) + 2\epsilon^{-2\theta} - 2\epsilon^{-\theta}\}}{2(\epsilon^{-\theta} - 1)(\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1)},$$

$$\begin{aligned} \delta(\theta) = & -n\{\theta(\epsilon^{-2\theta} + \epsilon^{-\theta}) + 2\epsilon^{-2\theta} - 2\epsilon^{-\theta}\}(\theta\epsilon^{-\theta} + \epsilon^{-\theta} - 1)\theta/\{(\epsilon^{-\theta} - 1)[\theta^2(2n\epsilon^{-2\theta}) \\ & + \theta(-\epsilon^{-3\theta} + 4n\epsilon^{-2\theta} - 4n\epsilon^{-\theta} + \epsilon^{-\theta}) - 2\epsilon^{-3\theta} + (2n + 4)\epsilon^{-2\theta} \\ & - (4n + 2)\epsilon^{-\theta} + 2n]\}. \end{aligned}$$

(v) **Série Logarítmica** ($0 < \theta < 1, y = 1, 2, \dots$): $\alpha(\theta) = -\log(\theta), \xi(\theta) = -\log(1 - \theta),$
 $d(y) = y, v(y) = -\log(y!);$

$$\beta(\theta) = -\frac{\theta}{(\theta - 1)\log(1 - \theta)},$$

$$B_1(\theta) = \frac{\theta(1 - \theta)\log(1 - \theta)[2\{\log(1 - \theta)\}^2 + (\theta + 2)\log(1 - \theta) + 2\theta]}{2\{\theta + \log(1 - \theta)\}^2},$$

$$\begin{aligned} B_2(\theta) = & -[24\{\theta + \log(1 - \theta)\}^5]^{-1}\theta(1 - \theta)\log(1 - \theta)[(48\theta - 24)\{\log(1 - \theta)\}^6 \\ & + (26\theta^2 + 164\theta - 60)\{\log(1 - \theta)\}^5 + (14\theta^3 + 296\theta^2 + 36\theta - 24)\{\log(1 - \theta)\}^4 \\ & + (3\theta^4 + 136\theta^3 + 312\theta^2 - 48\theta)\{\log(1 - \theta)\}^3 + (28\theta^4 + 276\theta^3)\{\log(1 - \theta)\}^2 \\ & + (60\theta^4 + 48\theta^3)\log(1 - \theta) + 24\theta^4], \end{aligned}$$

$$\begin{aligned} B_2^*(\theta) = & [24\{\theta + \log(1 - \theta)\}^5]^{-1}\theta(1 - \theta)\{\log(1 - \theta)\}^2[(24\theta - 24)\{\log(1 - \theta)\}^5 \\ & + (22\theta^2 + 52\theta - 72)\{\log(1 - \theta)\}^4 + (16\theta^3 + 136\theta^2 - 120\theta - 24)\{\log(1 - \theta)\}^3 \end{aligned}$$

$$\begin{aligned}
& +(3\theta^4 + 80\theta^3 - 72\theta)\{\log(1 - \theta)\}^2 + (20\theta^4 + 72\theta^3 - 72\theta^2)\log(1 - \theta) \\
& + 24\theta^4 - 24\theta^3], \\
V_1(\theta) &= -\frac{\theta(1 - \theta)^2\{\log(1 - \theta)\}^2}{\theta + \log(1 - \theta)} \\
V_2(\theta) &= \theta(1 - \theta)^2\{\log(1 - \theta)\}^2[(12\theta - 4)\{\log(1 - \theta)\}^4 + (10\theta^2 + 20\theta - 4)\{\log(1 - \theta)\}^3 \\
& + (3\theta^3 + 36\theta^2)\{\log(1 - \theta)\}^2 + (12\theta^3 + 12\theta^2)\{\log(1 - \theta)\} \\
& + 8\theta^5]/2\{\theta + \log(1 - \theta)\}^4, \\
\Delta(\theta) &= \theta(1 - \theta)^2\{\log(1 - \theta)\}^2[(20\theta - 8)\{\log(1 - \theta)\}^4 + (16\theta^2 + 32\theta - 8)\{\log(1 - \theta)\}^3 \\
& + (5\theta^3 + 60\theta^2 - 4\theta)\{\log(1 - \theta)\}^2 + (20\theta^3 + 16\theta^2)\log(1 - \theta) \\
& + 12\theta^3]/4\{\theta + \log(1 - \theta)\}^4, \\
\gamma(\theta) &= \frac{\theta\{2\{\log(1 - \theta)\}^2 + (\theta + 2)\log(1 - \theta) + 2\theta\}}{2(\theta - 1)(\theta + \log(1 - \theta))\log(1 - \theta)}, \\
\delta(\theta) &= n\theta(\theta + \log(1 - \theta))\{2\{\log(1 - \theta)\}^2 + (\theta + 2)\log(1 - \theta) + 2\theta\}/\{(\theta - 1)\log(1 - \theta) \\
& + (2\theta - 2)\{\log(1 - \theta)\}^3 + (\theta^2 + \theta + 2n - 2)\{\log(1 - \theta)\}^2 + (2\theta^2 + 4n\theta \\
& - 2\theta)\log(1 - \theta) + 2n\theta^2\}.
\end{aligned}$$

(vi) **Série de Potências** ($\theta > 0, a_y \geq y = 0, 1, 2, \dots$): $\alpha(\theta) = -\log(\theta), \xi(\theta) = \sum_{y=0}^{\infty} a_y \theta^y$,
 $d(y) = y, v(y) = \log(a_y)$;

$$\begin{aligned}
\beta(\theta) &= -\theta g, \\
B_1(\theta) &= -\frac{\theta(2g' + \theta g'')}{2(g + \theta g')^2}, \\
B_2(\theta) &= -\theta\{ \theta^4(3g'^2 g^{iv} + 33g''^3 - 26g'g''^2 g''') + \theta^3(108g'g''^2 - 36g'^2 g''' - 26gg''g''') \\
& + 6gg'g^{iv}) + \theta^2(204g'^2 g'' - 90gg''^2 + 3g^2 g^{iv} - 20gg'g''') + \theta(16g^2 g''' + 216g'^3 \\
& - 180gg'g'') - 48gg'^2 + 12g^2 g''\}/24(g + \theta g')^5, \\
B_2^*(\theta) &= \theta\{ \theta^4(3\theta^4 g'^2 g^{iv} + 15\theta^4 g''^3 - 16\theta^4 g'g''g''') + \theta^3(6\theta^3 gg'g^{iv} - 16\theta^3 gg''g''') \\
& + 24\theta^3 g'g''^2 - 12\theta^3 g'^2 g'') + \theta^2(3\theta^2 g^2 g^{iv} + 8\theta^2 gg'g''') + 20\theta^2 g^2 g''' + 36\theta^2 g'^2 g'' \\
& - 66\theta^2 gg''^2) + 48\theta g'^3 - 72gg' + 24g^2 g'' - 120gg'g''\}/24(g + \theta g')^5, \\
V_1(\theta) &= \frac{\theta}{g + \theta g'}, \\
V_2(\theta) &= -\frac{\theta(-5\theta^3 g''^2 + 2\theta^3 g'g''' + 2\theta^2 g''' - 12\theta^2 g'g'' - 16\theta g'^2 + 8\theta gg'' + 4gg''')}{2(g + \theta g')^4}, \\
\Delta(\theta) &= -\frac{\theta(-9\theta^3 g''^2 + 4\theta^3 g'g''' + 4\theta^2 gg''' - 20\theta^2 g'g'' - 28\theta g'^2 + 16\theta gg'' + 8gg'')}{4(g + \theta g')^4},
\end{aligned}$$

$$\gamma(\theta) = -\frac{\theta(2g' + \theta g'')}{2(g + \theta g')}.$$

$$\delta(\theta) = \frac{n\theta(g + \theta g')(2g' + \theta g'')}{2n\theta g'^2 + 4n\theta g g' + 2n g^2 + \theta g'' + 2g'}.$$

onde $g = g(\theta) = d \log \xi(\theta)/d\theta$.

(vii) **Zeta** ($\theta > 0, y = 1, 2, \dots$): $\alpha(\theta) = \theta + 1, \xi(\theta) = \text{Zeta}(\theta + 1), d(y) = y, v(y) = 0$;

$$\beta(\theta) = g,$$

$$B_1(\theta) = -\frac{g''}{2g'^2},$$

$$B_2(\theta) = -\frac{(33g''^3 - 26g'g''g''' + 3g'^2g^{iv})}{24g'^5},$$

$$B_2^*(\theta) = \frac{15g''^3 - 16g'g''g''' + 3g'^2g^{iv}}{24g'^5},$$

$$V_1(\theta) = \frac{1}{g'},$$

$$V_2(\theta) = \frac{5g''^2 - 2gg''''}{2g'^4},$$

$$\Delta(\theta) = \frac{9g''^2 - 4g'g''''}{4g'^4},$$

$$\gamma(\theta) = \frac{g''}{2g'},$$

$$\delta(\theta) = \frac{g''}{2g'},$$

onde $\text{Zeta}(\cdot)$ é a função Zeta de Riemann, isto é, $\text{Zeta}(\theta) = \sum_{i=1}^{\infty} i^{-\theta}$ (ver, e.g., Patterson, 1988) e $g = g(\theta) = d \log \text{Zeta}(\theta + 1)/d\theta$.

(viii) **Hipergeométrica Não-Central** ($\theta > 0, m_1, m_2, r$ inteiro positivo conhecido,

$a = \max\{0, r - m_2\} \leq y \leq \min\{m_1, r\} = b$): $\alpha(\theta) = \theta, \xi(\theta) = D_0(\theta), d(y) = -y,$

$$v(y) = \log \left\{ \binom{m_1}{y} \binom{m_2}{r-y} \right\};$$

$$\beta(\theta) = \frac{D_1(\theta)}{D_0(\theta)},$$

$$B_1(\theta) = \frac{D_0(-D_0^2 D_3 + 3D_0 D_1 D_2 - 2D_1^3)}{2(D_0 D_2 - D_1)^2},$$

$$B_2(\theta) = \{24(D_0 D_2 - D_1^2)^5\}^{-1} D_0(-24D_1^9 - 94D_0^4 D_1^3 D_3^2 - 92D_0^2 D_1^6 D_3 - 75D_0^3 D_1^3 D_2^3$$

$$\begin{aligned}
& -120D_0^2D_1^5D_2^2 + 108D_0D_1^7D_2 - 48D_0^5D_2^3D_3 + 144D_0^4D_1D_2^4 - 37D_0^3D_1^5D_4 \\
& -3D_0^6D_2^2D_5 - 3D_0^4D_1^4D_5 + 193D_0^5D_1D_2D_3^2 - 309D_0^4D_1^2D_2^2D_3 + 350D_0^3D_1^4D_2D_3 \\
& +26D_0^6D_2D_3D_4 - 63D_0D_5^5D_1D_2^2D_4 + 100D_0^4D_1^3D_2D_4 - 26D_0^5D_1^2D_3D_4 \\
& +6D_0^5D_1^2D_2D_5 - 33D_0^6D_3^3),
\end{aligned}$$

$$\begin{aligned}
B_2^*(\theta) = & \{24(D_0D_2 - D_1^2)^5\}^{-1}D_0^3(15D_0^4D_3^3 - 71D_0^3D_2D_1D_3^2 + 93D_0^2D_1^2D_2^2D_3 \\
& -82D_0D_1^4D_2D_3 - 16D_0^4D_2D_3D_4 + 33D_0^3D_1D_2^2D_4 - 50D_0^2D_1^3D_2D_4 \\
& +16D_0^3D_1^2D_3D_4 + 3D_0^2D_1^4D_5 - 6D_0^3D_1^2D_2D_5 + 26D_0^2D_1^3D_3^2 + 16D_1^6D_3 \\
& +51D_0D_1^3D_2^3 - 12D_1^5D_2^2 + 18D_0^3D_2^3D_3 - 54D_0^2D_1D_2^4 + 17D_0D_1^5D_2^4 + 3D_0^4D_2^2D_5),
\end{aligned}$$

$$V_1(\theta) = \frac{D_0^2}{(D_0D_2 - D_1^2)},$$

$$\begin{aligned}
V_2(\theta) = & \{24(D_0D_2 - D_1^2)^4\}^{-1}D_0^2(5D_0^4D_3^2 - 22D_0^3D_1D_2D_3 + 12D_0^2D_1^3D_3 \\
& +15D_0^2D_1^2D_2^2 - 24D_0D_1^4D_2 + 8D_1^6 - 2D_0^4D_2D_4 + 6D_0^3D_2^3 + 2D_0^3D_1^2D_4),
\end{aligned}$$

$$\begin{aligned}
\Delta(\theta) = & \{4(D_0D_2 - D_1^2)^4\}^{-1}D_0^2(9D_0^4D_3^2 - 38D_0^3D_1D_2D_3 + 20D_0^2D_1^3D_3 \\
& +21D_0^2D_1^2D_2^2 - 36D_0D_1^4D_2 + 12D_1^6 - 4D_0^4D_2D_4 + 12D_0^3D_2^3 + 4D_0^3D_1^2D_4),
\end{aligned}$$

$$\gamma(\theta) = \frac{D_0^2D_3 - 3D_0D_1D_2 + 2D_1^3}{2D_0(D_0D_2 - D_1^2)},$$

$$\delta(\theta) = \frac{D_0^2D_3 - 3D_0D_1D_2 + 2D_1^3}{2D_0(D_0D_2 - D_1^2)},$$

onde $D_p = D_p(\theta) = \sum_{y=a}^b y^p \binom{m_1}{y} \binom{m_2}{r-y} e^{\theta y}$, $p = 0, 1, 2, 3, 4, 5$.

(ix) **Maxwell** ($\theta > 0, y > 0$): $\alpha(\theta) = (2\theta^2)^{-1}$, $\xi(\theta) = \theta^3$, $d(y) = y^2$, $v(y) = \log(y^2\sqrt{2/\pi})$;

$$\begin{aligned}
\beta(\theta) &= -3\theta^2, B_1(\theta) = -\frac{\theta}{12}, B_2(\theta) = \frac{\theta}{288}, B_2^*(\theta) = -\frac{\theta}{288}, \\
V_1(\theta) &= \frac{\theta^2}{6}, V_2(\theta) = -\frac{\theta^2}{72}, \Delta(\theta) = -\frac{\theta^2}{48}, \gamma(\theta) = -\frac{\theta^2}{2}, \delta(\theta) = -\frac{2n\theta^2}{4n+1}, \\
\hat{\theta} &= \sqrt{\frac{\bar{y}}{3}}, \tilde{\theta}_1 = \sqrt{\frac{2n\bar{y}}{(6n-1)}}, \tilde{\theta}_2 = \sqrt{\frac{(4n+1)\bar{y}}{(12n+1)}},
\end{aligned}$$

(x) **Gama** ($k > 0, \theta > 0, y > 0$):

a) k conhecido: $\alpha(\theta) = \theta$, $\xi(\theta) = \theta^{-k}$, $d(y) = y$, $v(y) = (k-1)\log(y) - \log\{\Gamma(k)\}$;

$$\beta(\theta) = -\frac{k}{\theta}, B_1(\theta) = \frac{\theta}{k}, B_2(\theta) = \frac{\theta}{k^2}, B_2^*(\theta) = 0,$$

$$\begin{aligned}
V_1(\theta) &= \frac{\theta^2}{k}, V_2(\theta) = \frac{4\theta^2}{k^2}, \Delta(\theta) = \frac{3\theta^2}{k^2}, \gamma(\theta) = \delta(\theta) = -\frac{1}{\theta}, \\
\hat{\theta} &= \frac{k}{\bar{y}}, \hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta} - \frac{1}{n\bar{y}},
\end{aligned}$$

b) θ conhecido: $\alpha(\theta) = -k + 1, \xi(\theta) = \theta^{-k}\Gamma(k), d(y) = \log(y), v(y) = -\theta y;$

$$\begin{aligned}
\beta(k) &= -\psi(k) + \log(\theta), B_1(k) = -\frac{\psi''(k)}{2\psi'(k)^2}, \\
B_2(k) &= -\frac{(33\psi''(k)^3 - 2\psi'(k)\psi''(k)\psi'''(k) + 3\psi'(k)^2\psi^{iv}(k))}{24\psi'(k)^5}, \\
B_2^*(k) &= \frac{15\psi''(k)^3 - 16\psi'(k)\psi''(k)\psi'''(k) + 3\psi'(k)^2\psi^{iv}(k)}{24\psi'(k)^5}, \\
V_1(k) &= \frac{1}{\psi'(k)}, \\
V_2(k) &= \frac{5\psi''(k)^2 - 2\psi'(k)\psi'''(k)}{2\psi'(k)}, \\
\Delta(k) &= \frac{9\psi''(k)^2 - 4\psi'(k)\psi'''(k)}{4\psi'(k)^4}, \\
\gamma(k) &= -\frac{\psi''(k)}{2\psi'(k)}, \\
\delta(k) &= -\frac{\psi''(k)}{2\psi'(k)},
\end{aligned}$$

onde $\Gamma(\cdot)$ é a função gama e $\psi(\cdot)$ é a função digama.

(xi) **Distribuições do Sistema Burr** ($\theta > 0, b > 0, b$ conhecido, $y > 0$): $\alpha(\theta) = \theta,$
 $\xi(\theta) = c(\theta)/\theta, d(y) = -\log G(y), v(y) = \log\{ | d \log G(y)/dy | \};$

$$\begin{aligned}
\beta(\theta) &= \frac{\theta \log(c) - 1}{\theta}, B_1(\theta) = B_2(\theta) = \theta, B_2^*(\theta) = 0, \\
V_1(\theta) &= \theta^2, V_2(\theta) = 4\theta^2, \Delta(\theta) = 3\theta^2, \gamma(\theta) = -\frac{1}{\theta}, \delta(\theta) = -\frac{1}{\theta}, \\
\hat{\theta} &= \frac{1}{\log(c) + \bar{y}}, \hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta} - \frac{1}{n \log(c) + n\bar{y}},
\end{aligned}$$

onde $c(\cdot)$ e $G(\cdot)$ são funções reais positivas e $\bar{y} = \sum_{i=1}^n \log G(y_i)/n$. Diferentes escolhas para $c(\theta)$ e $G(y)$ levam a diferentes distribuições, entretanto, nem todas fazem parte da família exponencial; ver Burr (1942).

(xii) **Rayleigh** ($\theta > 0, y > 0$): $\alpha(\theta) = \theta^{-2}, \xi(\theta) = \theta^2, d(y) = y^2, v(y) = \log(2y)$;

$$\begin{aligned}\beta(\theta) &= -\theta^2, B_1(\theta) = -\frac{\theta}{8}, B_2(\theta) = \frac{\theta}{128}, B_2^*(\theta) = -\frac{\theta}{128}, \\ V_1(\theta) &= \frac{\theta^2}{4}, V_2(\theta) = -\frac{\theta^2}{32}, \Delta(\theta) = -\frac{3\theta^2}{64}, \gamma(\theta) = -\frac{\theta^2}{64}, \delta(\theta) = -\frac{2n\theta^2}{(8n+3)}, \\ \hat{\theta} &= \sqrt{\bar{y}}, \tilde{\theta}_1 = \sqrt{\frac{4n\bar{y}}{(4n-1)}}, \tilde{\theta}_2 = \sqrt{\frac{(8n+3)\bar{y}}{(8n+1)}},\end{aligned}$$

onde $\bar{y} = \sum_{i=1}^n y_i^2/n$.

(xiii) **Pareto** ($\theta > 0, k > 0, k$ conhecido, $y > k$): $\alpha(\theta) = \theta + 1, \xi(\theta) = (\theta k^\theta)^{-1}, d(y) = \log(y), v(y) = 0$;

$$\begin{aligned}\beta(\theta) &= -\frac{(1 + \theta \log(k))}{\theta}, B_1(\theta) = B_2(\theta) = \theta, B_2^*(\theta) = 0, \\ V_1(\theta) &= \theta^2, V_2(\theta) = 4\theta^2, \Delta(\theta) = 3\theta^2, \gamma(\theta) = \delta(\theta) = -\theta^{-1}, \\ \hat{\theta} &= \log(\phi/\tilde{y})^{-1}, \\ \tilde{\theta}_1 &= \tilde{\theta}_2 = \hat{\theta} - \frac{1}{n \log(\tilde{y}/k)},\end{aligned}$$

onde $\tilde{y}$ é a média geométrica dos $y_{i:s}$.

(xiv) **Weibull** ($\theta > 0, \phi > 0, \phi$ conhecido, $y > 0$): $\alpha(\theta) = \theta^{-\phi}, \xi(\theta) = \theta^\phi, d(y) = y^\phi, v(y) = \log(\phi) + (\phi - 1)\log(y)$;

$$\begin{aligned}\beta(\theta) &= -\theta^\phi, \\ B_1(\theta) &= \frac{\theta(1 - \phi)}{2\phi^2}, \\ B_2(\theta) &= -\frac{\theta(2\phi^3 - 9\phi^2 + 10\phi - 3)}{24\phi^4}, \\ B_2^*(\theta) &= -\frac{\theta(2\phi^3 - 3\phi^2 - 2\phi + 3)}{24\phi^4}, \\ V_1(\theta) &= \frac{\theta^2}{\phi^2}, V_2(\theta) = \frac{\theta^2(\phi^2 - 4\phi + 3)}{2\phi^4}, \\ \Delta(\theta) &= \frac{\theta^2(\phi^2 - 6\phi + 5)}{4\phi^4}, \gamma(\theta) = \frac{\theta^\phi(1 - \phi)}{2\phi}, \delta(\theta) = \frac{n\phi\theta^\phi(1 - \phi)}{2n\phi^2 + \phi^2 - 1}, \\ \hat{\theta} &= \bar{y}^{1/\phi}, \\ \tilde{\theta}_1 &= \left(\frac{2n\phi\bar{y}}{2n\phi - (\phi - 1)} \right)^{1/\phi}, \tilde{\theta}_2 = \left(\frac{(2n\phi^2 + \phi^2 - 1)\bar{y}}{2n\phi^2 + \phi - 1} \right)^{1/\phi},\end{aligned}$$

onde $\bar{y} = \sum_{i=1}^n y_i^\phi / n$.

(xv) **Potência** ($\theta > 0, \phi > 0, \phi$ conhecido, $0 < y < \phi$): $\alpha(\theta) = 1 - \theta, \xi(\theta) = \theta^{-1}\phi^\theta$.
 $d(y) = \log(y), v(y) = 0$;

$$\begin{aligned}\beta(\theta) &= -\frac{(1 + \theta \log(\phi))}{\theta}, B_1(\theta) = B_2(\theta) = \theta, B_2^*(\theta) = 0, \\ V_1(\theta) &= \theta^2, V_2(\theta) = 4\theta^2, \Delta(\theta) = 3\theta^2, \gamma(\theta) = \delta(\theta) = \theta^{-1}, \\ \hat{\theta} &= \log(\phi/\bar{y})^{-1}, \\ \tilde{\theta}_1 &= \tilde{\theta}_2 = \hat{\theta} - \frac{1}{n \log(\phi/\bar{y})},\end{aligned}$$

onde $\bar{y}$ é a média geométrica dos y_i 's.

(xvi) **Laplace** ($\theta > 0, -\infty < k < \infty, k$ conhecido, $y > 0$): $\alpha(\theta) = \theta^{-1}, \xi(\theta) = 2\theta$,
 $d(y) = |y - k|, v(y) = 0$;

$$\begin{aligned}\beta(\theta) &= -\theta, B_1(\theta) = B_2(\theta) = B_2^*(\theta) = 0, \\ V_1(\theta) &= \theta^2, V_2(\theta) = \Delta(\theta) = \gamma(\theta) = \delta(\theta) = 0, \\ \hat{\theta} &= \tilde{\theta}_1 = \tilde{\theta}_2 = \sum_{i=1}^n \frac{|y_i - k|}{n}.\end{aligned}$$

(xvii) **Valor Extremo** ($-\infty < \theta < \infty, \phi > 0, \phi$ conhecido, $-\infty < y < \infty$): $\alpha(\theta) = e^{\theta/\phi}$,
 $\xi(\theta) = \phi e^{-\theta/\phi}, d(y) = e^{-y/\phi}, v(y) = -y/\phi$;

$$\begin{aligned}\beta(\theta) &= e^{-\theta/\phi}, B_1(\theta) = \frac{\phi}{2}, B_2(\theta) = B_2^*(\theta) = \frac{\phi}{12}, \\ V_1(\theta) &= \phi^2, V_2(\theta) = \frac{\phi^2}{2}, \Delta(\theta) = \frac{\phi^2}{4}, \gamma(\theta) = -\frac{e^{-\theta/\phi}}{2}, \delta(\theta) = -\frac{ne^{-\theta/\phi}}{(2n+1)}, \\ \hat{\theta} &= -\phi \log(\bar{d}), \tilde{\theta}_1 = \hat{\theta} + \phi \log\left(1 - \frac{1}{2n}\right), \tilde{\theta}_2 = \hat{\theta} + \phi \log\left(1 - \frac{1}{(2n+1)}\right),\end{aligned}$$

onde $\bar{y} = \sum_{i=1}^n e^{-y_i/\phi}$.

(xviii) **Valor Extremo Truncada** ($\theta > 0, y > 0$): $\alpha(\theta) = \theta^{-1}, \xi(\theta) = \theta, d(y) = e^y - 1$,
 $v(y) = y$;

$$\begin{aligned}\beta(\theta) &= -\theta, B_1(\theta) = B_2(\theta) = B_2^*(\theta) = 0, \\ V_1(\theta) &= \theta^2, V_2(\theta) = \Delta(\theta) = \gamma(\theta) = \delta(\theta) = 0, \\ \hat{\theta} &= \tilde{\theta}_1 = \tilde{\theta}_2 = \frac{1}{n} \sum_{i=1}^n e^{y_i} - 1.\end{aligned}$$

(xix) **Lognormal** ($\theta > 0, \mu > 0, \mu$ conhecido, $y > 0$): $\alpha(\theta) = \theta^{-2}, \xi(\theta) = \theta$,

$d(y) = \{\log(y) - \mu\}^2/2, v(y) = -\log(y) - \{\log(2\pi)\}/2$;

$$\begin{aligned}\beta(\theta) &= -\frac{\theta^2}{2}, B_1(\theta) = -\frac{\theta}{4}, B_2(\theta) = \frac{\theta}{32}, B_2^*(\theta) = -\frac{\theta}{32}, \\ V_1(\theta) &= \frac{\theta^2}{2}, V_2(\theta) = -\frac{\theta^2}{8}, \Delta(\theta) = -\frac{3\theta^2}{16}, \gamma(\theta) = -\frac{\theta^2}{4}, \delta(\theta) = -\frac{n\theta^2}{(4n+3)}, \\ \hat{\theta} &= \left\{ \frac{1}{n} \sum_{i=1}^n \log(y_i - \mu)^2 \right\}^{1/2}, \tilde{\theta}_1 = \left\{ \frac{2n}{n(2n-1)} \sum_{i=1}^n \log(y_i - \mu)^2 \right\}^{1/2}, \\ \tilde{\theta}_2 &= \left\{ \frac{(4n+3)}{n(4n+1)} \sum_{i=1}^n \log(y_i - \mu)^2 \right\}^{1/2}.\end{aligned}$$

(xx) **Normal** ($\theta > 0, -\infty < \mu < \infty, -\infty < y < \infty$):

a) μ conhecido: $\alpha(\theta) = (2\theta)^{-1}, \xi(\theta) = \theta^{1/2}, d(y) = (y - \mu)^2, v(y) = -\{\log(2\pi)\}/2$;

$$\begin{aligned}\beta(\theta) &= -\theta, B_1(\theta) = B_2(\theta) = B_2^*(\theta) = 0, \\ V_1(\theta) &= 2\theta^2, V_2(\theta) = \Delta(\theta) = \gamma(\theta) = \delta(\theta) = 0, \\ \hat{\theta} &= \tilde{\theta}_1 = \tilde{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2.\end{aligned}$$

b) θ conhecido: $\alpha(\mu) = -\mu/\theta, \xi(\theta) = e^{\mu^2/2\theta}, d(y) = y, v(y) = -\{y^2 + \log(2\pi\theta)\}/2$;

$$\begin{aligned}\beta(\mu) &= -\mu, B_1(\mu) = B_2(\mu) = B_2^*(\mu) = 0, \\ V_1(\mu) &= \theta, V_2(\mu) = \Delta(\mu) = \gamma(\mu) = \delta(\mu) = 0, \\ \hat{\mu} &= \hat{\mu}_1 = \hat{\mu}_2 = \bar{y}.\end{aligned}$$

(xxi) **Normal Inversa** ($\theta > 0, \mu > 0, y > 0$):

a) μ conhecido: $\alpha(\theta) = \theta, \xi(\theta) = \theta^{-1/2}, d(y) = (y - \mu)^2/(2\mu^2 y), v(y) = -\{\log(2\pi y^3)\}/2$;

$$\begin{aligned}\beta(\theta) &= -(2\theta)^{-1}, B_1(\theta) = 2\theta, B_2(\theta) = 4\theta, B_2^*(\theta) = 0, \\ V_1(\theta) &= 2\theta^2, V_2(\theta) = 16\theta^2, \Delta(\theta) = 12\theta^2, \gamma(\theta) = \delta(\theta) = -\theta^{-1}, \\ \hat{\theta} &= \left\{ \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\mu^2 y_i} \right\}^{-1}, \\ \tilde{\theta}_1 &= \hat{\theta} - \frac{1}{n\bar{y}}, \tilde{\theta}_2 = \hat{\theta} - \frac{1}{n\bar{y}} - \frac{2}{n^2\bar{y}},\end{aligned}$$

onde $\bar{y} = \sum_{i=1}^n (y_i - \mu)^2 / (2n\mu^2 y_i)$.

b) θ conhecido: $\alpha(\mu) = \theta/2\mu^2, \xi(\theta) = e^{-\theta/\mu}, d(y) = y, v(y) = -\theta/2y + [\log\{\theta/(2\pi y^3)\}]/2$;

$$\begin{aligned}\beta(\mu) &= -\mu, B_1(\mu) = B_2(\mu) = B_2^*(\mu) = 0, \\ V_1(\mu) &= \mu^3/\theta, V_2(\mu) = \Delta(\mu) = \gamma(\mu) = \delta(\mu) = 0, \\ \hat{\mu} &= \tilde{\mu}_1 = \tilde{\mu}_2 = \bar{y}.\end{aligned}$$

(xxii) **McCullagh** ($\theta > -1/2, -1 \leq \mu \leq 1, \mu$ conhecido, $0 < y < 1$): $\alpha(\theta) = -\theta$,
 $\xi(\theta) = B(\theta + 1/2, 1/2)/4^\theta, d(y) = \log[y(1-y)/\{(1+\mu)^2 - 4\mu y\}], v(y) = -[\log\{y(1-y)\}]/2$;

$$\begin{aligned}\beta(\theta) &= \psi(\theta + 1/2) - \psi(\theta + 1) - \log(4), \\ B_1(\theta) &= -\frac{\psi''(\theta + 1/2) - \psi''(\theta + 1)}{2\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^2}, \\ B_2(\theta) &= [24\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^5]^{-1}[-33\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}^3 \\ &\quad + 26\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}\{\psi'''(\theta + 1/2) \\ &\quad - \psi'''(\theta + 1)\} - 3\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^2\{\psi^{iv}(\theta + 1/2) - \psi^{iv}(\theta + 1)\}], \\ B_2^*(\theta) &= [24\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^5]^{-1}[15\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}^3 \\ &\quad - 16\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}\{\psi'''(\theta + 1/2) \\ &\quad - \psi'''(\theta + 1)\} + 3\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^2\{\psi^{iv}(\theta + 1/2) - \psi^{iv}(\theta + 1)\}], \\ V_1(\theta) &= -\frac{1}{\psi'(\theta + 1/2) - \psi'(\theta + 1)}, \\ V_2(\theta) &= [2\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^4]^{-1}[5\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}^2 \\ &\quad - 2\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}\{\psi'''(\theta + 1/2) - \psi'''(\theta + 1)\}], \\ \Delta(\theta) &= [4\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}^4]^{-1}[9\{\psi''(\theta + 1/2) - \psi''(\theta + 1)\}^2 \\ &\quad - 4\{\psi'(\theta + 1/2) - \psi'(\theta + 1)\}\{\psi'''(\theta + 1/2) - \psi'''(\theta + 1)\}], \\ \gamma(\theta) &= -\frac{\psi''(\theta + 1/2) - \psi''(\theta + 1)}{2[\psi'(\theta + 1/2) - \psi'(\theta + 1)]}, \\ \delta(\theta) &= -\frac{\psi''(\theta + 1/2) - \psi''(\theta + 1)}{2[\psi'(\theta + 1/2) - \psi'(\theta + 1)]},\end{aligned}$$

onde $B(.,.)$ é a função beta (McCullagh, 1989).

(xxiii) **Von Mises** ($\theta > 0, 0 < \mu < 2\pi, \mu$ conhecido, $0 < y < 2\pi$): $\alpha(\theta) = -\theta, \xi(\theta) = 2\pi I_0(\theta)$,

$d(y) = \cos(y - \mu), v(y) = 0$;

$$\beta(\theta) = -r,$$

$$\begin{aligned}
B_1(\theta) &= -\frac{r''}{2r'^2}, \\
B_2(\theta) &= -\frac{(33r''^3 - 26r'r''r''' + 3r'^2r^{iv})}{24r'^5}, \\
B_2^*(\theta) &= \frac{15r''^3 - 16r'r''r''' + 3r'^2r^{iv}}{24r'^5}, \\
V_1(\theta) &= \frac{1}{r'}, \\
V_2(\theta) &= \frac{5r''^2 - 2r'r''}{2r'^4}, \\
\Delta(\theta) &= \frac{9r''^2 - 4r'r''}{4r'^4}, \\
\gamma(\theta) &= -\frac{r''}{2r'}, \\
\delta(\theta) &= -\frac{r''}{2r'}.
\end{aligned}$$

onde $I_\nu(\cdot)$ é a função Bessel modificada de ordem ν e $r = r(\theta) = I'_0(\theta)/I_0(\theta)$.

(xxiv) **Secante Hiperbólica Generalizada** ($-\pi/2 \leq \theta \leq \pi/2, 0 < y < 1, r > 0, r$ conhecido):

$$\alpha(\theta) = \theta, \xi(\theta) = \pi \{\sec(\theta)\}^r, d(y) = -\pi^{-1} - \log\{y/(1-y)\}, v(y) = -(1/2) \log\{y/(1-y)\};$$

$$\begin{aligned}
\beta(\theta) &= \frac{r \operatorname{sen}(\theta)}{\cos(\theta)}, \\
B_1(\theta) &= -\frac{\operatorname{sen}(\theta) \cos(\theta)}{r}, \\
B_2(\theta) &= \frac{\operatorname{sen}(\theta) \cos(\theta) \{10(\cos(\theta))^2 - 3\}}{3r^2}, \\
B_2^*(\theta) &= -\frac{2\operatorname{sen}(\theta)(\cos(\theta))^3}{3r^2}, \\
V_1(\theta) &= \frac{(\cos(\theta))^2}{r}, \\
V_2(\theta) &= -\frac{2(\cos(\theta))^2 \{3(\cos(\theta))^2 - 2\}}{r^2}, \\
\Delta(\theta) &= -\frac{(\cos(\theta))^2 \{5(\cos(\theta))^2 - 3\}}{r^2}, \\
\gamma(\theta) &= \frac{\operatorname{sen}(\theta)}{\cos(\theta)}, \\
\delta(\theta) &= \frac{\operatorname{sen}(\theta)}{\cos(\theta)},
\end{aligned}$$

$$\begin{aligned}\hat{\theta} &= \arctan \left\{ \frac{1}{r} \left\{ \frac{1}{\pi} + \frac{1}{n} \sum_{i=1}^n \log \left(\frac{y_i}{1-y_i} \right) \right\} \right\}, \\ \hat{\theta}_1 &= \hat{\theta}_2 = \arctan \left\{ \frac{n}{(nr-1)} \left(\frac{1}{\pi} + \frac{1}{n} \sum_{i=1}^n \log \left(\frac{y_i}{1-y_i} \right) \right) \right\}.\end{aligned}$$

2.4 Verificação das Fórmulas dos Vícios em Alguns Casos Especiais

Nesta seção, fazemos uma verificação da validade das equações (2.15) a (2.17) para alguns casos especiais. A idéia básica é que podemos encontrar o vício do E.M.V. $\hat{\theta}$, para estes casos, através de sua distribuição assintótica.

Seja, $\hat{\theta} = nk / \sum_{i=1}^n y_i$, o E.M.V. do parâmetro de escala para a distribuição gama, com $k > 0$ conhecido. Então, após alguns desenvolvimentos, temos que $\hat{\theta} = 2nk\theta/X$, onde $X \sim \chi_{2nk}^2$. Seja agora, $f(X) = 1/X$, onde $X \sim \chi_n^2$. Expandindo $f(X)$ em série de Taylor, em torno de μ , e tomando sua esperança, temos

$$E[f(X)] = f(\mu) + E(X-\mu)f'(\mu) + \frac{E(X-\mu)^2}{2!}f''(\mu) + \frac{E(X-\mu)^3}{3!}f'''(\mu) + \frac{E(X-\mu)^4}{4!}f^{(iv)}(\mu) + \dots$$

Utilizando algumas relações entre os momentos e cumulantes da distribuição χ_n^2 , ou seja, que $\mu_{3X} = \kappa_{3X} = 8n$ e $\mu_{4X} = \kappa_{4X} + 3\kappa_{2X}^2 = 48n + 12n^2$, temos depois de alguma álgebra, que

$$E[f(X)] = \frac{1}{n} + \frac{1}{n^2} + \frac{4}{n^3} + O(n^{-4}).$$

Logo,

$$E(\hat{\theta}) = 2nk\theta E[f(X)] = \theta + \frac{\theta}{(kn)} + \frac{\theta}{(kn)^2} + O(n^{-3}).$$

Desta forma, podemos verificar que para a distribuição gama com índice $k > 0$ conhecido e parâmetro de escala $\theta > 0$, as expressões para $B_1(\theta)$ e $B_2(\theta)$ estão de acordo com as obtidas de (2.15) e (2.16). Para verificar o termo $B_2^*(\theta)$ obtido de (2.17) percebe-se que

$$E(\hat{\theta}_3) = E \left\{ \hat{\theta} - \frac{B_1(\hat{\theta})}{n} - \frac{B_2^*(\hat{\theta})}{n^2} \right\} = \theta + O(n^{-3}).$$

Portanto, $\hat{\theta}_3$ é um estimador corrigido por vício até ordem n^{-2} .

Para a distribuição Pareto com $k > 0$ conhecido ($y > k$) e parâmetro escala $\theta > 0$, temos que $\hat{\theta} \sim 2n\theta/\chi_{2n}^2$ (Johnson e Kotz, 1970; pp. 233-249). Assim, utilizando o mesmo desenvolvimento que na distribuição gama, temos que $E(\hat{\theta}) = \theta + \theta/n + \theta/n^2 + O(n^{-3})$. Note que também, o vício de $\hat{\theta}_3$ é de ordem n^{-3} e, portanto, as expressões para $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ estão de acordo, respectivamente, com as expressões obtidas de (2.15) a (2.17). O E.M.V. $\hat{\theta}$ de, $\theta \in R$, para a distribuição de valor extremo com $\phi > 0$ conhecido é igual a $\hat{\theta} = -\phi \log(n^{-1} \sum_{i=1}^n e^{-y_i/\phi})$ e pode ser mostrado, facilmente, que $E(\hat{\theta}) = \theta + \phi/(2n) + \phi/(12n^2) + O(n^{-3})$ e que $E(\hat{\theta}_3 - \theta) = O(n^{-3})$ e, portanto, $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ estão de acordo com as correspondentes expressões obtidas através de (2.15) a (2.17).

Para a distribuição lognormal com parâmetro de locação $\mu > 0$ e parâmetro de escala $\theta > 0$, deve-se notar que $\hat{\theta} \sim \theta \sqrt{\chi_n^2/n}$. Desta forma, fazendo a expansão de $f(X) = \sqrt{X}$, onde $X \sim \chi_n^2$, em torno de μ , e tomando sua esperança temos, depois de alguma álgebra, que $E(\hat{\theta}) = \theta - \theta/(4n) + \theta/(32n^2) + O(n^{-3})$. Também, podemos mostrar que $E(\hat{\theta}_3 - \theta) = O(n^{-3})$ e, portanto, $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ estão de acordo, respectivamente, com as expressões dadas por (2.15) a (2.17). Temos também, que o E.M.V. de $\hat{\theta}$ para a distribuição normal inversa com média $\mu > 0$ e parâmetro de escala $\theta > 0$ tem distribuição $n\theta/\chi_n^2$. Assim, da mesma forma que no caso da distribuição gama, podemos mostrar que $E(\hat{\theta}) = \theta + 2\theta/n + 4\theta/n^2 + O(n^{-3})$ e que $E(\hat{\theta}_3) = \theta + O(n^{-3})$ e, portanto, $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ estão de acordo com as correspondentes expressões obtidas de (2.15) a (2.17).

É fácil verificar também que, para as distribuições binomial, Poisson, Laplace, valor extremo truncada e normal, $\hat{\theta}$ é um estimador não-viciado de θ e que, portanto, $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ devem ser nulos.

2.5 Expansões Assintóticas

Em algumas distribuições, a obtenção das correções de vício requer o cálculo de expressões longas e não-usuais, envolvendo funções tais como poligama, Bessel e zeta. Estas expressões podem ser avaliadas através de um “software” como o MATHEMATICA ou MAPLE ou, alternativamente, através de aproximações obtidas por expansões assintóticas. Outra utilidade de tais expansões é proporcionar uma indicação do comportamento das correções com funções do parâmetro θ . Expansões para os termos $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$ e $\Delta(\theta)$ foram obtidas para as distribuições de Poisson truncada, série logarítmica, zeta, gama (com θ co-

nhecido), McCullagh e von Mises, fazendo o uso do software MAPLE (Abell e Braselton, 1994).

Para obter as aproximações para $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$ e $\Delta(\theta)$ para a distribuição Poisson truncada, foi necessário fazer o uso da expansão $e^\theta = 1 + \theta + \theta^2/2 + \theta^3/6 + \theta^4/24 + \dots$. Usando esta expansão, obtemos, para θ pequeno,

$$\begin{aligned} B_1(\theta) &= -\frac{\theta}{3} - \frac{2\theta^2}{9} - \frac{7\theta^3}{90} + \frac{8\theta^4}{405} + O(\theta^5), \\ B_2(\theta) &= \frac{2\theta}{9} - \frac{19\theta^2}{45} + \frac{149\theta^3}{405} - \frac{1793\theta^4}{8505} + O(\theta^5), \\ B_2^*(\theta) &= -\frac{\theta}{3} + \frac{56\theta^2}{135} - \frac{112\theta^3}{405} + \frac{83\theta^4}{630} + O(\theta^5), \\ \Delta(\theta) &= -\frac{4\theta}{3} + \frac{7\theta^2}{18} - \frac{82\theta^3}{45} - \frac{389\theta^4}{405} + O(\theta^5). \end{aligned}$$

Considere agora a distribuição série logarítmica. Para valores pequenos de θ , temos que

$$\log(1 - \theta) = -\theta - \frac{\theta^2}{2} - \frac{\theta^3}{3} - \frac{\theta^4}{4} - \frac{\theta^5}{5} - \frac{\theta^6}{6} + O(\theta^7).$$

Usando este resultado, obtemos então para a distribuição série logarítmica com θ pequeno

$$\begin{aligned} B_1(\theta) &= -\frac{5\theta}{3} + \frac{19\theta^2}{18} + \frac{11\theta^3}{45} + \frac{17\theta^4}{162} + O(\theta^5), \\ B_2(\theta) &= \frac{23\theta}{9} - \frac{407\theta^2}{90} + \frac{59\theta^3}{162} - \frac{2654\theta^4}{8505} + O(\theta^5), \\ B_2^*(\theta) &= -\frac{7\theta}{3} + \frac{379\theta^2}{135} - \frac{53\theta^3}{324} - \frac{127\theta^4}{1134} + O(\theta^5), \\ \Delta(\theta) &= -\frac{20\theta}{3} + \frac{67\theta^2}{3} - \frac{818\theta^3}{45} + \frac{325\theta^4}{324} + O(\theta^5). \end{aligned}$$

Para a distribuição zeta, considere a expansão da função zeta de Riemann, denotada por $\zeta(\cdot)$, dada em Abramowitz e Stegun (1970, 23.2.5), escrita como

$$\zeta(\theta + 1) = \frac{1}{\theta} + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \gamma_j \theta^j,$$

onde

$$\gamma_j = \lim_{m \rightarrow \infty} \left\{ \sum_{k=1}^m \frac{(\log k)^j}{k} - \frac{(\log k)^{j+1}}{j+1} \right\},$$

$j = 0, 1, \gamma_0$ sendo a constante de Euler. A expansão acima implica que

$$\zeta'(\theta + 1) = -\frac{1}{\theta^2} - \gamma_1 + \gamma_2\theta - \frac{\gamma_3\theta^2}{2} + \frac{\gamma_4\theta^3}{6} - \frac{\gamma_5\theta^4}{24} + \frac{\gamma_6\theta^5}{120} - \frac{\gamma_7\theta^6}{720} + O(\theta^7).$$

Como $g(\theta) = d\log\zeta(\theta + 1)/d\theta$, segue que

$$g(\theta) = -\frac{1}{\theta} + \gamma_0 - (2\gamma_1 + \gamma_2^2)\theta + (\gamma_0^3 + 3\gamma_0\gamma_1 + \frac{3\gamma_2}{2})\theta^2 - (\gamma_0^4 + 4\gamma_0^2\gamma_1 + 2\gamma_0\gamma_2 + 2\gamma_1^2 + \frac{2\gamma_2}{3})\theta^3 + O(\theta^4).$$

Temos então, para valores pequenos de θ , que

$$\begin{aligned} B_1(\theta) &= \theta + 2(\gamma_0^2 + 2\gamma_1)\theta^3 + O(\theta^4), \\ B_2(\theta) &= \theta + 12(\gamma_0^2 + 2\gamma_1)\theta^3 + O(\theta^4), \\ B_2^*(\theta) &= -2(\gamma_0^2 + 2\gamma_1)\theta^3 + O(\theta^4), \\ \Delta(\theta) &= 3\theta^2 + O(\theta^4). \end{aligned}$$

Considere agora a distribuição gama (com θ conhecido). Para valores pequenos de k , temos que

$$\begin{aligned} B_1(k) &= k - \frac{\pi^2 k^3}{3} + 5\text{Zeta}(3)k^4 - \frac{\pi^4 k^5}{60} + O(k^6), \\ B_2(k) &= k - 2\pi^2 k^3 + 44\text{Zeta}(3)k^4 + \frac{47\pi^4 k^5}{180} + O(k^6), \\ B_2^*(k) &= \frac{\pi^2 k^3}{3} - 11\text{Zeta}(3)k^4 + \frac{\pi^4 k^5}{36} + O(k^6), \\ \Delta(k) &= 3k^2 - 3\pi^2 k^4 + 54\text{Zeta}(3)k^5 + O(k^6), \end{aligned}$$

onde $\text{Zeta}(3) \approx 1.2020569$. Para valores grandes de k , temos que (Abramowitz, 1970; 6.4.12)

$$\psi'(k) = \frac{1}{k} + \frac{1}{2k^2} + \frac{1}{6k^3} - \frac{1}{30k^5} + \frac{1}{42k^7} - \frac{1}{30k^9} + \frac{5}{66k^{11}} - \frac{691}{2730k^{13}} + \dots$$

Usando esta expansão obtemos, para valores grandes de k ,

$$\begin{aligned} B_1(k) &= \frac{1}{2} - \frac{1}{24k^2} - \frac{1}{24k^3} + \frac{1}{480k^4} + \frac{11}{240k^5} + O(k^{-6}), \\ B_2(k) &= -\frac{1}{24} - \frac{1}{48k^2} - \frac{1}{288k^3} + \frac{1}{192k^4} + \frac{557}{8640k^5} + O(k^{-6}), \\ B_2^*(k) &= -\frac{1}{24} - \frac{1}{48k^2} + \frac{23}{288k^3} + \frac{25}{192k^4} - \frac{547}{8640k^5} + O(k^{-6}), \\ \Delta(k) &= \frac{1}{4} + \frac{1}{8k^2} + \frac{1}{8k^3} - \frac{119}{960k^4} - \frac{179}{480k^5} + O(k^{-6}). \end{aligned}$$

Fazendo o uso de expansão direta através do “software” MAPLE (Abell e Braselton, 1994) obtemos, para θ grande, a seguinte expansão assintótica para a distribuição de McCullagh:

$$B_1(\theta) = 2\theta + \frac{1}{2} + \frac{1}{8\theta^3} + O(\theta^{-4}),$$

$$\begin{aligned}
B_2(\theta) &= 4\theta + 1 + \frac{1}{4\theta} - \frac{1}{16\theta^2} - \frac{9}{32\theta^3} + O(\theta^{-4}), \\
B_2^*(\theta) &= \frac{1}{4\theta} - \frac{1}{16\theta^2} - \frac{41}{32\theta^3} + O(\theta^{-4}), \\
\Delta(\theta) &= 12\theta^2 + 6\theta + \frac{9}{4} - \frac{3}{2\theta^2} + \frac{3}{4\theta^3} + O(\theta^{-4})
\end{aligned}$$

Para a distribuição de von Mises, $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$ e $\Delta(\theta)$ envolvem a função de Bessel modificada de primeira ordem. Entretanto, é possível obter aproximações que não envolvem essa função. Para valores pequenos de θ , é possível obter essas expansões fazendo o uso da equação (3.4.46) dada em Mardia (1972, pág. 63)

$$r(\theta) = \frac{\theta}{2} \left\{ 1 - \frac{\theta^2}{8} + \frac{\theta^4}{48} - \frac{11\theta^6}{3072} + \frac{19\theta^8}{30720} - \frac{473\theta^{10}}{4423680} \dots \right\}.$$

As expansões para $B_1(\theta)$, $B_2(\theta)$, $B_2^*(\theta)$ e $\Delta(\theta)$ são então dadas por:

$$\begin{aligned}
B_1(\theta) &= \frac{3\theta}{4} + \frac{7\theta^3}{48} + O(\theta^5), \\
B_2(\theta) &= \frac{19\theta}{16} + \frac{505\theta^3}{768} + O(\theta^5), \\
B_2^*(\theta) &= -\frac{\theta}{4} - \frac{53\theta^3}{768} + O(\theta^5), \\
\Delta(\theta) &= 3 + \frac{55\theta^2}{16} + \frac{121\theta^4}{128} + O(\theta^6).
\end{aligned}$$

Para valores grandes de θ , é possível escrever uma expansão para $r(\theta)$ fazendo uso de expansões em Abramowitz e Stegun (1970, 9.7.1), e então usar este resultado para obter, para θ grande,

$$\begin{aligned}
B_1(\theta) &= 2\theta - \frac{1}{2} + \frac{15}{16\theta^2} + \frac{39}{8\theta^3} + O(\theta^{-4}), \\
B_2(\theta) &= 4\theta - 1 - \frac{3}{4\theta} - \frac{3}{\theta^2} - \frac{171}{16\theta^3} + O(\theta^{-4}), \\
B_2^*(\theta) &= -\frac{3}{4\theta} - \frac{27}{4\theta^2} - \frac{765}{16\theta^3} + O(\theta^{-4}), \\
\Delta(\theta) &= 12\theta^2 - 6\theta - \frac{15}{4} - \frac{45}{4\theta} - \frac{891}{16\theta^2} - \frac{3129237}{16384\theta^3} + O(\theta^{-4}).
\end{aligned}$$

2.6 Análise Gráfica

Na Seção 2.3 mostramos que, para muitos casos especiais, $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ variam com o parâmetro desconhecido θ . Foi também apresentado que é muito comum, na prática, a utilização do E.Q.M. como critério de decisão para a escolha entre dois estimadores. É então importante examinar $\Delta(\theta)$, que dá a diferença entre os E.Q.M.'s de $\hat{\theta}$ e $\hat{\theta}_3$, até ordem n^{-2} . Até ordem n^{-2} , $\Delta > 0$ quando $\text{EQM}(\hat{\theta}) > \text{EQM}(\hat{\theta}_3)$ e $\Delta \leq 0$ quando $\text{EQM}(\hat{\theta}) \leq \text{EQM}(\hat{\theta}_3)$. Para algumas distribuições, a expressão para Δ é muito simples, como é o caso das distribuições Rayleigh, para a qual, $\Delta = -3\theta^2/64$ que é negativo para todos os valores de θ , Pareto e potência, para as quais, $\Delta = 3\theta^2$, que é positivo para todos os valores de θ . Entretanto, a expressão para Δ , pode ser bastante complexa. A idéia é então construir gráficos de Δ contra θ para examinar quais regiões do espaço paramétrico correspondem a valores positivos, zero e negativos de Δ .

Nesta seção, apresentamos para alguns casos especiais, gráficos de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ mostrando como essas quantidades variam com o parâmetro desconhecido θ . Gráficos de $\Delta(\theta) = \text{EQM}(\hat{\theta}) - \text{EQM}(\hat{\theta}_3)$ contra θ são também apresentados. As seguintes distribuições foram consideradas: Binomial negativa ($r = 1$), Poisson truncada, série logarítmica, gama, Rayleigh, McCullagh, von Mises e secante hiperbólica generalizada ($r = 1$). As Figuras 2.2 a 2.17 apresentam, respectivamente, os gráficos dos vícios $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ e de $\Delta(\theta)$ contra θ , para cada uma dessas distribuições.

Figura 2.2: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição Binomial negativa.

Figura 2.3: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição Binomial negativa.

Figura 2.4: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição Poisson truncada.

Figura 2.5: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição Poisson truncada.

Figura 2.6: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição série logarítmica.

Figura 2.7: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição série logarítmica.

Figura 2.8: Gráfico de $B_1(k)$, $B_2(k)$ e $B_2^*(k)$ contra k , para a distribuição gama.

Figura 2.9: Gráfico de $\Delta(k)$ contra k , para a distribuição gama.

Figura 2.10: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição Rayleigh.

Figura 2.11: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição Rayleigh.

Figura 2.12: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição McCullagh.

Figura 2.13: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição McCullagh.

Figura 2.14: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição von Mises.

Figura 2.15: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição von Mises.

Figura 2.16: Gráfico de $B_1(\theta)$, $B_2(\theta)$ e $B_2^*(\theta)$ contra θ , para a distribuição sec. hip. generalizada.

Figura 2.17: Gráfico de $\Delta(\theta)$ contra θ , para a distribuição sec. hip. generalizada.

Como pode ser observado nos gráficos apresentados nas Figuras 2.2 a 2.17, os vícios de ordem n^{-1} e n^{-2} e $B_2^*(\theta)$ podem variar substancialmente dependendo do verdadeiro valor do parâmetro θ . Exceto para a distribuição gama (Figura 2.8), o comportamento de $B_2(\theta)$ e de $B_2^*(\theta)$ são bem diferentes. Em relação à quantidade $\Delta(\theta)$, a Figura 2.5 mostra que para a distribuição Poisson truncada, o E.Q.M. do estimador corrigido por vício $\hat{\theta}_3$ é menor que o E.Q.M. do estimador de máxima verossimilhança, para (aproximadamente) $\theta > 1.4$ e maior para (aproximadamente) $\theta < 1.4$. As Figuras 2.13 e 2.15 mostram, respectivamente, que para as distribuições McCullagh e von Mises, $\hat{\theta}_3$ domina $\hat{\theta}$ em termos de E.Q.M. A Figura 2.17 mostra que para a distribuição secante hiperbólica generalizada, Δ é negativo quando θ está próximo de zero, e positivo quando θ está próximo de $-\pi/2$ ou $\pi/2$.

Capítulo 3

Estudos de Simulação

3.1 Introdução

Neste capítulo, apresentamos alguns estudos de simulação visando mostrar a importância das correções de vício desenvolvidas. As seções seguintes apresentam resultados de simulação comparando o desempenho do E.M.V. e suas versões corrigidas (analiticamente e por reamostragem) apresentadas nos Capítulos 1 e 2. Foram consideradas as distribuições binomial negativa, gama, Rayleigh, Pareto, lognormal. Em cada uma destas distribuições, 5000 amostras foram geradas para diversas combinações, variando o tamanho da amostra n e o valor do parâmetro θ . Para cada experimento de simulação foram calculados os estimadores de máxima verossimilhança, os estimadores corrigidos por vício $\hat{\theta}_1$ e $\hat{\theta}_3$ (Ferrari e outros, 1996) dados, respectivamente, em (2.2) e (2.4); os estimadores corrigidos por vício obtidos como solução de (2.22) e (2.23) (Firth, 1993) e denotados, respectivamente, por $\tilde{\theta}_1$ e $\tilde{\theta}_2$; o estimador $\bar{\theta}$ dado em (2.9); e os estimadores corrigidos por vício através dos processos de reamostragem Jackknife e Bootstrap, $\tilde{\theta}_J^1$ e $\tilde{\theta}_B^1$, dados, respectivamente, em (1.3) e (1.6). As estimativas bootstrap foram calculadas usando bootstrap não-paramétrico com o processo de reamostragem baseado na função distribuição empírica. Os E.M.V.'s corrigidos por vício baseado no bootstrap foram calculados usando $N = 200$ réplicas bootstrap. Todas as simulações foram executadas através de programas desenvolvidos no “software” estatístico S-Plus, 1993. O Apêndice 2 apresenta, a título de ilustração, o programa utilizado para as

simulações da distribuição binomial negativa. Programas similares foram desenvolvidos para a simulação das demais distribuições.

3.2 A Distribuição Binomial Negativa

Muitos modelos diferentes dão origem à distribuição binomial negativa, e consequentemente, existe uma variedade de definições na literatura. Formalmente, a distribuição Binomial negativa pode ser definida em termos da expansão da expressão binomial negativa $(Q - P)^{-r}$, onde $Q = 1 + P$, $P > 0$, e r é um número real positivo; o $(x + 1)$ -ésimo termo na expansão produz $Pr[X = x]$. Vários tipos de parametrização para esta distribuição podem ser encontrados em Johnson e Kotz (1992). A distribuição binomial negativa tem aplicações em diversas áreas tais como na indústria (Shapiro e Zahedi, 1990), em processos estocásticos, em modelos de contagem, etc. O modelo binomial negativo é uma extensão direta do modelo geométrico e pode ser usado, por exemplo, para determinar a probabilidade de que Y ensaios sejam requeridos antes da r -ésima unidade defeituosa ser encontrada em um programa de inspeção. A variável aleatória binomial negativa com parâmetro r pode então ser considerada como uma soma de r variáveis geométricas independentes, $X_1, \dots, X_r$, com probabilidade comum, onde X_1 é o número de ensaios para o primeiro “sucesso”, X_2 é o número de ensaios do primeiro “sucesso” para o segundo “sucesso”, e assim por diante. A distribuição binomial negativa pode resultar também de outro tipo de situação. Dado um processo que pode ser modelado por uma distribuição de Poisson, tal como o número de defeitos Y em um produto, onde o parâmetro do processo (a taxa de ocorrência) é aleatoriamente mudado (tem uma distribuição gama), então a distribuição resultante de Y é binomial negativa (generalizada). Apesar de ser uma distribuição discreta, ela é bastante usada na análise de dados contínuos quando se tem dados super-dispersos. A análise de dados contínuos com super-dispersão, via a distribuição binomial negativa, é justificável se considerarmos somas de quantidades aleatórias ao longo de um intervalo de tempo, também aleatório, ao contrário do tempo ser fixado.

Consideramos a distribuição binomial negativa com parâmetros θ e r , onde $0 \leq \theta < 1$, é a probabilidade de fracasso e $r > 0$, o número de sucessos. A distribuição binomial negativa, neste caso, representa o número de ensaios onde não ocorre sucesso antes de ocorrer sucesso pela r -ésima vez, onde o evento “sucesso” tem probabilidade constante $(1 - \theta)$. Quando $r = 1$,

temos caracterizada a distribuição geométrica, que pode ser pensada como representando o número de ensaios até a ocorrência do primeiro sucesso, em experimentos de Bernoulli independentes, cuja probabilidade de sucesso é $(1 - \theta)$. A geração de números aleatórios com distribuição binomial negativa foi executada utilizando o gerador de números aleatórios do “software” S-Plus, 1993. As Tabelas 3.1, 3.2 e 3.3 apresentam os resultados das simulações para a distribuição binomial negativa considerando, respectivamente, as probabilidades de sucesso $(1 - \theta)$ iguais a 0.9, 0.5 e 0.1. Em cada tabela, são apresentados os valores médios e os erros quadráticos médios das estimativas dos parâmetros corrigidas e não-corrigidas por vício.

Como pode ser observado, as Tabelas 3.1 a 3.3 mostram que para amostras pequenas o vício do E.M.V. $\hat{\theta}$ pode ser bem considerável. Quando $\theta = 0.1$ (ver Tabela 3.1) o vício deste estimador é praticamente eliminado pelos estimadores corrigidos por vício. O mesmo acontece quando θ é igual a 0.5 e 0.9 (Tabelas 3.2 e 3.3). Pode-se observar ainda, que à medida que o número de sucessos r aumenta, o vício de $\hat{\theta}$ torna-se menor. Isto fica bastante claro se observarmos as fórmulas dos termos de vícios de ordens n^{-1} e n^{-2} , $B_1(\theta)$ e $B_2^*(\theta)$, na Seção 2.3. Outro ponto importante de se destacar é em relação ao E.Q.M. dos estimadores corrigidos em relação ao E.M.V. Como pode ser observado na Figura 2.3 (Seção 2.6), o E.Q.M. dos estimadores corrigidos por vício é maior que o E.Q.M. do E.M.V. quando $\theta = 0.1$ e menor quando $\theta = 0.9$. Para $\theta = 0.5$, é esperado acontecer o mesmo que em $\theta = 0.9$. Entretanto, isto só pode ser observado para os estimadores $\hat{\theta}_3$ e $\tilde{\theta}_2$ (Tabela 3.2), no caso em que o tamanho da amostra n é igual a 6. Ainda em relação às Tabelas 3.1 a 3.3, quando r e n são iguais a 1, o estimador corrigido por vício $\tilde{\theta}_1$ não apresenta estimativas, representado por NA. O que acontece, neste caso, é que algumas estimativas de $\tilde{\theta}_1$ são iguais a infinito, uma vez que pode ocorrer divisão por zero no cálculo destas estimativas (ver expressão de $\tilde{\theta}_1$, Seção 2.3, caso (ii)). Também, para n igual a 1 não é possível obter estimativas através do processo Jackknife. Neste caso, é fácil observar que esta estimativa não existe devido ao próprio processo de obtenção deste estimador. Para finalizar, dentre os estimadores corrigidos por vício até ordem n^{-1} , $\tilde{\theta}_2$ é o estimador que é aparentemente o menos viciado. Entretanto, em alguns casos, seu E.Q.M. é maior se comparado com os outros estimadores corrigidos. É importante ressaltar que o estimador $\hat{\theta}_3$ teve um desempenho relativamente melhor que os estimadores corrigidos por vício até ordem n^{-1} , como esperado.

Tabela 3.1: Resultados da simulação para a distribuição Binomial negativa, $\theta = 0.1$.

r	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_3$		$\bar{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_j$		$\hat{\theta}_R^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
1	1	0.05	0.03	0.08	0.05	0.09	0.07	0.09	0.07	NA	NA	0.10	0.09	-	-	0.05	0.03
	2	0.07	0.02	0.09	0.04	0.10	0.04	0.07	0.03	0.10	0.04	0.09	0.04	0.09	0.03	0.08	0.03
	3	0.08	0.02	0.10	0.03	0.10	0.03	0.08	0.08	0.10	0.03	0.10	0.03	0.09	0.02	0.09	0.02
	4	0.08	0.01	0.10	0.02	0.10	0.02	0.08	0.02	0.10	0.02	0.10	0.02	0.10	0.02	0.09	0.02
	5	0.09	0.01	0.10	0.02	0.10	0.02	0.09	0.01	0.10	0.02	0.10	0.02	0.10	0.02	0.09	0.01
	6	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	10	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
2	1	0.07	0.02	0.09	0.04	0.10	0.04	0.08	0.03	0.10	0.04	0.09	0.04	-	-	0.07	0.02
	2	0.08	0.01	0.10	0.02	0.10	0.02	0.08	0.02	0.10	0.02	0.10	0.02	0.10	0.02	0.09	0.02
	3	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.09	0.01
	4	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	5	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	6	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	10	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00
5	1	0.08	0.01	0.10	0.01	0.10	0.01	0.08	0.01	0.10	0.02	0.10	0.02	-	-	0.08	0.01
	2	0.09	0.01	0.10	0.01	0.10	0.01	0.09	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	3	0.09	0.00	0.10	0.01	0.10	0.01	0.09	0.00	0.10	0.01	0.10	0.01	0.10	0.01	0.10	0.01
	4	0.09	0.00	0.10	0.00	0.10	0.00	0.09	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00
	5	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00
	6	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00
	10	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00

Tabela 3.2: Resultados da simulação para a distribuição Binomial negativa, $\theta = 0.5$.

r	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_3$		$\bar{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_j^1$		$\hat{\theta}_F^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
1	1	0.31	0.14	0.42	0.19	0.46	0.22	0.46	0.22	NA	NA	0.49	0.25	-	-	0.31	0.14
	2	0.38	0.08	0.47	0.09	0.49	0.10	0.45	0.10	0.50	0.10	0.47	0.09	0.46	0.10	0.42	0.09
	3	0.42	0.05	0.48	0.06	0.49	0.06	0.46	0.07	0.50	0.06	0.49	0.06	0.49	0.07	0.46	0.06
	4	0.44	0.04	0.49	0.04	0.50	0.04	0.48	0.05	0.50	0.04	0.49	0.04	0.49	0.05	0.47	0.04
	5	0.45	0.03	0.49	0.03	0.50	0.03	0.48	0.04	0.50	0.03	0.49	0.03	0.50	0.03	0.48	0.03
	6	0.46	0.03	0.50	0.03	0.50	0.02	0.48	0.03	0.50	0.02	0.50	0.03	0.50	0.03	0.49	0.03
	10	0.48	0.01	0.50	0.01	0.50	0.01	0.50	0.02	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01
2	1	0.38	0.08	0.47	0.09	0.49	0.10	0.45	0.10	0.50	0.10	0.48	0.09	-	-	0.38	0.08
	2	0.44	0.04	0.49	0.04	0.50	0.04	0.48	0.04	0.50	0.04	0.50	0.04	0.49	0.04	0.47	0.04
	3	0.46	0.02	0.50	0.02	0.50	0.02	0.49	0.03	0.50	0.02	0.50	0.02	0.50	0.03	0.48	0.03
	4	0.47	0.02	0.50	0.02	0.50	0.02	0.49	0.02	0.50	0.02	0.50	0.02	0.50	0.02	0.50	0.02
	5	0.48	0.01	0.50	0.01	0.50	0.01	0.49	0.02	0.50	0.01	0.50	0.01	0.50	0.01	0.49	0.01
	6	0.48	0.01	0.50	0.01	0.50	0.01	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.49	0.01
	10	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01
5	1	0.45	0.03	0.49	0.03	0.50	0.03	0.48	0.04	0.50	0.03	0.49	0.03	-	-	0.45	0.03
	2	0.47	0.01	0.50	0.01	0.50	0.01	0.49	0.02	0.50	0.01	0.50	0.01	0.50	0.02	0.49	0.01
	3	0.48	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.49	0.01
	4	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01
	5	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01
	6	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01
	10	0.49	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.01

Tabela 3.3: Resultados da simulação para a distribuição Binomial negativa, $\theta = 0.9$.

r	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_3$		$\bar{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_J^1$		$\hat{\theta}_R^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
1	1	0.75	0.10	0.87	0.08	0.89	0.08	0.89	0.08	NA	NA	0.91	0.08	-	-	0.75	0.10
	2	0.84	0.03	0.89	0.02	0.90	0.02	0.90	0.02	0.90	0.02	0.90	0.02	0.92	0.03	0.88	0.03
	3	0.86	0.01	0.90	0.01	0.90	0.01	0.90	0.01	0.90	0.01	0.90	0.01	0.92	0.01	0.90	0.01
	4	0.87	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.91	0.01	0.90	0.01
	5	0.88	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	6	0.88	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	10	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
2	1	0.83	0.03	0.89	0.02	0.90	0.02	0.90	0.02	0.90	0.02	0.89	0.02	-	-	0.83	0.03
	2	0.87	0.01	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.91	0.01	0.89	0.01
	3	0.88	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	4	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	5	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	6	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	10	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
5	1	0.88	0.00	0.90	0.00	0.90	0.02	0.90	0.02	0.90	0.02	0.90	0.00	-	-	0.88	0.00
	2	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	3	0.89	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	4	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	5	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	6	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00
	10	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00	0.90	0.00

3.3 A Distribuição Gama

A distribuição gama $G(\theta, k)$, com parâmetro de forma $k > 0$ e parâmetro de escala $\theta > 0$, é usada na análise de dados contínuos não-negativos, os quais apresentam uma variância crescente com a média e, mais fundamentalmente, quando o coeficiente de variação dos dados for aproximadamente constante. A distribuição gama aparece naturalmente da teoria associada com variáveis aleatórias normalmente distribuídas, como a distribuição da soma de quadrados de variáveis normais unitárias independentes. Se $U_1, U_2, \dots, U_\nu$ são variáveis aleatórias normais unitárias independentes, então $\sum_{j=1}^{\nu} U_j^2$ tem uma distribuição χ^2 com ν graus de liberdade. O caso especial da distribuição qui-quadrado aparece quando fazemos $\theta = 1/2$ e $k = n/2$. Em trabalhos aplicados, a distribuição gama dá representações úteis de muitas situações físicas. O fato de que a soma de variáveis aleatórias independentes exponencialmente distribuídas tem uma distribuição gama, leva esta distribuição a aparecer na teoria de contagem aleatória e outros tópicos associados com processos aleatórios no tempo, em particular, em processos de precipitação meteorológica. O caso da distribuição exponencial, quando fazemos $k = 1$, desempenha importante papel na teoria de confiabilidade e no estudo da lei de falhas. Ela tem sido usada na indústria para fazer ajustes “realísticos” da representação do tempo de vida de equipamentos eletrônicos e de outros materiais em situações de testes de vida. Neste caso, a vida média é θ e para se obter uma “boa” estimativa, o número de observações deve ser tão grande quanto possível. Entretanto, este número deve ser tão pequeno quanto possível se considerarmos o tempo e o custo dos testes de tais equipamentos. Podemos também considerar a variável aleatória obtida tomando-se o logaritmo da variável aleatória com distribuição gama $G(\theta, k)$. Neste caso, temos a distribuição log-gama, amplamente utilizada em análise de sobrevivência. A geração de números aleatórios com distribuição gama foi executada utilizando o gerador de números aleatórios do “software” S-Plus.

As Tabelas 3.4, 3.5 e 3.6 apresentam, respectivamente, os valores médios e os erros quadrático médios de todas as estimativas corrigidas e não-corrigidas por vício, considerando valores de θ iguais a 1, 10 e 50.

Tabela 3.4: Resultados da simulação para a distribuição gama, $k = 0.1$.

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_1 = \hat{\theta}_2$		$\hat{\theta}_1$		$\hat{\theta}_B$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
1	5	13499.6	422750385001.0	-13499.6	422750438999.0	-13499.6	422750438999.0	-210194128.0	1.711049e20	-3.864873e20	7.468459e62
	10	8.4	8212.3	0.0	1.0	0.0	1.0	-49.5	534946.5	-448869.1	3.389586e14
	15	2.9	147.9	1.0	16.0	1.0	16.0	-2.7	1095.5	-33.86	266754.7
	20	1.9	8.9	2.0	2.0	1.0	2.0	-0.3	138.8	-1.5	413.1
	25	1.7	4.2	1.0	1.4	1.0	1.4	0.5	7.8	0.3	9.9
	30	1.5	2.1	1.0	0.8	1.0	0.8	0.7	2.4	0.7	2.1
10	5	134995.8	4.227504e13	-134995.8	4.227504e13	-134995.8	4.227504e13	-2101941278.0	1.711049e22	-3.864873e30	7.468459e64
	10	84.0	821227.2	0.0	100.0	0.0	100.0	-495.2	53494653	-4488691	3.389586e16
	15	29.0	14794.4	9.7	1603.7	9.7	1603.7	-27.4	109544.6	-338.6	26675470
	20	19.3	892.8	9.7	202.0	9.7	202.0	-2.6	13878.6	-15.4	41312.0
	25	16.5	417.2	9.9	135.1	9.9	135.1	5.1	779.8	3.1	993.0
	30	14.8	205.5	9.9	81.3	9.9	81.3	7.2	243.9	6.9	213.3
50	5	674979.2	1.056876e15	-674979.2	1.056876e15	-674979.2	1.056876e15	-10509706388.0	4.277624e23	-1.932437e31	1.867115e66
	10	420.2	20530679.0	0.0	2500.0	0.0	2500.0	-2476.1	1337366316.0	-22443456.0	8.473965e17
	15	145.1	369859.2	48.4	40093.0	48.4	40093.0	-136.8	2738615.0	-1692.8	666886743.0
	20	96.3	22320.1	48.1	5048.8	48.1	5048.8	-13.2	346965.6	-77.0	1032800.0
	25	82.4	10429.9	49.4	3377.6	49.4	3377.6	25.3	19495.7	15.7	24826.1
	30	73.9	5138.6	49.2	2031.5	49.2	2031.5	36.0	6097.3	34.6	5332.0

Tabela 3.5: Resultados da simulação para a distribuição gama, $k = 1$.

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_j^1$		$\hat{\theta}_B^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
1	3	1.5	4.5	1.0	1.9	1.0	1.9	0.6	5.9	0.1	260.8
	4	1.3	1.0	1.0	0.5	1.0	0.5	0.8	1.6	0.8	6.5
	5	1.3	0.6	1.0	0.3	1.0	0.3	0.9	0.5	1.0	0.5
	10	1.1	0.2	1.0	0.1	1.0	0.1	1.0	0.1	1.0	0.1
	15	1.1	0.1	1.0	0.1	1.0	0.1	1.0	0.1	1.0	0.1
	20	1.1	0.1	1.0	0.1	1.0	0.1	1.0	0.1	1.0	0.1
10	25	1.0	0.1	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
	30	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
	3	15.2	447.4	10.1	186.8	10.1	186.8	5.8	593.3	1.1	26078.9
	4	13.4	101.7	10.1	50.7	10.1	50.7	8.3	161.3	8.1	651.1
	5	12.5	58.0	10.0	33.0	10.0	33.0	9.2	54.1	9.7	47.4
	10	11.1	16.6	10.0	12.5	10.0	12.5	9.9	13.3	10.0	13.2
50	15	10.7	9.5	10.0	7.9	10.0	7.9	9.9	8.1	10.0	8.2
	20	10.5	6.3	10.0	5.5	10.0	5.5	9.9	5.5	10.0	5.6
	25	10.4	4.9	10.0	4.4	10.0	4.4	9.3	4.4	10.0	4.5
	30	10.3	3.8	10.0	3.5	10.0	3.5	10.0	3.5	10.0	3.6
	3	76.0	11184.6	50.7	4670.9	50.7	4670.9	29.2	14832.8	5.4	651974.0
	4	67.1	2543.2	50.3	1266.8	50.3	1266.8	41.3	4031.9	40.3	16277.0
50	5	62.7	1449.2	50.2	824.2	50.2	824.2	46.1	1351.5	48.3	1185.7
	10	55.5	414.6	50.0	311.3	50.0	311.3	49.3	332.3	50.0	330.94
	15	53.5	238.5	49.9	197.0	49.9	197.0	49.7	201.7	50.0	204.1
	20	52.4	157.0	49.8	136.5	49.8	136.5	49.7	137.8	49.9	139.3
	25	51.9	122.9	49.8	110.1	49.8	110.1	49.7	110.6	49.8	111.7
	30	51.5	96.1	49.8	87.7	49.8	87.7	49.8	88.1	49.8	88.7

Tabela 3.6: Resultados da simulação para a distribuição gama, $k = 10$

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_j$		$\hat{\theta}_B^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
10	3	10.3	3.8	10.0	3.5	10.0	3.5	10.0	3.5	10.1	3.7
	4	10.2	2.7	10.0	2.5	10.0	2.7	10.0	2.7	10.0	2.6
	5	10.2	2.2	10.0	2.1	10.0	2.1	10.0	2.1	10.0	2.1
	10	10.1	1.0	10.0	1.0	10.0	1.0	10.0	1.0	10.0	1.0
	15	10.1	0.7	10.0	0.7	10.0	0.7	10.0	0.7	10.0	0.7
	20	10.1	0.5	10.0	0.5	10.0	0.5	10.0	0.5	10.0	0.5
50	25	10.0	0.4	10.0	0.4	10.0	0.4	10.0	0.4	10.0	0.4
	30	10.0	0.3	10.0	0.3	10.0	0.3	10.0	0.3	10.0	0.3
	3	51.5	96.1	49.8	87.7	49.8	87.7	49.7	91.9	50.3	91.4
	4	51.2	68.0	49.9	63.4	49.9	63.4	49.8	64.7	50.2	65.3
	5	50.9	54.2	49.9	51.4	49.9	51.4	49.8	52.0	50.0	52.2
	10	50.5	25.8	50.0	25.0	50.0	25.0	50.0	25.1	50.1	25.3
100	15	50.4	17.4	50.1	17.0	50.1	17.0	50.1	17.1	50.1	17.1
	20	50.3	13.0	50.0	12.8	50.0	12.8	50.0	12.8	50.0	12.8
	25	50.2	10.2	50.0	10.0	50.0	10.0	50.0	10.0	50.0	10.1
	30	50.2	8.4	50.0	8.4	50.0	8.4	50.0	8.4	50.0	8.4

As Tabelas 3.4 a 3.6 mostram que o E.M.V. $\hat{\theta}$ pode apresentar um vício bastante grande, especialmente se k e n são pequenos. Como pode ser observado na Tabela 3.4, quando $k = 0.1$, $\hat{\theta}$ é um estimador altamente viciado para valores de n iguais a 5 e 10. Note que nestes casos, a correção do vício não produziu resultados satisfatórios. Percebe-se também que, para tamanhos de amostra moderados, o vício deste estimador ainda apresenta uma ordem de grandeza considerável. Entretanto, os estimadores corrigidos por vício apresentam resultados bem satisfatórios, como pode ser observado para tamanhos de amostras maiores e iguais a 15. No caso especial onde $k = 1$ (distribuição exponencial), pode ser observado que $\hat{\theta}$ ainda é muito viciado. Note que o vício de $\hat{\theta}$ torna-se “desprezível” com o aumento de k (ver Tabela 3.6).

É importante destacar que para a distribuição gama, o vício de ordem n^{-2} é igual a zero e, portanto, $\hat{\theta}_3$ é igual ao estimador $\hat{\theta}_1$. Também, neste caso, os estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ obtidos como solução de (2.22) e (2.23) são iguais ao estimador corrigido por vício $\hat{\theta}_1$. Quanto aos estimadores corrigidos por vício através de métodos de reamostragem, $\hat{\theta}_J^1$ apresenta um desempenho melhor que $\hat{\theta}_B^1$ para valores pequenos de n , entretanto, com E.Q.M.’s maiores. Está claro também, que os estimadores corrigidos por vício através de métodos analíticos funcionam muito melhor neste caso que os obtidos por reamostragem.

3.4 A Distribuição de Rayleigh

A distribuição de Rayleigh foi derivada, originalmente, por Lord Rayleigh num problema em acústica e tem sido muito utilizada no campo da engenharia elétrica. A geração de números aleatórios com distribuição de Rayleigh foi realizada utilizando o método de inversão da função distribuição (Devroye, 1986; pp. 27-29). A Tabela 3.7 apresenta os resultados da simulação para a distribuição de Rayleigh considerando valores de θ iguais a 10 e 50. Como pode ser observado, $\hat{\theta}$ apresenta vício pequeno e os estimadores corrigidos apresentam vício desprezível mesmo para n bem pequeno. Perceba-se que os estimadores corrigidos por vício são bem eficientes neste caso, entretanto, eles apresentam maiores E.Q.M.’s que o E.M.V. $\hat{\theta}$. Entre os estimadores analíticos, $\hat{\theta}_2$ é o menos viciado e com menor E.Q.M. até ordem n^{-1} . Note também que o estimador corrigido por vício através do método Bootstrap, $\hat{\theta}_B^1$, apresenta um menor E.Q.M. se comparado com os estimadores corrigidos por vício até ordem n^{-1} , para n pequeno.

Tabela 3.7: Resultados da simulação para a distribuição de Rayleigh.

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_3$		$\bar{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_j^1$		$\hat{\theta}_R^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
10	3	9.6	8.0	10.0	8.5	10.0	8.5	9.6	8.0	10.0	8.6	10.0	8.5	10.0	8.7	9.9	8.4
	4	9.7	6.1	10.0	6.4	10.0	6.4	9.7	6.1	10.0	6.4	10.0	6.4	10.0	6.5	9.9	6.3
	5	9.7	4.9	10.0	5.1	10.0	5.1	9.7	4.9	10.0	5.1	10.0	5.1	10.0	5.1	9.9	5.1
	6	9.8	4.0	10.0	4.1	10.0	4.1	9.8	4.0	10.0	4.1	10.0	4.1	10.0	4.1	9.9	4.1
	7	9.8	3.5	10.0	3.6	10.0	3.6	9.8	3.5	10.0	3.6	10.0	3.6	10.0	3.6	9.9	3.6
	8	9.8	3.1	10.0	3.2	10.0	3.2	9.8	3.1	10.0	3.2	10.0	3.2	10.0	3.2	9.9	3.6
	10	9.9	2.5	10.0	2.5	10.0	2.5	9.9	2.5	10.0	2.5	10.0	2.5	10.0	3.2	10.0	3.2
	15	9.9	1.7	10.0	1.7	10.0	1.7	9.9	1.7	10.0	1.7	10.0	1.7	10.0	1.7	10.0	1.7
	20	9.9	1.2	10.0	1.2	10.0	1.2	9.9	1.1	10.0	1.2	10.0	1.2	10.0	1.2	10.0	1.2
50	3	48.1	200.3	50.1	213.2	50.1	213.6	48.1	200.3	50.2	214.4	49.9	212.2	50.0	216.3	49.2	208.8
	4	48.4	153.4	49.9	160.4	50.0	160.6	48.4	153.4	50.0	160.9	49.9	160.0	49.9	161.6	49.4	158.4
	5	48.7	123.2	49.9	127.6	49.9	127.7	48.7	123.2	50.0	127.9	49.9	127.4	49.9	128.5	49.6	127.0
	6	48.9	99.3	49.9	102.3	49.9	102.3	48.9	99.3	50.0	102.4	49.9	102.2	49.9	102.8	49.7	102.3
	7	49.0	87.6	49.9	89.8	49.9	89.8	49.0	87.6	49.9	89.9	49.8	89.7	49.8	90.2	49.7	89.6
	8	49.1	77.4	49.9	79.1	49.9	79.1	49.1	77.4	49.9	79.1	49.9	79.0	49.9	79.3	49.7	79.2
	10	49.3	62.0	49.9	63.0	49.9	63.0	49.3	62.0	49.9	63.1	49.9	63.0	49.9	63.2	49.8	63.1
	15	49.5	41.3	49.9	41.8	49.9	41.8	49.5	41.3	49.9	41.8	49.9	41.8	49.9	41.8	49.9	41.9
	20	49.6	30.1	49.9	30.3	49.9	30.3	49.6	30.1	49.9	30.3	49.9	30.3	49.9	30.3	49.9	30.4

3.5 A Distribuição de Pareto

A distribuição de Pareto foi nomeada devido a um professor suíço de economia, nascido na Itália, Vilfredo Pareto (1848-1923). As Leis de Pareto, como formuladas por ele, tratam da distribuição de renda de uma população. Muitas quantidades sócio-econômicas e outras naturais são distribuídas de acordo com certas distribuições estatísticas com caudas à direita muito pesadas. Exemplos de alguns desses fenômenos empíricos são distribuições do tamanho da população de cidades, ocorrência de recursos naturais, tamanho de empresas, renda pessoal e outros. A distribuição de Pareto tem representado a maior parte dessas investigações. Sua função distribuição acumulada é frequentemente usada para representar a chance da renda de um indivíduo ser limitada por y , quando a renda mínima possível é k .

A geração de números aleatórios com distribuição Pareto foi executada utilizando o método de inversão da função distribuição (Devroye, 1986; pp. 27-29). A Tabela 3.8 apresenta os resultados da simulação para a distribuição de Pareto considerando valores do parâmetro θ iguais a 5, 10, 25 e 50, quando $k = 1$. Como se pode observar, o E.M.V. $\hat{\theta}$ apresenta um vício muito grande. Como no caso da distribuição gama, o termo $B_2^*(\theta)$ é igual a zero e, portanto, $\hat{\theta}_3$ é igual ao estimador $\hat{\theta}_1$. Note mais uma vez que os estimadores corrigidos por vício até ordem n^{-1} são muito pouco viciados e apresentam E.Q.M. menor que o estimador $\hat{\theta}$. Os estimadores corrigidos por vício através de métodos de reamostragem não produziram boas correções para valores pequenos de n como, por exemplo, n igual a 5. Note também que $\tilde{\theta}_J^1$ é menos viciado que o estimador corrigido por vício $\tilde{\theta}_B^1$ pelo método de Bootstrap.

Tabela 3.8: Resultados da Simulação para a Distribuição de Pareto, $k = 1$.

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_J^1$		$\hat{\theta}_B^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
5	3	7.6	111.9	5.1	46.7	5.1	46.7	5.1	46.7	2.9	148.3	0.5	6519.7
	4	7.6	25.5	5.0	12.7	5.0	12.7	5.0	12.7	4.1	40.3	4.0	162.8
	5	6.3	14.5	5.0	8.2	5.0	8.2	5.0	8.2	4.6	13.5	4.8	11.9
	10	5.6	4.2	5.0	3.1	5.0	3.1	5.0	3.1	4.9	3.3	5.0	3.3
	15	5.4	2.4	5.0	2.0	5.0	2.0	5.0	2.0	5.0	2.0	5.0	2.0
	20	5.2	1.6	5.0	1.4	5.0	1.4	5.0	1.4	5.0	1.4	5.0	1.4
10	25	5.2	1.2	5.0	1.1	5.0	1.1	5.0	1.1	5.0	1.1	5.0	1.1
	30	5.2	1.0	5.0	0.9	5.0	0.9	5.0	0.9	5.0	0.9	5.0	0.9
	3	15.2	447.4	10.1	186.8	10.1	186.8	10.1	186.8	5.8	593.3	1.1	26078.9
	4	13.4	101.7	10.1	50.7	10.1	50.7	10.1	50.7	8.3	161.3	8.1	651.1
	5	12.5	58.0	10.0	33.0	10.0	33.0	10.0	33.0	9.2	54.1	9.7	47.4
	10	11.1	16.6	10.0	12.5	10.0	12.5	10.0	12.5	9.9	13.3	10.0	13.2
25	15	10.7	9.5	10.0	7.9	10.0	7.9	10.0	7.9	9.9	8.1	10.0	8.2
	20	10.5	6.3	10.0	5.5	10.0	5.5	10.0	5.5	9.9	5.5	10.0	5.6
	25	10.4	4.9	10.0	4.4	10.0	4.4	10.0	4.4	9.9	4.4	10.0	4.5
	30	10.3	3.8	10.0	3.5	10.0	3.5	10.0	3.5	10.0	3.5	10.0	3.6
	3	38.0	2796.1	25.3	1167.7	25.3	1168.0	25.3	1168.0	14.6	3708.2	2.7	162993.5
	4	33.5	635.8	25.2	316.7	25.2	316.7	25.2	316.7	20.7	1008.0	20.1	4069.3
50	5	31.4	362.3	25.1	206.1	25.1	206.1	25.1	206.1	23.1	337.9	24.2	296.4
	10	27.8	103.6	25.0	77.8	25.0	77.8	25.0	77.8	24.7	83.1	25.0	82.7
	15	26.8	59.6	25.0	49.3	25.0	49.3	25.0	49.3	24.8	50.4	25.0	51.0
	20	26.2	39.3	24.9	34.1	24.9	34.1	24.9	34.1	24.8	34.5	24.9	34.8
	25	25.9	30.7	24.9	27.5	24.9	27.5	24.9	27.5	24.8	27.7	24.9	27.9
	30	25.8	24.0	24.9	21.9	24.9	21.9	24.9	21.9	24.9	22.0	24.9	22.2
50	3	76.0	11184.6	50.7	4670.9	50.7	4670.9	50.7	4670.9	29.2	14832.8	5.4	651974.0
	4	67.1	2543.2	50.3	1266.8	50.3	1266.8	50.3	1266.8	41.3	4031.9	40.3	16277.0
	5	62.7	1449.2	50.2	824.2	50.2	824.2	50.2	824.2	46.1	1351.5	48.3	1185.7
	10	55.5	414.6	50.0	311.3	50.0	311.3	50.0	311.3	49.3	332.3	50.0	330.94
	15	53.5	238.5	49.9	197.0	49.9	197.0	49.9	197.0	49.7	201.7	50.0	204.1
	20	52.4	157.0	49.8	136.5	49.8	136.5	49.8	136.5	49.7	137.8	49.9	139.3
50	25	51.9	122.9	49.8	110.1	49.8	110.1	49.8	110.1	49.7	110.6	49.8	111.7
	30	51.5	96.1	49.8	87.7	49.8	87.7	49.8	87.7	49.8	88.1	49.8	88.7

3.6 A Distribuição Lognormal

A distribuição lognormal, com parâmetros de escala μ e de forma θ , pode ser definida como a distribuição de uma variável aleatória cujo logaritmo é normalmente distribuído. Tal variável é necessariamente positiva. Visto que muitas variáveis na vida real, como tamanhos de organismos e o número de espécies em biologia, pluviosidade na meteorologia e tamanho de rendas em economia, são inerentemente positivas, a distribuição lognormal tem sido amplamente aplicada de uma forma empírica para ajuste desses dados. Johnson e Kotz (1970) e Shimizu e Crow (1988) descrevem uma breve história da distribuição lognormal e apresentam várias áreas de aplicação.

Em aplicações, a distribuição lognormal é usada com processos pontuais. Neste caso, por exemplo, depois do início de uma doença, a distribuição lognormal pode representar o período de tempo decorrido até que um paciente se recupere ou morra, ou o tempo de sobrevivência de pacientes de câncer, ou ainda, o período de incubação de uma doença infecciosa. Ainda, em relação a processos pontuais, o tempo de chamadas telefônicas pode ser distribuído lognormalmente. Existem indicações de que a distribuição lognormal representa numerosos aspectos da vida econômica. O resultado aceito mais facilmente na distribuição de renda é paradoxal: a lognormal aproxima a renda na amplitude média mas falha na cauda superior, onde a distribuição de Pareto é mais apropriada. Na indústria, a distribuição lognormal tem sido usada amplamente como um modelo da distribuição do tempo de vida. A distribuição é especificada dizendo que o tempo de vida de X é lognormal se o logaritmo, $Y = \log X$, do tempo de vida é normal com média μ e variância σ^2 . Distribuições lognormais tem sido usadas, extensivamente, em ciências atmosféricas para descrever variáveis que tomam valores não negativos, tais como, tempos de chuva de longo período, diárias, de tempestades, quantidade de neve e granito, tamanho de partículas (aerosol), concentração de poluentes, flutuações da velocidade do ar, etc. A Hidrologia é uma usuária de grande escala de métodos estatísticos, incluindo estimação e modelagem com distribuição lognormal, particularmente, para prever precipitação, frequência de enchentes, fornecimento de água, etc. Na Geologia, a distribuição lognormal tem sido amplamente reconhecida como um modelo útil no ajuste de dados geológicos com distribuição extremamente assimétricas. O conteúdo de metal em um pedaço de rocha de um depósito mineral, a dimensão de partículas de areia e cascalho na base de um rio e o tamanho da jazida de óleo ou gás em uma região geográfica são exemplos de variáveis geológicas que tem sido modelados usando a distribuição lognormal. A distribuição

também é usada para modelar a distribuição do tamanho de depósitos minerais em grande áreas.

A geração de números aleatórios com distribuição lognormal foi realizada utilizando o gerador de números aleatórios do software S-Plus, 1993. A Tabela 3.9 apresenta os resultados da simulação para a distribuição lognormal considerando valores do parâmetro de dispersão θ iguais a 10 e 50, quando o parâmetro de locação μ é igual a 10.

Como se pode observar, $\hat{\theta}$ é um estimador muito viciado à medida que cresce o valor do parâmetro θ . Em relação aos estimadores corrigidos por vício até ordem n^{-1} , todos são praticamente não viciados, entretanto, com E.Q.M.'s maiores que o E.M.V., $\hat{\theta}$. Dos estimadores corrigidos por vício até ordem n^{-1} , $\tilde{\theta}_2$ é o que apresenta menor E.Q.M. para valores pequenos de n . Já entre os estimadores corrigidos por vício por reamostragem, $\tilde{\theta}_B^1$ apresenta E.Q.M.'s menores que os estimadores corrigidos por processos analíticos, entretanto, estes últimos corrigem melhor o vício do estimador $\hat{\theta}$ do que o estimador $\tilde{\theta}_B^1$. Dentre os estimadores corrigidos por vício até ordem n^{-1} , o estimador Jackknife $\tilde{\theta}_J^1$ foi o que teve pior desempenho.

Tabela 3.9: Resultados da Simulação para a Distribuição lognormal, $\mu = 10$.

θ	n	$\hat{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_3$		$\bar{\theta}$		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_1^1$		$\hat{\theta}_R^1$	
		Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM	Média	EQM
10	3	9.2	15.7	10.0	17.7	10.0	17.8	9.2	15.7	10.1	18.1	9.9	17.4	9.9	18.1	9.6	16.8
	4	9.3	12.2	9.9	13.3	9.9	13.3	9.3	12.2	10.0	13.5	9.9	13.2	9.9	13.6	9.7	12.9
	5	9.5	9.7	9.9	10.4	9.9	10.4	9.5	9.7	10.0	10.5	9.9	10.3	9.9	10.6	9.8	10.3
	6	9.5	8.2	9.9	8.7	9.9	8.7	9.5	8.2	10.0	8.7	9.9	8.7	9.9	8.8	9.8	8.6
	7	9.6	7.0	9.9	7.4	9.9	7.4	9.6	7.0	9.9	7.4	9.9	7.3	9.9	7.4	9.8	7.3
	8	9.6	6.3	9.9	6.6	9.9	6.6	9.6	6.3	9.9	6.6	9.9	6.5	9.9	6.6	9.9	6.5
	9	9.7	5.6	9.9	5.8	9.9	5.8	9.7	5.6	10.0	5.8	9.9	5.8	9.9	5.8	9.9	5.8
	10	9.7	5.0	10.0	5.2	10.0	5.2	9.7	5.0	10.0	5.2	10.0	5.2	10.0	5.2	9.9	5.2
	11	9.7	4.7	10.0	4.9	10.0	4.9	9.7	4.7	10.0	4.9	10.0	4.9	10.0	4.9	9.9	4.9
	12	9.8	4.2	10.0	4.3	10.0	4.3	9.8	4.2	10.0	4.3	10.0	4.3	10.0	4.3	10.0	4.3
	15	9.8	3.5	10.0	3.5	10.0	3.5	9.8	3.5	10.0	3.5	10.0	3.5	10.0	3.6	10.0	3.5
50	3	45.9	393.0	49.8	441.9	49.9	444.6	45.9	393.0	50.3	451.8	49.3	434.8	49.5	451.8	48.0	421.1
	4	46.7	305.1	49.6	332.4	49.7	333.5	46.7	305.1	49.9	336.3	49.4	329.3	49.5	339.7	48.5	323.5
	5	47.3	243.3	49.7	260.2	49.7	260.8	47.3	243.3	49.8	262.1	49.5	258.6	49.6	265.6	48.9	256.9
	6	47.6	205.9	49.6	217.5	49.7	217.9	47.6	205.9	49.8	218.6	49.5	216.6	49.6	220.1	49.1	214.5
	7	47.9	175.9	49.6	184.0	49.6	184.2	47.9	175.9	49.7	184.7	49.5	183.4	49.6	185.9	49.2	182.7
	8	48.1	157.6	49.6	164.0	49.7	164.1	48.1	157.6	49.7	164.4	49.6	163.6	49.6	165.4	49.3	163.2
	9	48.4	140.0	49.7	145.1	49.7	145.2	48.4	140.0	49.8	145.4	49.6	144.8	49.7	145.5	49.4	144.2
	10	48.6	125.6	49.8	129.9	49.8	130.0	48.6	125.6	49.9	130.2	49.8	129.7	49.8	130.4	49.6	129.4
	11	48.7	118.6	49.8	122.4	49.8	122.4	48.7	118.6	49.9	122.5	49.8	122.2	49.8	122.7	49.6	122.0
	12	48.9	105.0	49.9	108.1	49.9	108.2	48.9	105.0	49.9	108.3	49.9	108.0	49.9	108.5	49.7	108.2
	15	49.1	86.4	49.9	88.4	49.9	88.4	49.1	86.4	49.9	88.5	49.9	88.3	49.9	88.7	49.8	88.1

Conclusão

Os resultados apresentados neste trabalho ilustram a importância prática da utilização de correções de vícios de estimadores de máxima verossimilhança. Um dos enfoques principais deste trabalho foi obter, para alguns modelos especiais, através das metodologias sugeridas por Ferrari e outros (1996) e Firth (1993), correções analíticas para o vício até ordens n^{-1} e n^{-2} dos E.M.V.'s. Também, alguns resultados de simulação foram apresentados, comparando estimadores corrigidos por vício obtidos através destes métodos analíticos e por métodos de reamostragem. Como pode ser observado, a metodologia de correção de vícios de estimadores é, de uma forma geral, de fácil utilização, podendo ser implementada sem grandes dificuldades. Entretanto, é importante ressaltar que o uso de tais técnicas deve ser acompanhado com bastante cuidado, uma vez que, sua utilização nem sempre possa ser conveniente.

Nos exemplos apresentados neste trabalho, os estimadores corrigidos por vício através de métodos analíticos superaram os obtidos por processos de reamostragem (Jackknife e Bootstrap). Os estimadores analíticos tiveram praticamente o mesmo desempenho, sendo que em alguns casos (binomial negativa, Rayleigh, lognormal) o estimador $\tilde{\theta}_2$, obtido através da metodologia sugerida por Firth, apresentou uma ligeira superioridade, para valores pequenos de n . Com relação aos métodos de reamostragem, o estimador corrigido por vício $\tilde{\theta}_J^1$ obtido pelo processo Jackknife mostrou ser superior, para valores pequenos de n , ao estimador corrigido por vício $\tilde{\theta}_B^1$ obtido através do processo Bootstrap.

Finalizando, vários pontos importantes podem ser sugeridos como futuras pesquisas. Primeiro, a obtenção de um estimador, baseado na metodologia sugerida por Firth (1993), que corrija o vício até ordem n^{-2} . É importante destacar que a metodologia sugerida por Firth proporciona fórmulas bastante complexas para algumas distribuições. Desta forma, os estimadores sugeridos por Ferrari e outros (1996) levam vantagem por serem obtidos

mais facilmente. Outro ponto importante é o estudo de estimadores corrigidos por vício até ordem n^{-2} , obtidos através dos processos Jackknife e Bootstrap. É fácil perceber que a teoria envolvida nestes procedimentos é bastante simples, no entanto, o estudo de simulação de tais métodos demanda um custo computacional ainda bastante elevado.

Apêndice 1

‘Este programa calcula as formulas para Beta(t), B1(t), B2(t), B2*(t), V1(t), V2(t),
‘Delta(t), gama(t) e delta(t), para a Distribuicao Binomial’;
‘t: Representa o parametro theta’;

alpha(theta):=-log(theta);

xi(theta):=1/(1-theta)^r;

beta(t):=simplify(diff(xi(t),t)/(xi(t)*diff(alpha(t),t)));

B1(t):=simplify(-diff(beta(t),t^2)/(2*diff(alpha(t),t)*(diff(beta(t),t))^2));

B2(t):=simplify((-12*diff(beta(t),t)*diff(alpha(t),t^2)*(diff(beta(t),t^2))^2
-33*diff(alpha(t),t)*diff(beta(t),t^2)^3
+4*diff(beta(t),t)^2*diff(alpha(t),t^2)*diff(beta(t),t^3)
+26*diff(alpha(t),t)*diff(beta(t),t)*diff(beta(t),t^2)
*diff(beta(t),t^3)-3*diff(alpha(t),t)*diff(beta(t),t)^2
*diff(beta(t),t^4))/(24*diff(alpha(t),t)^3*diff(beta(t),t)^5));

B2e(t):=simplify((12*diff(beta(t),t)^2*diff(alpha(t),t^2)^2*diff(beta(t),t^2)
+18*diff(alpha(t),t)*diff(beta(t),t)*diff(alpha(t),t^2)
*diff(beta(t),t^2)^2+15*diff(alpha(t),t)^2*diff(beta(t),t^2)^3
-6*diff(alpha(t),t)*diff(beta(t),t)^2*diff(beta(t),t^2)
*diff(alpha(t),t^3)-8*diff(alpha(t),t)*diff(beta(t),t)^2
*diff(alpha(t),t^2)*diff(beta(t),t^3)-16*diff(alpha(t),t)^2
*diff(beta(t),t)*diff(beta(t),t^2)*diff(beta(t),t^3)

```
+3*diff(alpha(t),t)^2*diff(beta(t),t)^2*diff(beta(t),t$4))/(24*
diff(alpha(t),t)^4*diff(beta(t),t)^5));
```

```
V1(t):=simplify(1/(diff(alpha(t),t)*diff(beta(t),t)));
```

```
V2(t):=simplify((2*diff(beta(t),t)*diff(alpha(t),t$2)*diff(beta(t),t$2)
+5*diff(alpha(t),t)*diff(beta(t),t$2)^2-2*diff(alpha(t),t)
*diff(beta(t),t)*diff(beta(t),t$3))/(2*diff(alpha(t),t)^3
*diff(beta(t),t)^4));
```

```
Delta(t):=simplify((4*diff(beta(t),t)*diff(alpha(t),t$2)*diff(beta(t),t$2)
+9*diff(alpha(t),t)*diff(beta(t),t$2)^2-4*diff(alpha(t),t)
*diff(beta(t),t)*diff(beta(t),t$3))/(4*diff(alpha(t),t)^3
*diff(beta(t),t)^4));
```

```
Gamma(t):=simplify((diff(beta(t),t$2))/(2*diff(beta(t),t)*diff(alpha(t),t)));
```

```
delta(t):=simplify((diff(alpha(t),t)*diff(beta(t),t)*diff(beta(t),t$2))/(2
*diff(alpha(t),t)^2*diff(beta(t),t)^2-(1/n)*diff(alpha(t),t$2)
*diff(beta(t),t$2)));
```

‘Calculo do E.M.V.’;

```
eq1:= beta(t)+d=0;
```

```
Solucao1:= solve(eq1,t);
```

‘Calculo do E.M.V. corrigido usando a informacao esperada (Firth, 1993)’;

```
eq2:= beta(t)+d-Gamma(t)/n=0;
```

```
Solucao2:= solve(eq2,t);
```

‘Calculo do E.M.V. corrigido usando a informacao observada (Firth, 1993)’;

```
eq3:= beta(t)+d-delta(t)/n=0;
```

```
Solucao3:= solve(eq3,t);
```

Apêndice 2

```
SIMU.AUX<- function(n=5,k=1,psucesso=0.1,thetatrue=0.9,r=5000)
{
#
# Esta funcao executa o procedimento de simulacao para o calculo dos EMV's,
# viciado e corrigidos por vicio, baseada em 5000 replicas.
# k e psucesso: valores dos parametros para geracao de numeros
# aleatorios com distribuicao Binomial negativa
# thetatrue: verdadeiro valor do parametro
#
temp<- matrix(0,r/10,110)
one<- matrix(1,n,1)
#
# Geracao de numeros aleatorios com distribuicao Binomial negativa
#
for(i in 1:10){
set.seed(123+11*i)
Y<- matrix(rnbinom(n*(r/10),k,psucesso),n,r/10)
#
# Calculo de EMV corrigidos por vicio atraves de metodos analiticos
#
# Estimadores obtidos por Ferrari e outros (1995)
#
temp[,i]<- (1/n)*(t(Y)%*%one)
```

```

temp[,i+10]<- ((1/n)*(t(Y)%*%one))/(k+(1/n)*(t(Y)%*%one))
temp[,i+20]<- temp[,i+10]+((temp[,i+10]*(1-temp[,i+10]))/(k*n))
temp[,i+30]<- temp[,i+10]+((temp[,i+10]*(1-temp[,i+10]))/(k*n))
+((temp[,i+10]*((1-temp[,i+10])^2)))/((k*n)^2))
#
# Calculo do estimador dado em Silva (1996), expressao (2.9)
#
temp[,i+40]<- (temp[,i+10]*(5*temp[,i+10]-2)*((1-temp[,i+10])^2))/(k*k)
for (j in 1:(r/10)){
  if (temp[j,i+40] <= 0){
    temp[j,i+50]<- temp[j,i+10]
  } else
  {
    temp[j,i+50]<- temp[j,i+30]
  }
temp[,i+60]<- cbind(temp[(1:(r/10)),i+50])
temp[,i+60]
}
#
# Estimadores obtidos por David Firth (1993)
#
temp[,i+70]<- temp[,i]/(k+temp[,i]-(1/n))
temp[,i+80]<- ((n*k*temp[,i]+2*temp[,i]+n*(k*k))
-(((n*n)*(k*k)*(temp[,i]^2)))+(2*(n*n)*(k^3)*temp[,i])
+((n*n)*(k^4))-(sqrt(4*k*temp[,i])))/(2*(temp[,i]+k))
#
# Calculo dos Estimadores corrigidos por processo de reamostragem
# Execucao da funcao jackknife
#
theta1.jack<- matrix(Jackknife1(Y,r=r,n=n,k=k,theta2=theta.mle),r/10,1)
temp[,i+90]<- temp[,i+10]-(n-1)*(theta1.jack-temp[,i+10])
#

```



```

# Execucao da funcao bootstrap
#
boot1.res<- apply(Y,2,bootstrap.n=n,k=k)
theta.star1<- matrix(apply(boot1.res,2,mean),r/10,1)
temp[,i+100]<- 2*temp[,i+10]-theta.star1
}
#
# Impressao dos resultados
#
mle<- rbind(temp[, (11:20)])
hat1<- rbind(temp[, (21:30)])
hat2<- rbind(temp[, (31:40)])
thetab<- rbind(temp[, (61:70)])
tilde1<- rbind(temp[, (71:80)])
tilde2<- rbind(temp[, (81:90)])
jack1<- rbind(temp[, (91:100)])
boot1<- rbind(temp[, (101:110)])
result<- c(mean(mle),mean((mle-thetatru)^2),mean(hat1),
            mean((hat1-thetatru)^2),mean(hat2),mean((hat2-thetatru)^2),
            mean(thetab),mean((thetab-thetatru)^2),mean(tilde1),
            mean((tilde1-thetatru)^2),mean(tilde2),mean((tilde2-thetatru)^2),
            mean(jack1),mean((jack1-thetatru)^2),
            mean(boot1),mean((boot1-thetatru)^2))
result<- matrix(result,1,length(result),dimnames=list(n,c("MLE","EQM","HAT1",
"EQM","HAT2","EQM","TILDE","EQM","TILDE1","EQM",
"TILDE2","EQM","JACK1","EQM","BOOT1","EQM")))
result
round(result,2)
}

bootstrap<- function(x,nboot=200,n,k,theta=theta.mle,...)
{

```

```

#
# Esta funcao executa o procedimento Bootstrap nao-parametrico com 1 iteracao.
# A execucao desta funcao so e possivel para tamanhos de amostra n maiores que 1.
# Para n=1, a estimativa bootstrap e igual a estimativa de maxima verossimilhanca.
#
data<- matrix(sample(x,size=length(x)*nboot,replace=T),nrow=nboot)
res<- apply(data,1,theta,n,k)
res
}

Jackknife1<- function(x,r,n,k,theta=theta.mle,...)
{
# Esta funcao executa o procedimento Jackknife. A execucao desta funcao
# so e possivel para tamanhos de amostra n maiores que 1.
#
jack<- matrix(0,n,r/10)
for(i in 1:n){
dados<- x[-i,]
jack[i,]<- matrix(apply(dados,2,theta,n-1,k),n-1,r/10)
}
thetajack<- rbind(jack[(1:n),])
thetajack1<- apply(thetajack,2,mean)
thetajack1
}

theta.mle<- function(x,n,k=1)
{
# Esta funcao calcula o EMV para ser usado nas funcoes Bootstrap e Jackknife
#
emle<- ((1/n)*sum(x))/(k+((1/n)*sum(x)))
emle
}

```

Referências Bibliográficas

- ABELL, M. L.; BRASELTON, J. P. (1994). The Maple Handbook. New York: AP Professional.
- ABRAMOWITZ, M.; STEGUN, I. R. (1976). Handbook of Mathematical Functions. Washington, D. C.: National Bureau of Standards.
- ADAMS, J. E.; GRAY, H. L.; WATKINS, T. A. (1971). An asymptotic characterization of bias reduction by jackknifing. *The Annals of Mathematical Statistics*, 42, 5, 1606-1612.
- ANDERSON, J. A.; RICHARDSON, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, 21, 1, 71-78.
- AUBIN, E. C. Q. (1995). Aperfeiçoamento de Testes da Razão de Verossimilhança em Modelos Especiais com Algumas Aplicações e Extensões. São Paulo, 232p. Tese (Doutorado) - Instituto de Matemática e Estatística da Universidade de São Paulo.
- BARTLETT, M. S. (1953). Approximate confidence intervals II. *Biometrika*, 40, 306-317.
- BEAN, S. J.; TSOKOS, C. P. (1980). Developments in nonparametric density estimation. *Institute Statistical Review*, 48, 267-287.
- BOX, M. J. (1971). Bias in nonlinear estimation. *Journal of the Royal Statistical Society, B*, 32, 171-201.
- BURR, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, 13, 215-232.
- COOK, R. D.; TSAI, C. L.; WEI, B. C. (1986). Bias in nonlinear regression. *Biometrika*, 73, 615-623.
- CORDEIRO, G. M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, A*, 22(1), 169-188.

- CORDEIRO, G. M.; McCULLAGH, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society*, B, 53, 3. 629-643.
- COX, D. R.; SNELL, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society*, B, 30, 248-275.
- CRIBARI-NETO, F.; BOTTER, D. A.; CORDEIRO, G. M.; FERRARI, S. L. P. (1995). Bias reduction in one-parameter exponential family models. Relatório Técnico, RTMAE nº 9532. Instituto de Matemática e Estatística da Universidade de São Paulo.
- CROW, E. L.; SHIMIZU, K. (1988). Lognormal Distributions - Theory and Applications. Statistics, textbooks and monographs, vol. 88.
- DEVROYE, L. (1986). Non-Uniform Random Variate Generation. New York: Springer-Verlag.
- DOSS, H.; SETHURAMAN, J. (1989). The price of bias reduction when there is no unbiased estimate. *The Annals of Statistics*, 17, 1, 440-442.
- EFRON, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF. Regional Conference Series in Applied Mathematics, Philadelphia.
- EFRON, B.; TIBSHIRANI, R. (1993). An Introduction to the Bootstrap. Chapman and Hall, London.
- FERRARI, S. L. P.; CRIBARI-NETO, F. (1996). On bootstrap and analytical bias corrections. Relatório Técnico, RTMAE nº 9621. Instituto de Matemática e Estatística da Universidade de São Paulo.
- 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*. A aparecer.
- FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.
- HALL, P. (1992). The Bootstrap and Edgeworth Expansion. Springer-Verlag, New-York.
- JOHNSON, N. L.; KOTZ, S. (1970a). Discrete Univariate Distributions. Boston, Houghton Mifflin.
- JOHNSON, N. L.; KOTZ, S. (1970b). Continuous Univariate Distributions, V. 1. Boston, Houghton Mifflin.
- JOHNSON, N. L.; KOTZ, S. (1970c). Continuous Univariate Distributions, V. 2. Boston, Houghton Mifflin.

- JOHNSON, N. L; KOTZ, S. (1992). Discrete Univariate Distributions. 2nd ed. New York: Wiley.
- LAWLEY, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, 43, 295-303.
- MARDIA, K. V. (1972). Statistics of Directional Data. London: Academic Press.
- McCULLAGH, P. (1987). Tensor Methods in Statistics. London. Chapman and Hall.
- McCULLAGH, P. (1989). Some statistical properties of a family of continuous univariate distributions. *Journal of the American Statistical Association*, 84, 125-129.
- McLACHLAN, G. J. (1980). A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics*, 22, 4, 621-627.
- MILLER, R. G. (1964). A trustworthy Jackknife. *Annals of Mathematical Statistics*, 35, 1594-1605.
- MILLER, R. G. (1974). The jackknife - a review. *Biometrika*, 61, 1-15.
- PATTERSON, S. J. (1988). An Introduction to the Theory of the Riemann Zeta-function. New York, Cambridge University Press.
- PAULA, G. A. (1991). Bias correction for exponential family nonlinear models. *Journal of Statistical Computation and Simulation*, 40, 43-54.
- PICARD, D.; PRUM, B. (1992). The bias of the EMV, an example of the behaviour of different corrections in genetic models. *Statistics*, 23, 159-169.
- QUENOUILLE, M. H. (1949). Approximate tests, of correlation in time-Series. *Journal of the Royal Statistical Society*, B, 25, 305-317.
- QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.
- RAO, J. N. K. (1965). A note on estimation of ratios by Quenouille's method. *Biometrika*, 52, 647-649.
- RAO, J. N. K.; WEBSTER, J. T. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika*, 53, 571-577.
- SCHABE, H. (1990). Bias reduction of maximum likelihood estimators using kernel estimators. *Statistics*, 12, 4, 577-590.

- SCHUCANY, W. R.; GRAY, H. L.; OWEN, O. B. (1971). On bias reduction in estimation. *Journal of the American Statistical Association*, 66, 524-533.
- SHENTON, L. R.; BOWMAN, K. (1963). Higher moments of a maximum likelihood estimate. *Journal of the Royal Statistical Society, B*, 25, 305-317.
- SHENTON, L. R.; BOWMAN, K. (1977). Maximum likelihood estimation in small samples. London, Griffin.
- SHAPIRO, S. S.; ZAHEDI, H. (1990). Bernoulli trials and discrete distributions. *Journal of Quality Technology*, 22, 3.
- SHENTON, L. R.; WALLINGTON, P. A. (1962). The bias of moments estimators with an application to the negative binomial distribution. *Biometrika*, 49, 193-204.
- SHIMIZU, K.; CROW, E. (1988). Lognormal Distributions: Theory and Applications. Statistics, textbooks and monographs; vol. 88. Marcel Dekker, Inc. New York and Bassel.
- S-PLUS Guide to Statistical and Mathematical Analysis. (1993). Version 3.2. StatSci. A Division of MathSoft, Inc. Seattle, Washington.
- WITHERS, C. S. (1987). Bias reduction by Taylor series. *Communications in Statistics, A*, 16(8), 2369-2383.
- YOUNG, D. H.; BAKIR, S. T. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, 29, 2, 183-191.
- YOUNG, G. A.; DANIELS, H. E. (1990). Bootstrap bias. *Biometrika*, 77, 1, 179-185.