

**Aprimoramento de métodos
assintóticos em classes
de modelos estatísticos**

Francisco Moisés Cândido de Medeiros

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

Área de concentração: Estatística

Orientador: Prof. Dra. Silvia Lopes de Paula Ferrari

São Paulo, outubro de 2016

**Aprimoramento de métodos
assintóticos em classes
de modelos estatísticos**

Esta é a versão original da tese elaborada pelo
candidato Francisco Moisés Cândido de Medeiros, tal como
submetida à Comissão Julgadora.

Aprimoramento de métodos assintóticos em classes de modelos estatísticos

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 28/11/2016. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof^a. Dr^a. Silvia Lopes de Paula Ferrari (orientadora) - IME-USP
- Prof^a. Dr^a. Denise Aparecida Botter - IME-USP
- Prof. Dr. Fábio Mariano Bayer - UFSM
- Prof. Dr. Miguel Angel Uribe Opazo - UNIOESTE
- Prof. Dr. Artur José Lemonte - UFRN

Resumo

Estudamos testes de hipóteses em modelos de regressão lineares simétricos, log-simétricos e em modelos de dispersão quando o tamanho da amostra é pequeno ou moderado. Focamos o estudo nos testes de Wald, razão de verossimilhanças, escore e gradiente. Esses testes usam aproximações assintóticas e não são confiáveis quando o tamanho da amostra não é grande o suficiente para garantir uma boa concordância entre a distribuição exata da estatística de teste e a correspondente distribuição qui-quadrado de referência. As correções de Bartlett e tipo-Bartlett tipicamente atenuam a distorção de tamanho dos testes. As correções estão disponíveis na literatura para as estatísticas da razão de verossimilhanças e escore em modelos de regressão lineares simétricos e em modelos de dispersão. Derivamos um fator de correção tipo-Bartlett para a estatística gradiente nessas classes de modelos e verificamos que o fator de correção obtido para os modelos de regressão lineares simétricos é válido para os modelos de regressão lineares log-simétricos. Comparamos numericamente a probabilidade de erro tipo I do teste gradiente corrigido com as de outros tests propostos na literatura e de testes corrigidos via bootstrap. Nossos resultados sugerem que os testes corrigidos, incluindo o teste gradiente corrigido proposto nesta tese, apresentam probabilidades de erro tipo I mais próximas do nível nominal sem perda de poder.

Adicionalmente, apresentamos um estudo de viés dos estimadores de máxima verossimilhança em distribuições log-simétricas. Obtivemos ajustes que reduzem o viés e o viés mediano desses estimadores.

Apresentamos aplicações a dados reais para ilustrar a utilidade de nossos resultados.

Palavras-chave: Bootstrap, correção tipo-Bartlett, estatística da razão de verossimilhanças, estatística escore, estatística gradiente, estatística de Wald, estimadores modificados, redução de viés, redução de viés mediano.

Abstract

This thesis deals with the issue of testing hypotheses in symmetric and log-symmetric linear regression models and dispersion models in small and moderate-sized samples. We focus on four tests, namely the Wald, likelihood ratio, score, and gradient tests. These tests rely on asymptotic results and are unreliable when the sample size is not large enough to guarantee a good agreement between the exact distribution of the test statistic and the corresponding chi-squared asymptotic distribution. Bartlett and Bartlett-type corrections typically attenuate the size distortion of the tests. These corrections are available in the literature for the likelihood ratio and score tests in symmetric linear regression models and dispersion models. Here, we derive a Bartlett-type correction for the gradient test in these classes of models. We show that the correction factor obtained for symmetric linear regression models is also valid for the class of log-symmetric linear regression models. We numerically compare the type I error probability of the corrected gradient test with the type I error probability of other tests proposed in the literature and bootstrapped tests. Our results suggest that the corrected tests, including the corrected gradient test proposed in this thesis, exhibit type I error probability closer to the chosen nominal level with virtually no power loss.

In addition, we present a study of the bias of maximum likelihood estimators in log-symmetric distributions. We obtain adjustments that reduce the bias and the median bias of these estimators.

We present applications to real data to illustrate the usefulness of our results.

Keywords: Bartlett-type correction, bias reduction, bootstrap, gradient statistic, likelihood ratio statistic, median bias reduction, modified estimators, score statistic, Wald statistic.

Sumário

Resumo	i
Abstract	ii
Lista de Figuras	v
Lista de Tabelas	vi
1 Introdução	1
2 Aprimoramento de testes de hipóteses em modelos de regressão lineares simétricos e log-simétricos	4
2.1 Introdução	5
2.2 Modelos de regressão lineares simétricos, estimação e testes	8
2.3 Testes de hipóteses aperfeiçoados	12
2.4 Modelos de regressão lineares log-simétricos e testes aperfeiçoados	16
2.5 Estudo de simulação	18
2.6 Aplicações	28
2.7 Considerações finais	33
2.8 Detalhes técnicos	34
2.8.1 Prova da Proposição 1	34
2.8.2 Cálculo dos δ 's	39
3 Aprimoramento de testes de hipóteses em modelos de dispersão	44
3.1 Introdução	45
3.2 Modelos de dispersão, estimação e testes	47

SUMÁRIO

3.3	Estatísticas de teste aprimoradas	51
3.4	Estudo de simulação	56
3.5	Aplicações	65
3.6	Considerações finais	68
3.7	Detalhes técnicos: Prova da Proposição 2	68
4	Diferentes métodos de redução de viés de estimadores em distribuições log-simétricas	74
4.1	Introdução	74
4.2	A classe de distribuições log-simétricas	76
4.3	Métodos de redução de viés	79
4.4	Estudo de simulação	84
4.5	Aplicação	89
4.6	Considerações finais	91
4.7	Detalhes técnicos	91
5	Considerações finais e pesquisas futuras	94
	Referências Bibliográficas	96

Lista de Figuras

2.1	Poder dos testes para $n = 30$, $p = 4$, $q = 3$, $\phi = 3$ e $\alpha = 10\%$	27
2.2	Gráfico normal de probabilidades e envelope gerado para as distribuições (a) log-normal, (b) log-t-Student e (c) log-logística II ajustadas ao modelo (2.10).	29
2.3	Gráfico normal de probabilidades e envelope simulado para o modelo (2.11).	30
2.4	Gráfico normal de probabilidades e envelope simulado para as distribuições (a) normal, (b) t-Student e (c) logística II ajustadas ao modelo (2.12).	31
2.5	Gráfico normal de probabilidades e envelope simulado para o modelo (2.14).	33
3.1	Poder dos testes para $n = 30$, $p = 4$, $q = 3$ e $\alpha = 10\%$; von Mises (esquerda) e log-gama (direita).	64
4.1	Função de densidade estimada para os estimadores de η ($n = 40$); modelo log-normal.	87
4.2	Função de densidade estimada para os estimadores de η ($n = 40$); modelo log-t-Student.	88
4.3	Função de densidade estimada para os estimadores de ϕ ($n = 40$); modelo log-normal.	88
4.4	Função de densidade estimada para os estimadores de ϕ ($n = 40$); modelo log-t-Student.	89
4.5	Gráficos da função densidade (a) e da função de distribuição acumulada (b) estimados.	90

Lista de Tabelas

2.1	Função geradora de densidades e expressões de w , ξ , δ_{20000} e δ_{20002} para algumas distribuições simétricas. ^a	10
2.2	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo normal.	21
2.3	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo normal.	22
2.4	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo t-Student.	23
2.5	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo t-Student.	24
2.6	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo logístico II.	25
2.7	Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ para $p = 6$; modelo logístico II.	26
2.8	Valor da estatística de teste e p -valor para as hipóteses $\mathcal{H}_0 : \beta_4 = 0$, $\mathcal{H}_0 : \beta_5 = 0$ e $\mathcal{H}_0 : \beta_6 = 0$ no modelo (2.12).	32
3.1	Expressões de $t(y, \theta)$ e $a(y, \theta)$ para alguns modelos de dispersão. ^a	48
3.2	Expressões de D_{2l} , D_{3l} e D_{4l} para alguns modelos de dispersão.	49
3.3	Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo von Mises.	59
3.4	Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo von Mises.	60
3.5	Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo log-gamma.	62
3.6	Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo log-gama.	63

3.7	Teste de $\mathcal{H}_0 : \beta_1 = \beta_{10}$ contra $\mathcal{H}_0 : \beta_1 \neq \beta_{10}$ (p -valor em parênteses).	66
3.8	Teste de $\mathcal{H}_0 : \beta_1 = 0$	67
4.1	Função geradora de densidades e expressões de w , $\alpha_{2,0}$ e $\alpha_{2,2}$ para algumas distribuições log-simétricas. ^a	78
4.2	$\alpha_{3,1}$ e $\alpha_{3,3}$ para algumas distribuições log-simétricas. ^a	81
4.3	Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc); modelo log-t-Student - estimação do parâmetro de escala.	86
4.4	Média das 10000 estimativas (media), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc); modelo log-t-Student - estimação do parâmetro de forma.	86
4.5	Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc): model log-normal - estimação do parâmetro de escala.	86
4.6	Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc): model log-normal - estimação do parâmetro de forma.	87
4.7	Estimativas pontuais e intervalares para o parâmetro η	90
4.8	Estimativas pontuais e intervalares para o parâmetro ϕ	90

LISTA DE TABELAS

Capítulo 1

Introdução

Em inferência estatística é comum a utilização das estatísticas de Wald, razão de verossimilhanças e escore para se testar hipóteses de interesse em modelos paramétricos. Estas estatísticas foram nomeadas por Rao (2005) como “the Holy Trinity”. Recentemente, a estatística gradiente (Terrel, 2002) tem recebido atenção especial na literatura pois, ao contrário das estatísticas de Wald e escore, não necessita do cálculo, estimação ou inversão da matriz de informação de Fisher. Esta estatística mede a distância entre os estimadores de máxima verossimilhança irrestrito, $\hat{\theta}$, e restrito (sob a hipótese nula), $\tilde{\theta}$, de uma perspectiva diferente: mede a projeção ortogonal do vetor escore (avaliado sob a hipótese nula) sobre o vetor $\hat{\theta} - \tilde{\theta}$. Além disso, a estatística gradiente tem as mesmas propriedades assintóticas de primeira ordem que as estatísticas de Wald, razão de verossimilhanças e escore.

Para a maioria dos modelos paramétricos, em amostra finitas, a distribuição exata das estatísticas de Wald, razão de verossimilhanças, escore e gradiente é desconhecida. Por essa razão, faz-se necessário o uso de teoria assintótica para aproximar a distribuição exata destas estatísticas. Sob condições de regularidade, a distribuição limite destas estatísticas de teste é χ^2 central com erro de aproximação de ordem $\mathcal{O}(n^{-1})$ sob a hipótese nula. A aproximação pela distribuição χ^2 é válida quando o número de observações na amostra tende a infinito e assim, em amostras finitas (pequenas ou moderadas), a aproximação pela distribuição χ^2 pode levar a distorção na probabilidade de erro tipo I. Para solucionar este problema, uma alternativa é obter correções para as estatísticas de testes baseadas em teoria assintótica de segunda ordem. Além disso, as estatísticas de Wald, razão de verossimilhanças, escore

e gradiente diferem nas propriedades assintóticas de segunda ordem e podem apresentar diferenças consideráveis de comportamento em amostras finitas.

Nas últimas décadas, algumas propostas surgiram na literatura para melhorar a taxa de convergência da distribuição das estatísticas de teste pela distribuição χ^2 . Como exemplo, podemos citar as correções de Bartlett para a estatística da razão de verossimilhanças e tipo-Bartlett para as estatísticas score e gradiente. Maiores detalhes sobre estas correções podem ser encontrados em [Cordeiro e Cribari-Neto \(2016\)](#) e [Lemonte \(2016\)](#). As correções conduzem a testes com probabilidades de erro tipo I menos distorcidas quando o tamanho da amostra é pequeno ou moderado. Sendo assim, um dos objetivos desta tese é derivar expressões matriciais para o fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de regressão lineares simétricos e log-simétricos, e na classe dos modelos de dispersão a partir dos resultados gerais de [Vargas et al. \(2013\)](#).

Um outro tópico importante em inferência estatística é a estimação pontual dos parâmetros. Sob condições de regularidade, o estimador de máxima verossimilhança é consistente e assintoticamente normal, porém pode ser viesado. Em geral, o viés é utilizado como um critério de seleção de estimadores. Para o estimador de máxima verossimilhança, o viés é de ordem $\mathcal{O}(n^{-1})$ e seu erro padrão assintótico é de ordem $\mathcal{O}(n^{-1/2})$. Por este motivo, o viés é considerado desprezível em amostras de tamanho grande. No entanto, em amostras pequenas ou moderadas pode ser não desprezível. Sendo assim, considerando diferentes técnicas, muitos artigos foram publicados nos últimos anos na área de correção de viés e possibilitam a obtenção de estimadores corrigidos que reduzem os vieses dos estimadores. Um outro critério para seleção de um estimador é o viés mediano. Dizemos que um estimador é não viesado por mediana se a mediana da distribuição deste estimador é igual ao verdadeiro valor do parâmetro. Em geral, em amostras pequenas ou moderadas, o viés mediano do estimador pode ser não desprezível. Sendo assim, um outro objetivo desta tese é propor estimadores corrigidos para os parâmetros que indexam a classe de distribuições log-simétricas que reduzem o viés e o viés mediano dos estimadores.

No que segue, esta tese contém cinco capítulos. Nos Capítulos 2 e 3, tratamos da correção tipo-Bartlett para a estatística gradiente em algumas classes de modelos de regressão. No Capítulo 4, investigamos algumas técnicas de correção de viés e de correção de viés mediano.

Cada capítulo desta tese é independente dos demais, ou seja, em cada capítulo introduzimos as notações e comentários intrínsecos a cada um deles. Isto possibilita a leitura individual, em qualquer ordem, dos Capítulos 2, 3 e 4. Além disso, na introdução de cada capítulo, é feita uma revisão bibliográfica de cada tema e ao final de cada capítulo são apresentados os detalhes técnicos.

No Capítulo 2 derivamos uma expressão matricial para o fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de regressão lineares simétricos e verificamos que este resultado é válido para a classe dos modelos de regressão lineares log-simétricos. Além disso, comparamos numericamente a probabilidade de erro tipo I e o poder do teste proposto com as probabilidades de erro tipo I e os poderes de outros testes existentes na literatura e dos testes corrigidos via bootstrap. Apresentamos também duas aplicações a conjuntos de dados reais.

No Capítulo 3 obtivemos uma expressão matricial para o fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de dispersão. Estudos de simulação são apresentados para comparar o tamanho do teste proposto com os tamanhos de outros testes propostos na literatura e dos testes corrigidos via bootstrap. Além disso, avaliamos numericamente o poder de cada teste. Para fins ilustrativos, apresentamos duas aplicações a conjuntos de dados reais.

No Capítulo 4, propomos estimadores corrigidos para os parâmetros que indexam a classe de distribuições log-simétricas que reduzem o viés e o viés mediano. Também comparamos numericamente os estimadores propostos e apresentamos uma aplicação a um conjunto de dados reais.

No Capítulo 5 apresentamos as considerações finais e sugestões de pesquisas futuras.

Capítulo 2

Aprimoramento de testes de hipóteses em modelos de regressão lineares simétricos e log-simétricos

Resumo

Neste capítulo, estudamos testes de hipóteses nos modelos de regressão lineares simétricos (MRLS) e nos modelos de regressão lineares log-simétricos (MRLLS) quando o tamanho da amostra é pequeno ou moderado. A distribuição qui-quadrado de referência fornece uma boa aproximação para a distribuição nula (i.e. sob a hipótese nula) das estatísticas Wald, razão de verossimilhanças (LR), escore e gradiente quando o tamanho da amostra é grande, mas pode apresentar resultados distorcidos em amostras pequenas ou moderadas. Para reduzir a distorção do tamanho do teste gradiente, derivamos um fator de correção tipo-Bartlett para classe dos MRLS e verificamos que esses resultados são válidos para a classe dos MRLLS. Além disso, comparamos numericamente a probabilidade de erro tipo I do teste proposto com a probabilidade de erro tipo I de outros testes propostos na literatura. Também avaliamos via simulação o poder de cada teste. Nossos estudos de simulações sugerem que, em amostras de tamanho pequeno ou moderado, os testes corrigidos apresentam probabilidade de erro tipo I mais próxima do verdadeiro nível nominal sem perda de poder e, portanto, são mais confiáveis para testar hipóteses nos MRLS e MRLLS do que os testes usuais. Também apresentamos aplicações a conjuntos de dados reais para fins ilustrativos.

Palavras-chave: Modelos de regressão lineares simétricos, correção tipo-Bartlett, correção de Bartlett, modelos de regressão lineares log-simétricos, estatística gradiente, estatística escore, estatística da razão de verossimilhanças, estatística Wald.

2.1 Introdução

Os modelos de regressão lineares normais são amplamente utilizados em diversas áreas do conhecimento e várias extensões têm sido propostas para flexibilizar a suposição de normalidade dos erros. Em particular, a suposição de erros simétricos (seguindo uma distribuição na classe simétrica de distribuições) tem sido amplamente investigada nas últimas décadas; ver [Villegas et al. \(2013\)](#), [Lemonte \(2012\)](#), [Paula e Cysneiros \(2009\)](#), [Cysneiros et al. \(2007\)](#), [Galea et al. \(2005\)](#) e [Liu \(2000\)](#). Os modelos de regressão lineares simétricos possibilitam a modelagem de dados com características de caudas mais leves ou caudas mais pesadas do que as da distribuição normal englobando diferentes distribuições para o erro aleatório como por exemplo, t-Student, logística I, logística II, normal contaminada, exponencial potência e slash. Já os modelos log-simétricos, estudados por [Vanegas e Paula \(2016\)](#), são derivados a partir de uma transformação exponencial de uma variável aleatória simétrica. Essa classe de modelos, que tem suporte no conjunto dos números reais positivos, inclui distribuições com caudas mais leves e caudas mais pesadas do que as da distribuição log-normal. [Vanegas e Paula \(2015a\)](#) propuseram um modelo de regressão semiparamétrico supondo que a variável resposta tem distribuição log-simétrica. Esses autores consideraram uma estrutura de regressão não linear para a mediana e uma estrutura não paramétrica para modelar o parâmetro de assimetria. O modelo de regressão proposto pelos autores flexibiliza a suposição de erros log-normais incluindo diferentes distribuições como, por exemplo, log-t-Student, log-logística I, log-logística II, log-normal-contaminada, log-exponencial-potência e log-slash.

Um dos objetivos na modelagem inferencial consiste em testar hipóteses sobre os parâmetros do modelo. As estatísticas comumente utilizadas são Wald, razão de verossimilhanças e escore. Recentemente, uma outra estatística, a estatística gradiente ([Terrell, 2002](#)), tem recebido atenção especial pois não necessita do cálculo nem da inversão da matriz de infor-

mação de Fisher e compartilha as mesmas propriedades assintóticas de primeira ordem com as demais (Lemonte e Ferrari, 2012a).

Considere o problema de testar uma hipótese simples ou composta \mathcal{H}_0 contra uma hipótese alternativa \mathcal{H}_1 . Sob condições de regularidade, as estatísticas Wald, LR, escore e gradiente têm, sob a hipótese nula, distribuição assintótica qui-quadrado central (χ_q^2), em que q é o número de parâmetros a serem testados. Quando o tamanho da amostra é pequeno ou moderado a aproximação χ_q^2 para as distribuições das estatísticas de teste pode apresentar resultados distorcidos. Uma alternativa para melhorar essa aproximação em amostras pequenas e moderadas é a utilização de fatores de correção de Bartlett para o teste LR e tipo-Bartlett para os testes escore e gradiente. Um teste LR modificado foi proposto inicialmente por Bartlett (1937) e estudado de maneira mais geral por Lawley (1956). Em problemas regulares, Lawley (1956) obteve um fator de correção que envolve momentos de derivadas até a quarta ordem do logaritmo da função de verossimilhança e mostrou que a estatística LR modificada tem todos os momentos iguais aos da distribuição χ_q^2 até ordem $\mathcal{O}(n^{-1})$. Para o teste escore, Cordeiro e Ferrari (1991) propuseram uma correção que envolve um polinômio de segundo grau na própria estatística escore para melhorar a taxa de convergência para a distribuição χ_q^2 . Recentemente, para a estatística gradiente, Vargas et al. (2013) obtiveram um fator de correção tipo-Bartlett por meio de um polinômio de segundo grau na estatística gradiente, que melhora a aproximação da distribuição desta estatística de teste pela distribuição χ_q^2 . É importante ressaltar que, no caso geral, não existe um fator de correção tipo-Bartlett para a estatística de Wald. Estas correções se tornaram amplamente utilizadas para melhorar a aproximação da distribuição qui-quadrado de referência em vários modelos paramétricos; ver, por exemplo, Chan et al. (2014), Silva-Junior et al. (2014), Vargas et al. (2014), Bayer e Cribari-Neto (2013), Lemonte e Ferrari (2011), Lemonte et al. (2010), Lagos et al. (2010), da Silva e Cordeiro (2009), Melo et al. (2009), Barroso e Cordeiro (2005).

Diversos trabalhos têm focado na derivação de fórmulas matriciais para os fatores de correção de Bartlett e tipo-Bartlett nos modelos de regressão simétricos. Por exemplo, Ferrari e Uribe-Opazo (2001) desenvolveram uma correção de Bartlett para o teste LR considerando uma estrutura de regressão linear para o vetor de médias e Cordeiro (2004) estende esses resultados considerando uma estrutura de regressão não-linear. Para o teste escore,

Uribe-Opazo et al. (2008) derivaram um fator de correção tipo-Bartlett nos modelos de regressão lineares simétricos, enquanto que Cysneiros et al. (2010) obtiveram uma correção tipo-Bartlett para a estatística escore na situação em que o componente sistemático é não-linear. Quanto aos modelos de regressão lineares log-simétricos, não encontramos nenhum estudo na literatura acerca das correções de Bartlett para a estatística LR e tipo-Bartlett para as estatísticas escore e gradiente. Embora, em geral, as fórmulas algébricas dos fatores de correção de Bartlett e tipo-Bartlett sejam complicadas, nos modelos de regressão lineares simétricos e log-simétricos essas expressões não são complicadas e podem ser facilmente implementadas em linguagens de programação como Ox (Doornik, 2013), R (R Core Team, 2013) e MAPLE (Rafter et al., 2003).

Neste capítulo, derivamos um fator de correção tipo-Bartlett, em notação matricial, para o teste gradiente nos modelos de regressão lineares simétricos. Além disso, derivamos os fatores de correção de Bartlett para a estatística da razão de verossimilhanças e tipo-Bartlett para as estatísticas escore e gradiente em modelos de regressão lineares log-simétricos. Por meio de simulações de Monte Carlo, avaliamos e comparamos numericamente a probabilidade de erro tipo I dos testes gradiente, escore e LR modificados bem como dos testes usuais de Wald, LR, escore e gradiente nos modelos de regressão normal, t-Student e logístico II. Nossos estudos de simulações sugerem que a probabilidade de erro tipo I do teste gradiente aperfeiçoado é bem próxima do verdadeiro nível nominal. Portanto, o teste gradiente modificado se mostra uma alternativa atraente aos testes assintóticos clássicos e aos testes corrigidos (LR e escore) nos MRLS e MRLLS quando o número de observações é pequeno ou moderado.

Este capítulo está organizado como segue. Na Seção 2.2 definimos os modelos de regressão lineares simétricos e discutimos a estimação e testes de hipóteses para os parâmetros da regressão. Na Seção 2.3 apresentamos uma correção de Bartlett para o teste LR, tipo-Bartlett para o teste escore e derivamos um fator de correção tipo-Bartlett para o teste gradiente nos MRLS. Na Seção 2.4 apresentamos os modelos de regressão lineares log-simétricos e derivamos os fatores de correções de Bartlett e tipo-Bartlett para essa classe de modelos. Na Seção 2.5 apresentamos estudos de simulação de Monte Carlo para comparar o tamanho e o poder de todos os testes e testes bootstrap. Na Seção 2.6 consideramos duas aplicações

a conjuntos de dados reais e, por fim, na Seção 2.7 apresentamos as considerações finais. Detalhes técnicos são apresentados na Seção 2.8.

2.2 Modelos de regressão lineares simétricos, estimação e testes

Seja y uma variável aleatória simétrica contínua com parâmetro de locação $\mu \in \mathbb{R}$ e parâmetro de escala $\phi > 0$, cuja função densidade é dada por

$$\pi(y; \mu, \phi) = \frac{1}{\phi} h\left(\left(\frac{y - \mu}{\phi}\right)^2\right), \quad y \in \mathbb{R}, \quad (2.1)$$

para alguma função $h : \mathbb{R} \rightarrow [0, \infty)$ tal que $\int_0^\infty u^{-1/2} h(u) du = 1$. Esta condição garante que $\pi(\cdot; \mu, \phi)$ seja de fato uma legítima função densidade. A função $h(\cdot)$ é denominada função geradora de densidades e pode depender de parâmetros desconhecidos. Se y tem uma distribuição simétrica com parâmetros μ e ϕ , escrevemos $y \sim S(\mu, \phi^2)$. São membros dessa classe as distribuições normal, Cauchy, t-Student, logística I, logística II, exponencial potência, entre outras; ver Tabela 2.1.

Algumas propriedades clássicas da distribuição normal são válidas para a classe das distribuições simétricas. Por exemplo, se $y \sim S(\mu, \phi^2)$, então $a + by \sim S(a + b\mu, b^2\phi^2)$, sendo $a, b \in \mathbb{R}$ com $b \neq 0$ e, em particular, $z = (y - \mu)/\phi \sim S(0, 1)$ com função densidade $\pi(z; 0, 1) = h(z^2)$, $z \in \mathbb{R}$. Desde que existam, $E(y) = \mu$ e $\text{Var}(y) = \phi\xi$, para uma constante $\xi > 0$. Se $\psi(t) = E(e^{ity})$ denota a função característica de y então $\psi(t) = e^{it\mu}\varphi(t^2\phi^2)$ para $t \in \mathbb{R}$ e alguma função real φ . Assim, $\xi = -2\frac{d\varphi(u)}{du}\big|_{u=0}$ e não depende dos parâmetros de locação e escala. A quantidade ξ para algumas distribuições é apresentada na Tabela 2.1. Outros resultados e propriedades para os modelos simétricos são explorados em Berkane e Bentler (1986), Rao (1990) e Fang et al. (1990).

Considere y_1, \dots, y_n variáveis aleatórias independentes com $y_l \sim S(\mu_l, \phi^2)$, para $l = 1, \dots, n$. Os modelos de regressão lineares simétricos são definidos como

$$y_l = \mathbf{x}_l^\top \boldsymbol{\beta} + \phi \epsilon_l, \quad l = 1, \dots, n \quad (2.2)$$

sendo $\mathbf{x}_l = (x_{l1}, \dots, x_{lp})^\top$ um vetor de variáveis explicativas associadas à l -ésima resposta, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ um vetor de parâmetros desconhecidos a ser estimado e $\epsilon_1, \dots, \epsilon_n$, variáveis aleatórias independentes com $\epsilon_l \sim S(0, 1)$. Assumimos que o parâmetro de precisão é desconhecido porém o mesmo para todas as observações e que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ é uma matriz $n \times p$ de posto coluna completo, isto é, $\text{posto}(\mathbf{X}) = p$. Assumimos também que as condições de regularidade usuais para estimação de máxima verossimilhança e inferência em grandes amostras são satisfeitas; ver Cox e Hinkley (1974, Cap. 9).

A suposição de que $\epsilon_l \sim S(0, 1)$ flexibiliza a suposição de normalidade dos erros possibilitando distribuições com caudas mais pesadas (t-Student, logística II, exponencial potência para $0 < k < 1$) ou caudas mais leves (logística I, exponencial potência para $-1 < k < 0$) do que as da distribuição normal. O modelo de regressão linear normal é um caso particular dos modelos de regressão lineares simétricos definidos em (2.2).

O logaritmo da função de verossimilhança para o vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ é dado por

$$l(\boldsymbol{\theta}) = -n \log(\phi) + \sum_{l=1}^n g(z_l), \quad (2.3)$$

sendo $g(z_l) = \log h(z_l^2)$ e $z_l = (y_l - \mathbf{x}_l^\top \boldsymbol{\beta})/\phi$ o erro padronizado relativo à l -ésima observação. O vetor escore de $\boldsymbol{\beta}$ e ϕ é dado por $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_\beta(\boldsymbol{\theta})^\top, U_\phi(\boldsymbol{\theta})^\top)^\top$ com

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \phi^{-2} \mathbf{X}^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad U_\phi(\boldsymbol{\theta}) = \phi^{-1} (\phi^{-2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - n),$$

em que $\mathbf{y} = (y_1, \dots, y_n)^\top$ e $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ com $w_l = -2d \log h(u)/du|_{u=z_l^2}$.

Os estimadores de máxima verossimilhança (EMV) de $\boldsymbol{\beta}$ e ϕ , obtidos resolvendo simultaneamente as equações $\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{0}$ e $U_\phi(\boldsymbol{\theta}) = 0$, são solução de $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{y}$ e $\widehat{\phi}^2 = n^{-1} \widehat{\mathbf{e}}^\top \widehat{\mathbf{W}} \widehat{\mathbf{e}}$ sendo $\widehat{\mathbf{e}} = \mathbf{y} - \mathbf{X} \widehat{\boldsymbol{\beta}}$ o vetor de resíduos. Para cada l , w_l pode ser interpretado como o peso da l -ésima observação nas estimativas dos parâmetros. Na Tabela 2.1 apresentamos w_l para algumas distribuições simétricas. Note que a distribuição normal assume o mesmo peso para todas as observações ($w_l = 1, l = 1, \dots, n$) na estimação de máxima verossimilhança dos parâmetros $\boldsymbol{\beta}$ e ϕ . Para as distribuições Cauchy, t-Student, logística II e exponencial potência ($0 < k < 1$), os pesos w_l são funções decrescentes em

$|z_l|$. Isso significa dizer que os EMV de β e ϕ são robustos à presença de outliers no sentido de que observações com um valor grande para $|z_l|$ têm pesos menores. Nas distribuições logística I e exponencial potência ($-1 < k < 0$), os pesos w_l são funções crescentes em $|z_l|$ pois essas distribuições têm caudas mais leves do que as da distribuição normal.

Sejam $\delta_{abcde} = E(g^{(1)a} g^{(2)b} g^{(3)c} g^{(4)d} z^e)$, para $a, b, c, d, e \in \{0, 1, 2, 3, 4\}$, $g^{(r)} = d^r g(z)/dz^r$ e $z \sim S(0, 1)$. Para algumas distribuições simétricas, os δ 's são dados em [Uribe-Opazo et al. \(2008\)](#) (ver detalhes técnicos na Seção 2.8.2) e satisfazem relações de regularidade tais como $\delta_{01001} = \delta_{10000} = 0$, $\delta_{20000} = -\delta_{01000}$, $\delta_{00010} = -\delta_{10100}$, $\delta_{40000} = -3\delta_{21000}$, $\delta_{01002} = 2 - \delta_{20002}$, $\delta_{11001} + \delta_{00101} + \delta_{01000} = 0$, $2\delta_{00101} + \delta_{00012} + \delta_{10102} = 0$ e $3\delta_{01002} + \delta_{11003} + \delta_{00103} = 0$. A matriz de informação de Fisher para $\theta = (\beta^\top, \phi)^\top$ é bloco diagonal e é dada por $\mathbf{K} = \text{diag}\{\mathbf{K}_\beta, K_\phi\}$, sendo $\mathbf{K}_\beta = \phi^{-2} \delta_{20000} \mathbf{X}^\top \mathbf{X}$ e $K_\phi = n\phi^{-2}(\delta_{20002} - 1)$. Portanto, β e ϕ são globalmente ortogonais e seus EMV são assintoticamente não correlacionados. As quantidades δ_{20000} e δ_{20002} são dadas na Tabela para algumas distribuições da classe simétrica.

Tabela 2.1: Função geradora de densidades e expressões de w , ξ , δ_{20000} e δ_{20002} para algumas distribuições simétricas.^a

Modelo	$h(u), u > 0$	w	ξ	δ_{20000}	δ_{20002}
normal	$\frac{1}{\sqrt{2\pi}} e^{-u/2}$	1	1	1	3
Cauchy	$\frac{1}{\pi} (1+u)^{-1}$	$\frac{2}{1+z^2}$	não existe	$\frac{1}{2}$	$\frac{3}{2}$
t-Student	$\frac{\nu^{\nu/2}}{B(1/2, \nu/2)} (\nu+u)^{-\frac{\nu+1}{2}}$, $\nu > 0$	$\frac{\nu+1}{\nu+z^2}$	$\frac{\nu}{1+\nu}, \nu > 2$	$\frac{\nu+1}{\nu+3}$	$\frac{3(\nu+1)}{\nu+3}$
logística I	$c \frac{e^{-u}}{(1+e^{-u})^2}, c \cong 1.4843$	$2 \tanh(z^2/2)$	$\cong 0.79569$	$\cong 1.47724$	$\cong 4.01378$
logística II	$\frac{e^{-\sqrt{u}}}{(1+e^{-\sqrt{u}})^2}$	$\frac{e^{- z -1}}{ z (1+e^{- z })}$	$\frac{\pi^2}{3}$	$\frac{1}{3}$	$\cong 2.42996$
exp. potência	$\frac{1}{C(k)} \exp\{-\frac{1}{2}u^{1/(1+k)}\}$, $-1 < k \leq 1$	$\frac{1}{(1+k)z^{2k/(1+k)}}$	$2^{1+k} \frac{\Gamma(\frac{3}{2}(1+k))}{\Gamma(\frac{1+k}{2})}$	$2^{1-k} \frac{\Gamma(\frac{3-k}{2})}{(1+k)^2 \Gamma(\frac{1+k}{2})}$	$\frac{3+k}{1+k}$

^a $B(\cdot, \cdot)$ e $\Gamma(\cdot)$ são as funções beta e gama, respectivamente, e $C(k) = \Gamma(1 + \frac{1+k}{2}) 2^{1+(1+k)/2}$.

As equações $\mathbf{U}_\beta(\theta) = \mathbf{0}$ e $U_\phi(\theta) = 0$ não podem ser resolvidas explicitamente, exceto para o modelo normal. Portanto, devemos utilizar algum processo iterativo para determinar os EMV de β e ϕ . Como a matriz de informação de Fisher é bloco diagonal, o método scoring de Fisher para estimação simultânea dos parâmetros é dado pela resolução iterativa

das equações

$$\begin{aligned}\boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + \frac{1}{\delta_{20000}}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(m)}), \\ \phi^{(m+1)} &= \phi^{(m)} + \frac{1}{\phi^{(m)}(\delta_{20002} - 1)} \left(\frac{1}{n}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(m)})\mathbf{W}^{(m)}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(m)}) - \phi^{(m)2} \right),\end{aligned}$$

para $m = 0, 1, \dots$. O processo iterativo pode ser inicializado considerando $\boldsymbol{\beta}^{(0)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$, o estimador de mínimos quadrados ordinário de $\boldsymbol{\beta}$, e $\phi^{(0)} = ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(0)})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(0)})/n)^{1/2}$. Algumas distribuições simétricas podem ser obtidas como uma mistura no parâmetro de precisão da distribuição normal como, por exemplo, as distribuições t-Student e exponencial potência ($0 < k < 1$), ver [Andrews e Mallows \(1974\)](#) e [West \(1987\)](#). Portanto, o algoritmo EM (Dempster et al. 1977) pode ser utilizado como um método iterativo para estimação dos parâmetros. No software R dispomos do pacote `ssym` desenvolvido por [Vanegas e Paula \(2015b\)](#) para ajustar MRLLS; esse pacote pode ser utilizado também para obter as estimativas dos parâmetros nos MRLS.

Consideramos o problema de testar a hipótese nula composta $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ contra a hipótese alternativa $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$, em que o vetor de parâmetros desconhecidos $\boldsymbol{\beta}$ é particionado como $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, sendo $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ o vetor de parâmetros de interesse, $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$ o vetor de parâmetros de perturbação e $\boldsymbol{\beta}_{10}$ um vetor especificado de constantes de dimensão q . A partição do vetor $\boldsymbol{\beta}$ induz às partições $\mathbf{U}_\beta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top)^\top$, com $\mathbf{U}_{\beta_1} = \phi^{-2} \mathbf{X}_1^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ e $\mathbf{U}_{\beta_2} = \phi^{-2} \mathbf{X}_2^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ e

$$\mathbf{K}_\beta = \begin{bmatrix} \mathbf{K}_{\beta_{11}} & \mathbf{K}_{\beta_{12}} \\ \mathbf{K}_{\beta_{21}} & \mathbf{K}_{\beta_{22}} \end{bmatrix} = \frac{\delta_{20000}}{\phi^2} \begin{bmatrix} \mathbf{X}_1^\top \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{X}_2 \end{bmatrix},$$

com a matriz \mathbf{X} particionada da forma $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$, sendo \mathbf{X}_1 uma matriz $n \times q$ e \mathbf{X}_2 uma matriz $n \times (p - q)$.

Para testar a hipótese $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ contra $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$, as estatísticas Wald, LR, escore e gradiente são dadas, respectivamente, por

$$S_W = \delta_{20000} \hat{\phi}^{-2} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^\top (\mathbf{R}^\top \mathbf{R}) (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}),$$

$$S_{LR} = 2\{\ell(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\phi}) - \ell(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2, \widetilde{\phi})\},$$

$$S_R = \frac{1}{\widetilde{\phi}^2 \delta_{20000}^{-1}} (\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})^\top \widetilde{\mathbf{W}} \mathbf{X}_1 (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{X}_1^\top \widetilde{\mathbf{W}} (\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}),$$

$$S_T = \widetilde{\phi}^{-2} (\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})^\top \widetilde{\mathbf{W}} \mathbf{X}_1 (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}),$$

em que $(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\phi})$ é o EMV irrestrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$, $(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2, \widetilde{\phi})$ é o EMV restrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$ e $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_1$. A notação chapéu e til é utilizada para indicar que as funções são avaliadas nos estimadores irrestrito e restrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$, respectivamente. Sob a hipótese nula, a distribuição limite de todas as estatísticas de teste é χ_q^2 . Nota-se que, ao contrário da estatística de Wald e escore, a estatística gradiente não depende de nenhuma inversão matricial. Pode-se mostrar que, para erros com distribuição normal, as estatísticas escore e gradiente coincidem.

2.3 Testes de hipóteses aperfeiçoados

Sabe-se que, sob \mathcal{H}_0 e condições de regularidade, as estatísticas S_W , S_{LR} , S_R e S_T são assintoticamente equivalentes. Em particular, têm a mesma distribuição limite χ_q^2 com erro de aproximação de ordem $\mathcal{O}(n^{-1})$. Porém, em amostras pequenas e moderadas, a utilização da aproximação pela χ^2 pode causar distorções no tamanho dos testes. Para solucionar esse problema, uma alternativa é obter correções para as estatísticas de testes baseadas em teoria assintótica de segunda ordem. Nos MRLS, um fator de correção de Bartlett para o teste LR foi obtido em Ferrari e Uribe-Opazo (2001) e um fator de correção tipo-Bartlett para o teste escore foi obtido em Uribe-Opazo et al. (2008). Neste capítulo derivamos um fator de correção tipo-Bartlett para o teste gradiente nos modelos de regressão lineares simétricos utilizando os resultados gerais de Vargas et al. (2013). Esses resultados são novos e representam contribuições para o aprimoramento de testes de hipóteses nos MRLS.

Inicialmente definimos as matrizes: $\mathbf{Z} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$, $\mathbf{Z}_2 = \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$ (se $q < p$; $\mathbf{Z}_2 = \mathbf{0}_{n \times n}$, se $q = p$), $\mathbf{Z}_d = \text{diag}\{z_{11}, \dots, z_{nn}\}$, $\mathbf{Z}_{2d} = \text{diag}\{z_{211}, \dots, z_{2nn}\}$. As matrizes $\phi^2 \delta_{20000}^{-1} \mathbf{Z}$ e $\phi^2 \delta_{20000}^{-1} \mathbf{Z}_2$ representam as estruturas de covariância assintótica de $\mathbf{X}\widehat{\boldsymbol{\beta}}$ e $\mathbf{X}_2\widetilde{\boldsymbol{\beta}}_2$, respectivamente. Sejam $\rho_{ZZ} = n\text{tr}(\mathbf{Z}_d \mathbf{Z}_d)$, $\rho_{Z_2 Z_2} = n\text{tr}(\mathbf{Z}_{2d} \mathbf{Z}_{2d})$ e $\rho_{ZZ_2} = n\text{tr}(\mathbf{Z}_d \mathbf{Z}_{2d})$, em que tr é o operador traço.

Dos resultados gerais de Lawley (1956), Ferrari e Uribe-Opazo (2001) obtiveram um fator de correção de Bartlett para o teste LR para testar $\mathcal{H}_0 : \beta_1 = \beta_{10}$ nos modelos de regressão lineares simétricos como

$$S_{LR}^* = S_{LR}(1 - a_{LR}), \quad (2.4)$$

sendo $a_{LR} = A_{LR} + A_{LR,\beta\phi}$,

$$A_{LR} = \frac{d_0}{nq}(\rho_{ZZ} - \rho_{Z_2Z_2}), \quad A_{LR,\beta\phi} = \frac{d_1}{n} + \frac{d_2}{n} \frac{2p - q}{2},$$

e

$$d_0 = \frac{\delta_{00010}}{4\delta_{20000}^2}, \quad d_1 = -\frac{m_2 m_3}{2m_1^2} - \frac{2m_3 + m_3^2 + m_4}{2m_1}, \quad d_2 = -\frac{m_3^2}{2m_1},$$

com

$$m_1 = \delta_{01002} - 1, \quad m_2 = 4 - \delta_{00103} - 6\delta_{01002}, \quad m_3 = \frac{\delta_{00101} + 2\delta_{01000}}{\delta_{20000}}, \quad m_4 = \frac{\delta_{00012} - 6\delta_{11001}}{\delta_{20000}}.$$

O fator de correção $1 - a_{LR}$ é comumente denominado de fator de correção de Bartlett. Com esse fator de correção, sob a hipótese nula, a estatística S_{LR}^* tem distribuição assintótica χ_q^2 com erro de aproximação da ordem $\mathcal{O}(n^{-2})$.

Uma correção tipo-Bartlett para o teste escore para testar $\mathcal{H}_0 : \beta_1 = \beta_{10}$ nos MRLS foi derivada por Uribe-Opazo et al. (2008) utilizando os resultados gerais de Cordeiro e Ferrari (1991). A estatística escore modificada é dada por

$$S_R^* = S_R [1 - (c_R + b_R S_R + a_R S_R^2)], \quad (2.5)$$

em que $a_R = A_{R3}/[12q(q+2)(q+4)]$, $b_R = (A_{R22} - 2A_{R3})/[12q(q+2)]$, $c_R = (A_{R11} - A_{R22} + A_{R3})/(12q)$, $A_{R11} = A_{R1} + A_{R1,\beta\phi}$, $A_{R22} = A_{R2} + A_{R2,\beta\phi}$,

$$A_{R1} = \frac{12b_0}{n}(\rho_{ZZ_2} - \rho_{Z_2Z_2}), \quad A_{R2} = -\frac{9b_0}{n}(\rho_{ZZ} - 2\rho_{ZZ_2} + \rho_{Z_2Z_2}), \quad A_{R3} = 0,$$

$$A_{R1,\beta\phi} = \frac{12b_1}{n}q(p-q) - \frac{6b_2}{n}q, \quad A_{R2,\beta\phi} = -\frac{12b_3}{n}q(q+2),$$

com

$$b_0 = \frac{\delta_{21000}}{\delta_{20000}^2} + 1, \quad b_1 = \frac{\delta_{11001}(\delta_{11001} - \delta_{01000})}{\delta_{20000}^2(\delta_{20002} - 1)}, \quad b_3 = \frac{\delta_{11001}^2}{\delta_{20000}^2(\delta_{20002} - 1)} \quad e$$

$$b_2 = \frac{1}{\delta_{20000}^2(\delta_{20002} - 1)^2} [2\delta_{11001}(2\delta_{01002} + \delta_{00103}) + (\delta_{20002} - 1)(4\delta_{30001} + \delta_{40002} + \delta_{21002} - 2\delta_{01000})].$$

O fator de correção $[1 - (c_R + b_R S_R + a_R S_R^2)]$ se reduz a $[1 - (c_R + b_R S_R)]$ em (2.5) pois $a_R = 0$ em MRLS e é chamado de fator de correção tipo-Bartlett para o teste escore. A distribuição da estatística modificada S_R^* é melhor aproximada pela distribuição χ_q^2 com erro de aproximação reduzido de ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-2})$.

Recentemente, uma correção tipo-Bartlett para o teste gradiente foi proposta por Vargas et al. (2013). Os autores obtiveram uma estatística gradiente modificada por um polinômio de segundo grau na própria estatística gradiente cuja distribuição assintótica, sob a hipótese nula, é qui-quadrado com erro de ordem $\mathcal{O}(n^{-2})$. Assim, espera-se que o teste gradiente aperfeiçoado forneça resultados inferenciais mais acurados em amostras pequenas e moderadas do que o teste gradiente usual.

O fator de correção tipo-Bartlett derivado por Vargas et al. (2013) é bem geral e pode ser particularizado para cada modelo específico de interesse. Por envolver funções dos momentos de derivadas do logaritmo da função de verossimilhança, as expressões para um particular modelo de regressão podem ser difíceis de serem obtidas. Como veremos a seguir, derivamos expressões fechadas para o fator de correção tipo-Bartlett para o teste gradiente nos MRLS.

Proposição 1. *Considere o problema de testar a hipótese nula $\mathcal{H}_0 : \beta_1 = \beta_{10}$ em modelos de regressão lineares simétricos. A correção tipo-Bartlett para a estatística gradiente é dada por*

$$S_T^* = S_T [1 - (c_T + b_T S_T + a_T S_T^2)], \quad (2.6)$$

sendo a_T, b_T e c_T da ordem $\mathcal{O}(n^{-1})$ e dados por $a_T = A_{T3}/[12q(q+2)(q+4)]$, $b_T = (A_{T22} - 2A_{T3})/[12q(q+2)]$ e $c_T = (A_{T11} - A_{T22} + A_{T3})/(12q)$, com $A_{T11} = A_{T1} + A_{T1,\beta\phi}$ e $A_{T22} =$

$A_{T2} + A_{T2,\beta\phi}$,

$$A_{T1} = \frac{6c_0}{n}(\rho_{ZZ_2} - \rho_{Z_2Z_2}), \quad A_{T2} = -\frac{3c_0}{n}(\rho_{ZZ} - 2\rho_{ZZ_2} + \rho_{Z_2Z_2}), \quad A_{T3} = 0,$$

$$A_{T1,\beta\phi} = \frac{6c_1}{n}q(p-q) + \frac{6c_2}{n}q, \quad A_{T2,\beta\phi} = -\frac{3c_1}{n}q(q+2),$$

$$c_0 = \frac{\delta_{00010}}{\delta_{20000}^2}, \quad c_1 = -\frac{m_3^2}{m_1}, \quad c_2 = -\frac{m_2m_3 + 2m_1m_3}{m_1^2} - \frac{m_4}{m_1}.$$

Quando ϕ é conhecido, $A_{T1,\beta\phi} = A_{T2,\beta\phi} = 0$.

Prova. Ver Seção 2.8.1.

Como $A_{T3} = 0$, o fator de correção $[1 - (c_T + b_T S_T + a_T S_T^2)]$ se reduz a $[1 - (c_T + b_T S_T)]$ em (2.6) e é chamado de fator de correção tipo-Bartlett para o teste gradiente. A distribuição da estatística modificada S_T^* é melhor aproximada pela distribuição χ_q^2 com erro de aproximação reduzido de ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-2})$.

As quantidades A_{LR} , A_{R1} , A_{R2} , A_{T1} e A_{T2} que definem as correções dos testes são decompostas em duas partes, uma dependendo da distribuição assumida para os dados através dos δ 's e outra dependendo da matriz do modelo \mathbf{X} através de \mathbf{Z} e \mathbf{Z}_2 . As quantidades $A_{LR,\beta\phi}$, $A_{R1,\beta\phi}$, $A_{R2,\beta\phi}$, $A_{T1,\beta\phi}$ e $A_{T2,\beta\phi}$ são as contribuições geradas pelo fato de que o parâmetro ϕ é desconhecido e estimado a partir dos dados. Além disso, essas quantidades dependem do tamanho da amostra, do número de parâmetros da regressão (p), do número de parâmetros testados (q) e da distribuição assumida para os dados através dos δ 's.

Para calcular o valor das estatísticas corrigidas basta especificar os valores dos d 's, b 's e c 's para cada modelo. A seguir, apresentamos essas quantidades para algumas distribuições da classe simétrica.

Normal. $d_0 = 0$, $d_1 = 1$, $d_2 = 1$, $b_0 = 0$, $b_1 = 1$, $b_2 = 0$, $b_3 = 1/2$, $c_0 = 0$, $c_1 = 2$, $c_2 = 0$. Como esperado, $S_R^* = S_T^*$.

t-Student.

$$d_0 = \frac{3(\nu+2)(\nu+3)^2}{2\nu(\nu+1)(\nu+5)(\nu+7)}, \quad d_1 = \frac{(\nu+3)(\nu^3+11\nu^2+20\nu+4)}{\nu(\nu+7)(\nu+5)^2}, \quad d_2 = \frac{(\nu+3)(\nu+2)^2}{\nu(\nu+5)^2},$$

$$\begin{aligned}
 b_0 &= \frac{6(\nu^2 + 4\nu - 1)}{\nu(\nu + 5)(\nu + 7)}, & b_1 &= \frac{(\nu - 1)(\nu + 2)(\nu + 3)}{\nu(\nu + 5)^2}, & b_2 &= -\frac{12(\nu^2 + 3\nu + 2)(\nu + 3)}{\nu(\nu + 7)(\nu + 5)^2}, \\
 b_3 &= \frac{(\nu - 1)^2(\nu + 3)}{2\nu(\nu + 5)^2}, & c_0 &= \frac{6(\nu + 2)(\nu + 3)^2}{\nu(\nu + 1)(\nu + 5)(\nu + 7)}, & c_1 &= \frac{2(\nu + 2)^2(\nu + 3)}{\nu(\nu + 5)}, \\
 c_2 &= -\frac{24(\nu + 2)(\nu + 3)}{\nu(\nu + 7)(\nu + 5)^2}.
 \end{aligned}$$

Logística I. $d_0 = -0.0767, d_1 \approx 1.4706, d_2 \approx 1.3626, b_0 = -0.9035, b_1 \approx 1.7744, b_2 \approx 0.5690, b_3 \approx 1.1552, c_0 = -0.3069, c_1 \approx 2.7253, c_2 \approx 0.2158.$

Logística II. $d_0 = 3/20, d_1 \approx 0.7460, d_2 \approx 0.7867, b_0 = 2/5, b_1 \approx 0.5245, b_2 \approx -0.5835, b_3 \approx 0.1748, c_0 = 3/5, c_1 \approx 1.5735, c_2 \approx -0.0815.$

Exponencial potência. Para $-1 < k < 1/3,$

$$\begin{aligned}
 d_0 &= \frac{k(1-k)\Gamma\left(\frac{1-3k}{2}\right)\Gamma\left(\frac{1+k}{2}\right)}{8\Gamma\left(\frac{3-k}{2}\right)^2}, & d_1 &= d_2 = \frac{1}{1+k}, & b_0 &= 1 - \frac{(1-k)\Gamma\left(\frac{3-3k}{2}\right)\Gamma\left(\frac{1+k}{2}\right)}{2\Gamma\left(\frac{3-k}{2}\right)^2}, \\
 b_1 &= \frac{1-k}{1+k}, & b_2 &= \frac{2k(1-k)}{1+k}, & b_3 &= \frac{(1-k)^2}{2(1+k)}, & c_0 &= \frac{k(1-k)\Gamma\left(\frac{1-3k}{2}\right)\Gamma\left(\frac{1+k}{2}\right)}{8\Gamma\left(\frac{3-k}{2}\right)^2}, \\
 c_1 &= \frac{2}{1+k}, & c_2 &= 0.
 \end{aligned}$$

2.4 Modelos de regressão lineares log-simétricos e testes aperfeiçoados

Sejam $y \sim S(\mu, \phi^2)$ e $t = \exp(y)$. A função densidade de t é dada por

$$\pi(t; \eta, \phi) = \frac{h(\tilde{t}^2)}{t\phi}, \quad t > 0, \quad (2.7)$$

sendo $\tilde{t} = \log\left[(t/\eta)^{\frac{1}{\phi}}\right]$, $\eta = \exp(\mu)$ o parâmetro de escala (mediana), $\phi > 0$ o parâmetro de forma e $\int_0^\infty u^{-1/2}h(u)du = 1$. Essa classe de distribuição gerada pela transformação t foi estudada por [Vanegas e Paula \(2016\)](#) e é denominada classe log-simétrica, denotada por $t \sim LS(\eta, \phi^2)$, cujo suporte é $(0, \infty)$. A classe log-simétrica comporta distribuições com caudas mais leves e caudas mais pesadas do que as da distribuição log-normal e modelos que apresentam bimodalidade. São membros dessa classe de modelos as distribuições log-normal, log-t-Student, log-logística I, log-logística II, log-exponencial-potência entre outras. Para detalhes sobre as propriedades estatísticas e métodos de estimação clássico e bayesiano,

ver Vanegas e Paula (2016).

Se ξ é uma variável aleatória com distribuição log-simétrica padrão, isto é, $\xi \sim LS(1, 1)$ e

$$t = \eta \xi^\phi, \quad (2.8)$$

temos que $t \sim LS(\eta, \phi^2)$, sendo η a mediana de t e $\phi > 0$ o parâmetro de forma. Com essa propriedade, Vanegas e Paula (2015a) propuseram um modelo de regressão para o logaritmo da variável aleatória t modelando simultaneamente a mediana e o parâmetro de assimetria.

Sejam t_1, \dots, t_n variáveis aleatórias independentes com $t_l \sim LS(\eta_l, \phi^2)$ para $l = 1, \dots, n$, $\mathbf{x}_l = (x_{l1}, \dots, x_{lp})^\top$ um vetor de variáveis explicativas associado à l -ésima resposta e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ um vetor de parâmetros desconhecidos a ser estimado. Usando as propriedades das distribuições log-simétricas (Vanegas e Paula, 2016) e a equação (2.8), define-se o modelo de regressão log-simétrico como

$$y_l = \log t_l = \log \eta_l + \phi \log(\xi_l), \quad l = 1, \dots, n, \quad (2.9)$$

em que $\log \eta_l = \mu_l = \mathbf{x}_l^\top \boldsymbol{\beta}$, $\phi > 0$ é desconhecido, porém o mesmo para todas as observações, e $\log(\xi_l) = \epsilon_l \sim S(0, 1)$.

O modelo de regressão em (2.9) é equivalente ao dado em (2.2) para os modelos de regressão lineares simétricos. Portanto, as expressões para as correções de Bartlett para o teste LR dado em (2.4), tipo-Bartlett para o teste score dado em (2.5) e tipo-Bartlett para o teste gradiente dado em (2.6) são válidas para os modelos de regressão lineares log-simétricos definidos em (2.9). Esses resultados permitem testar hipóteses sobre os parâmetros da regressão utilizando os testes não corrigidos (S_W , S_{LR} , S_R e S_T) e os testes corrigidos (S_{LR}^* , S_R^* e S_T^*).

Na seção seguinte apresentamos alguns estudos de simulação de Monte Carlo para avaliar o desempenho do teste baseado na estatística gradiente corrigida e comparar com os testes baseados nas estatísticas S_{LR}^* e S_R^* , bem como com os testes baseados nas estatísticas não corrigidas S_W , S_{LR} , S_R e S_T e corrigidas via bootstrap nos MRLS e nos MRLLS. Os poderes dos testes serão comparados.

2.5 Estudo de simulação

Nessa seção apresentamos estudos de simulação de Monte Carlo para compararmos o desempenho dos testes Wald (S_W), razão de verossimilhanças (S_{LR}), escore (S_R) e gradiente (S_T) e dos testes aperfeiçoados da razão de verossimilhança (S_{LR}^*), escore (S_R^*) e gradiente (S_T^*) em amostras pequenas e moderadas em modelos de regressão lineares simétricos. Também foram considerados as correções bootstrap do teste Wald, LR, escore e gradiente (S_W^b , S_{LR}^b , S_R^b e S_T^b) para efeito de comparação. Consideramos o modelo

$$y_l = \beta_0 + \beta_1 x_{l1} + \cdots + \beta_{p-1} x_{l,p-1} + \phi \epsilon_l, \quad l = 1, \dots, n,$$

em que o parâmetro de escala $\phi > 0$ é desconhecido e o mesmo para todas as observações e os erros ϵ_l são variáveis aleatórias independentes e identicamente distribuídas com distribuição simétrica padrão segundo as distribuições normal, t-Student (com $\nu = 4$) e logística II. As covariáveis foram extraídas de uma distribuição $\mathcal{U}(0, 1)$, todos os parâmetros da regressão, exceto os fixados na hipótese nula, são iguais a um e o parâmetro de escala é fixado em $\phi = 3$. Também foram considerados diferentes valores para o número de parâmetros da regressão (p), número de parâmetros da hipótese nula (q) e diferentes tamanhos de amostra ($n = 20, 25$ e 30).

O número de réplicas de Monte Carlo foi fixado em 15000 e foram considerados os seguintes níveis nominais $\alpha = 10\%, 5\%$ e 1% . Todas as simulações foram realizadas utilizando a linguagem de programação matricial Ox (Doornik, 2013), que é distribuída livremente para fins acadêmicos e disponível em <http://www.doornik.com>. Todas as maximizações da função de log-verossimilhança para os parâmetros do modelo foram realizadas utilizando o método quasi-Newton (BFGS) com primeiras derivadas analíticas através da sub-rotina `MaxBFGS`.

Avaliamos via simulação a probabilidade de erro tipo I da hipótese nula $\mathcal{H}_0 : \beta_1 = \cdots = \beta_q = 0$ dos testes S_W , S_{LR} , S_R , S_T , S_{LR}^* , S_R^* , S_T^* , isto é, a proporção de vezes em que o valor calculado para cada estatística é maior do que o quantil $1 - \alpha$ da distribuição χ_q^2 . Para os testes corrigidos via bootstrap paramétrico, utilizamos as estatísticas não corrigidas (S_W , S_{LR} , S_R e S_T) e os quantis estimados através de 600 réplicas bootstrap para calcular a probabilidade de erro tipo I. Os resultados são apresentados nas Tabelas 2.2 e 2.3 para o

modelo normal, nas Tabelas 2.4 e 2.5 para o modelo t-Student e nas Tabelas 2.6 e 2.7 para o modelo logístico II.

As Tabelas 2.2 a 2.7 indicam que o teste de Wald (S_W) rejeita a hipótese nula com mais frequência do que o esperado com base no nível nominal selecionado. Por exemplo, para $p = 4$, $\alpha = 5\%$ e $n = 20$, as taxas de rejeição da hipótese nula para o modelo normal foram 14.63% (para $q = 3$), 12.39% (para $q = 2$) e 9.87% (para $q = 1$); para o modelo t-Student esses valores foram 18.94% (para $q = 3$), 15.81% (para $q = 2$) e 12.35% (para $q = 1$) e para o modelo logístico II foram 15.35% (para $q = 3$), 13.25% (para $q = 2$) e 10.47% (para $q = 1$) como mostram as Tabelas 2.2, 2.4 e 2.6, respectivamente.

Note também que o teste LR (S_{LR}) apresenta taxas de rejeição acima do nível nominal considerado. Por exemplo, para $p = 4$, $\alpha = 5\%$ e $n = 20$ as taxas de rejeição sob o modelo normal foram 9.51% (para $q = 3$), 9.09% (para $q = 2$) e 8.58% (para $q = 1$), para o modelo t-Student, 10.28% (para $q = 3$), 10.11% (para $q = 2$) e 9.49% (para $q = 1$) e para o modelo logístico II, 9.51% (para $q = 3$), 9.31% (para $q = 2$) e 8.71% (para $q = 1$) de acordo com as Tabelas 2.2, 2.4 e 2.6, respectivamente. Apesar dessas taxas de rejeição serem menores do que as do teste de Wald, quando comparadas com os testes escore e gradiente esses valores são bem maiores. Considerando o mesmo exemplo anterior, as taxas de rejeição do teste escore foram 4.33%, 5.83% e 6.90% no modelo normal, 4.94%, 6.55% e 7.17% no modelo t-Student e 4.91%, 6.39% e 6.97% no modelo logístico II, enquanto que para o teste gradiente esses valores foram 4.33%, 5.83% e 6.89% sob o modelo normal, 4.21%, 6.09% e 7.20% sob o modelo t-Student e 4.46%, 5.94% e 6.93% sob o modelo logístico II.

Os testes escore e gradiente são menos liberais do que os testes Wald e LR. No entanto, as taxas de rejeição da hipótese nula podem desviar-se consideravelmente do nível nominal do teste. Por exemplo, se $p = 6$, $q = 2$, $\alpha = 10\%$ e $n = 20$, a taxa de rejeição sob o modelo normal foi 15.90% para ambos os testes; para o modelo t-Student esses valores foram 15.61% (S_R) e 17.10% (S_T) e para o modelo logístico II foram 15.61% (S_R) e 16.25% (S_T), como mostram as Tabelas 2.3, 2.5 e 2.7, respectivamente. Para todos os modelos considerados, quando o número de parâmetros da regressão está fixado e o número de parâmetros testados decresce, as taxas de rejeição dos testes S_R e S_T tendem a aumentar. Por exemplo, para $p = 6$, $\alpha = 5\%$ e $n = 25$ as taxas de rejeição do modelo t-Student foram 5.58% (S_R) e

5.11% (S_T) (para $q = 3$), 6.86% (S_R) e 6.57% (S_T) (para $q = 2$) e 7.78% (S_R) e 8.11% (S_T) (para $q = 1$) como mostra a Tabela 2.5. Comportamento semelhante foi observado para os modelos normal e logístico II como indicam as Tabelas 2.3 e 2.7, respectivamente. Para os testes Wald e LR, quando o número de parâmetros da regressão está fixado e o número de parâmetros testados diminui, as taxas de rejeição de ambos os testes tendem a diminuir. Como exemplo, considere $p = 6$, $\alpha = 5\%$ e $n = 25$; os valores obtidos para as taxas de rejeição dos testes S_W e S_{LR} no modelo t-Student foram, respectivamente, 23.65% e 12.39% (para $q = 3$), 20.97% e 12.23% (para $q = 2$) e 17.69% e 11.99% (para $q = 1$), como indica a Tabela 2.5. Esse comportamento também foi observado para os modelos normal e logístico II como indicam as Tabelas 2.3 e 2.7.

Por outro lado, os testes S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b e S_T^b apresentaram taxas de rejeição menos distorcidas que os testes Wald, LR, escore e gradiente. Enquanto o número de parâmetros da regressão e o número de parâmetros a serem testados influenciam significativamente nos tamanhos dos testes S_W , S_{LR} , S_R e S_T , nos testes corrigidos essas quantidades têm pouca influência nas taxas de rejeição. Por exemplo, para $p = 4$, $\alpha = 5\%$ e $n = 20$, as taxas de rejeição sob o modelo t-Student foram 4.99% (S_{LR}^*), 4.82% (S_R^*) e 4.68% (S_T^*), 5.31% (S_W^b), 5.14% (S_{LR}^b), 5.14% (S_R^b) e 5.18% (S_T^b) para $q = 3$, 5.21% (S_{LR}^*), 5.29% (S_R^*), 5.07% (S_T^*), 5.18% (S_W^b), 5.37% (S_{LR}^b), 5.34% (S_R^b) e 5.28% (S_T^b) para $q = 2$ e 5.08% (S_{LR}^*), 5.25% (S_R^*), 5.12% (S_T^*), 5.23% (S_W^b), 5.28% (S_{LR}^b), 5.25% (S_R^b) e 5.24% (S_T^b) para $q = 1$, como mostra a Tabela 2.4. Comportamento análogo pode ser observado para os modelos normal e logístico II; ver Tabelas 2.2 e 2.6. Por fim, como esperado, conforme cresce o tamanho da amostra, as taxas de rejeição de todos os testes se aproximam do verdadeiro nível nominal.

Agora, avaliamos numericamente os poderes de todos os testes em amostra finitas. Como nossos estudos de simulação evidenciam, alguns testes apresentam diferentes tamanhos sob \mathcal{H}_0 . Para compararmos numericamente o poder dos testes é necessário que eles apresentem o mesmo tamanho sob a hipótese nula. Para os testes que apresentaram distorções no tamanho, sob a hipótese nula, foram consideradas 500,000 réplicas de Monte Carlo para estimar o valor crítico exato de cada teste de acordo com o nível nominal selecionado. Para $n = 30$, $p = 4$, $q = 3$, $\phi = 3$ e $\alpha = 10\%$, calculamos a taxa de rejeição da hipótese

Tabela 2.2: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo normal.

q	n	$\alpha(\%)$	S_W	S_{LR}	$S_R = S_T$	S_{LR}^*	$S_R^* = S_T^*$	S_W^b	S_{LR}^b	S_R^b	S_T^b	
3	20	10	21,32	16,32	10,72	10,34	10,13	10,35	10,35	10,37	10,37	
		5	14,63	9,51	4,33	5,10	4,63	5,12	5,15	5,15	5,15	
		1	6,12	2,62	0,37	1,09	0,73	1,13	1,17	1,19	1,19	
	25	10	18,94	14,83	10,49	10,22	10,09	10,26	10,27	10,27	10,27	10,27
		5	12,16	8,32	4,65	5,22	4,93	5,33	5,35	5,36	5,36	5,36
		1	4,69	2,25	0,53	1,05	0,79	1,17	1,18	1,19	1,19	1,19
	30	10	16,97	13,60	10,23	9,97	9,85	9,99	10,00	10,01	10,01	10,01
		5	10,50	7,54	4,67	4,99	4,86	5,03	5,04	5,06	5,06	5,06
		1	3,73	1,77	0,49	0,96	0,78	1,03	1,03	1,04	1,04	1,04
2	20	10	19,16	16,11	12,40	10,07	10,22	10,15	10,16	10,16	10,16	
		5	12,39	9,09	5,83	5,00	4,95	5,05	5,07	5,08	5,08	
		1	4,84	2,57	0,81	1,02	0,89	1,13	1,14	1,15	1,15	
	25	10	17,25	14,61	12,12	10,39	10,49	10,46	10,46	10,47	10,47	
		5	10,86	8,41	5,81	5,07	5,06	5,14	5,15	5,16	5,16	
		1	3,89	2,25	0,79	1,00	0,91	1,03	1,05	1,05	1,05	
	30	10	15,65	13,67	11,37	9,78	9,82	9,82	9,83	9,83	9,83	
		5	9,17	7,34	5,40	4,85	4,83	4,93	4,93	4,94	4,94	
		1	3,07	1,79	0,79	0,95	0,88	1,01	1,02	1,03	1,03	
1	20	10	16,23	14,69	13,24	9,98	10,22	10,05	10,05	10,07	10,07	
		5	9,87	8,58	6,90	5,05	5,19	5,15	5,17	5,17	5,17	
		1	3,73	2,43	1,29	1,04	1,02	1,11	1,12	1,14	1,14	
	25	10	14,80	13,75	12,71	10,23	10,35	10,27	10,27	10,29	10,29	
		5	8,82	7,74	6,60	5,23	5,31	5,29	5,31	5,31	5,31	
		1	3,06	2,22	1,31	1,15	1,14	1,24	1,24	1,24	1,24	
	30	10	13,79	12,87	12,02	10,08	10,13	10,11	10,11	10,11	10,11	
		5	8,19	7,27	6,38	5,16	5,23	5,24	5,24	5,24	5,24	
		1	2,49	1,87	1,17	1,03	1,02	1,11	1,11	1,11	1,11	

Tabela 2.3: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo normal.

q	n	$\alpha(\%)$	S_W	S_{LR}	$S_R = S_T$	S_{LR}^*	$S_R^* = S_T^*$	S_W^b	S_{LR}^b	S_R^b	S_T^b
4	20	10	29,52	20,92	11,51	9,67	9,43	9,58	9,61	9,63	9,63
		5	21,04	12,89	4,56	4,80	4,33	4,79	4,79	4,79	4,79
		1	10,33	4,03	0,26	1,07	0,59	1,08	1,09	1,11	1,11
	25	10	24,43	17,93	10,97	9,71	9,56	9,67	9,67	9,67	9,67
		5	16,47	10,44	4,94	5,03	4,75	5,03	5,04	5,05	5,05
		1	7,44	3,17	0,38	0,94	0,64	1,01	1,03	1,06	1,06
	30	10	21,94	16,71	11,19	9,95	9,84	9,95	9,95	9,97	9,97
		5	14,29	9,43	4,97	5,00	4,78	5,02	5,02	5,03	5,03
		1	5,68	2,67	0,54	0,97	0,79	1,01	1,03	1,04	1,04
3	20	10	26,49	20,71	14,17	9,89	10,07	9,87	9,89	9,90	9,90
		5	18,40	12,83	6,61	4,95	4,88	4,95	4,97	4,98	4,98
		1	8,87	4,15	0,75	1,13	0,82	1,17	1,18	1,19	1,19
	25	10	22,61	18,22	13,22	10,17	10,35	10,21	10,21	10,21	10,21
		5	15,33	10,99	6,32	5,07	5,01	5,10	5,10	5,11	5,11
		1	6,34	3,03	0,75	0,99	0,83	1,03	1,05	1,06	1,06
	30	10	19,74	16,27	12,27	9,68	9,79	9,74	9,75	9,75	9,75
		5	12,51	9,21	5,87	4,83	4,83	4,85	4,85	4,85	4,85
		1	4,72	2,51	0,82	0,92	0,84	1,00	1,00	1,01	1,01
2	20	10	23,18	19,77	15,90	9,81	10,35	9,87	9,87	9,89	9,89
		5	15,83	12,26	8,25	4,89	5,11	4,95	4,95	4,95	4,95
		1	7,07	3,91	1,33	1,07	1,00	1,13	1,14	1,14	1,14
	25	10	19,88	17,21	14,38	10,14	10,39	10,18	10,19	10,19	10,19
		5	12,94	10,37	7,29	4,89	5,11	4,99	5,00	5,01	5,01
		1	4,99	2,90	1,24	1,00	0,97	1,07	1,09	1,10	1,10
	30	10	17,65	15,55	13,33	10,03	10,22	10,09	10,09	10,09	10,09
		5	11,19	9,07	6,85	5,07	5,15	5,11	5,11	5,13	5,13
		1	4,17	2,56	1,17	0,95	0,93	1,01	1,02	1,03	1,03

Tabela 2.4: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo t -Student.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	26,57	17,53	11,55	10,67	10,33	10,42	10,06	10,45	10,45	10,64	10,54
		5	18,94	10,28	4,94	4,21	4,99	4,82	4,68	5,31	5,14	5,14	5,18
		1	9,24	2,81	0,52	0,33	0,99	0,77	0,74	1,05	1,07	1,12	1,13
	25	10	22,04	15,19	10,82	10,09	9,79	9,99	9,80	9,97	9,95	10,08	9,97
		5	14,73	8,58	5,01	4,64	5,20	4,95	5,03	5,25	5,33	5,19	5,29
		1	6,64	2,37	0,68	0,45	1,00	0,91	0,78	1,09	1,07	1,14	1,09
	30	10	19,78	14,23	10,58	10,04	9,73	9,86	9,65	9,74	9,89	9,97	9,88
		5	12,69	7,79	4,91	4,54	4,82	4,85	4,87	5,04	4,88	5,01	5,09
		1	5,09	1,85	0,67	0,55	0,87	0,83	0,89	1,12	0,96	0,98	1,03
2	20	10	23,27	17,15	12,73	12,89	10,22	10,21	10,23	10,47	10,47	10,14	10,37
		5	15,81	10,11	6,55	6,09	5,21	5,29	5,07	5,18	5,37	5,34	5,28
		1	7,22	2,84	0,99	0,74	0,92	0,91	0,82	1,04	1,07	1,12	1,11
	25	10	19,63	14,98	11,80	11,83	9,80	10,09	9,95	9,93	9,97	10,08	10,06
		5	12,75	8,51	5,68	5,59	5,00	4,86	4,91	5,19	5,13	4,99	5,09
		1	5,34	2,47	0,99	0,75	1,00	1,01	0,95	1,11	1,14	1,16	1,17
	30	10	18,06	14,26	11,55	11,73	10,03	10,12	10,11	10,01	10,14	10,12	10,17
		5	11,37	7,95	5,69	5,51	4,82	4,93	4,87	5,06	4,99	5,01	5,01
		1	4,09	2,09	1,00	0,87	1,01	0,98	0,94	1,22	1,15	1,09	1,09
1	20	10	18,51	15,79	13,35	13,79	10,20	10,48	10,37	10,37	10,35	10,36	10,39
		5	12,35	9,49	7,17	7,20	5,08	5,25	5,12	5,23	5,28	5,25	5,24
		1	4,97	2,71	1,23	1,19	0,86	0,90	0,89	1,04	1,07	1,09	1,03
	25	10	16,14	13,73	11,98	12,44	9,64	9,75	9,81	9,97	9,75	9,72	9,81
		5	10,05	7,80	6,03	6,39	4,79	4,86	4,83	4,81	4,88	4,87	4,91
		1	3,43	2,05	1,15	1,23	0,96	0,93	0,93	1,11	1,05	1,03	1,05
	30	10	14,93	13,01	11,65	11,99	9,61	9,75	9,76	9,77	9,77	9,74	9,78
		5	8,73	7,24	6,12	6,24	4,92	4,91	4,97	4,87	5,05	4,91	5,02
		1	2,83	1,84	1,31	1,31	1,05	1,05	1,09	1,22	1,11	1,11	1,16

Tabela 2.5: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo t -Student.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
4	20	10	38,77	23,31	11,91	11,51	9,72	9,59	9,20	9,81	9,90	9,75	9,72
		5	30,31	14,41	4,98	4,53	4,76	4,37	4,39	4,87	4,99	4,81	4,94
		1	17,60	4,73	0,44	0,35	0,87	0,61	0,66	1,05	0,96	1,01	1,06
	25	10	31,99	20,18	12,16	11,71	10,03	10,28	9,89	10,01	10,21	10,40	10,27
		5	23,65	12,39	5,58	5,11	5,09	4,91	4,73	5,07	5,22	5,21	5,22
		1	12,33	3,85	0,64	0,47	0,95	0,79	0,77	1,13	1,05	1,11	1,13
	30	10	26,99	17,93	11,69	11,19	9,90	10,06	9,77	10,07	10,05	10,10	10,04
		5	18,99	10,31	5,37	4,90	4,99	4,89	4,67	5,07	5,17	5,02	4,97
		1	8,87	2,90	0,70	0,46	0,90	0,82	0,74	1,10	0,99	1,01	1,05
3	20	10	35,30	23,46	14,09	14,59	9,61	10,23	9,76	9,90	9,95	9,99	9,77
		5	26,91	14,39	6,62	6,41	4,63	4,73	4,62	4,81	4,85	4,87	4,83
		1	14,91	4,66	0,89	0,66	0,78	0,78	0,73	0,93	0,85	1,01	1,02
	25	10	29,04	20,01	13,41	13,60	9,90	10,37	10,15	10,03	10,17	10,33	10,31
		5	20,97	12,23	6,86	6,57	4,80	5,13	4,97	5,09	5,04	5,25	5,21
		1	10,31	3,77	1,03	0,83	0,90	0,90	0,83	1,06	1,07	1,17	1,06
	30	10	24,57	17,77	12,95	13,13	9,97	10,37	9,99	10,03	10,17	10,30	10,05
		5	16,66	10,29	6,39	6,29	4,87	4,95	4,87	5,01	5,03	4,99	5,05
		1	7,50	2,77	1,06	0,85	0,97	0,97	0,85	1,11	1,07	1,11	1,06
2	20	10	30,09	22,26	15,61	17,10	9,86	10,29	10,48	10,28	10,23	9,97	10,23
		5	22,35	14,09	7,77	8,54	4,47	4,91	4,88	4,71	4,81	4,83	4,89
		1	11,71	4,37	1,32	1,20	0,78	0,85	0,73	0,87	0,91	0,98	0,97
	25	10	25,24	19,27	14,87	15,75	10,19	10,71	10,52	10,23	10,53	10,50	10,49
		5	17,69	11,99	7,78	8,11	5,01	5,33	5,19	5,11	5,21	5,33	5,29
		1	8,29	3,55	1,33	1,38	0,94	0,95	1,02	1,05	1,05	1,12	1,13
	30	10	21,99	17,11	13,74	14,46	10,03	10,39	10,17	10,03	10,33	10,25	10,15
		5	14,55	9,97	6,81	7,29	5,05	5,06	5,23	5,23	5,23	5,05	5,26
		1	6,12	2,91	1,31	1,35	0,98	0,94	0,93	1,15	1,09	1,07	1,15

Tabela 2.6: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo logístico II.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	22,69	16,28	10,99	10,42	9,68	9,90	9,83	10,25	10,15	10,11	10,15
		5	15,35	9,51	4,91	4,46	4,90	4,87	4,84	5,13	5,29	5,21	5,29
		1	6,95	2,77	0,39	0,35	0,96	0,77	0,71	1,14	1,11	1,11	1,10
	25	10	19,72	14,79	10,96	10,40	9,73	10,17	9,92	10,01	10,07	10,22	10,06
		5	12,71	8,23	5,09	4,74	5,07	5,06	5,02	5,17	5,26	5,22	5,25
		1	5,08	2,27	0,76	0,52	1,09	1,01	1,02	1,23	1,21	1,23	1,23
	30	10	17,53	13,73	10,63	10,09	9,51	9,96	9,73	9,79	9,81	10,01	9,91
		5	10,96	7,31	4,67	4,51	4,72	4,65	4,74	4,86	4,98	4,79	4,95
		1	3,91	1,71	0,54	0,49	0,77	0,74	0,68	0,93	0,88	0,92	0,91
2	20	10	19,91	16,24	12,75	12,61	9,69	10,24	10,11	10,23	10,15	10,12	10,13
		5	13,25	9,31	6,39	5,94	4,96	5,23	5,10	5,21	5,22	5,29	5,17
		1	5,48	2,83	0,92	0,83	1,02	0,85	0,94	1,27	1,21	1,08	1,21
	25	10	17,59	14,52	12,10	11,88	9,81	10,14	10,19	10,19	10,11	10,11	10,20
		5	11,08	8,29	5,89	5,77	4,83	5,03	5,03	5,22	5,11	5,07	5,15
		1	4,16	2,13	0,95	0,84	0,91	0,93	0,94	1,06	1,03	1,07	1,09
	30	10	15,95	13,25	11,50	11,35	9,49	9,92	9,85	9,87	9,84	9,88	9,87
		5	9,67	7,26	5,57	5,41	4,54	4,91	4,73	4,86	4,81	4,97	4,81
		1	3,20	1,79	0,91	0,79	0,87	0,89	0,85	0,99	0,97	0,98	1,03
1	20	10	16,93	15,00	13,21	13,45	9,69	10,27	10,13	10,04	10,11	10,15	10,11
		5	10,47	8,71	6,97	6,93	4,58	4,95	4,92	4,95	4,93	4,93	4,95
		1	3,74	2,35	1,35	1,35	1,06	1,04	1,08	1,23	1,21	1,19	1,22
	25	10	15,09	13,75	12,50	12,76	9,62	10,16	9,93	9,89	9,93	10,12	9,91
		5	8,96	7,58	6,51	6,52	4,77	4,99	5,00	4,84	5,05	4,99	5,03
		1	2,83	1,88	1,22	1,19	0,94	0,92	0,99	1,08	1,09	0,97	1,06
	30	10	14,18	12,99	11,88	12,09	9,65	9,96	9,86	9,97	9,86	9,91	9,86
		5	8,21	6,94	6,05	6,08	4,65	4,91	4,87	4,82	4,89	4,91	4,87
		1	2,33	1,65	1,05	0,99	0,79	0,86	0,81	0,90	0,91	0,91	0,88

Tabela 2.7: Taxas de rejeição (%) da hipótese $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ para $p = 6$; modelo logístico II.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b	
4	10	10	32,47	21,81	12,08	11,76	9,47	10,12	9,90	10,12	10,01	10,29	10,23	
		20	23,95	13,57	5,03	4,82	4,79	4,61	4,63	5,24	5,26	5,11	5,30	
		1	12,67	4,4	0,39	0,29	0,95	0,67	0,61	1,02	1,13	1,09	1,07	
	25	10	26,35	18,37	11,85	11,79	9,88	9,93	10,22	10,25	10,37	9,95	10,38	
		5	18,58	11,29	5,64	5,09	4,78	4,99	4,79	5,17	5,17	5,26	5,14	
		1	8,87	3,32	0,61	0,46	0,94	0,84	0,78	1,18	1,09	1,15	1,15	
	30	10	23,31	16,75	11,37	10,99	9,55	9,99	9,67	9,83	9,91	10,07	9,78	
		5	15,63	9,65	4,89	4,80	4,57	4,50	4,66	5,07	4,93	4,71	4,94	
		1	6,60	2,61	0,69	0,55	0,93	0,81	0,81	1,09	1,17	1,05	1,14	
	3	10	10	29,27	21,79	14,33	14,27	9,40	10,15	10,15	10,04	10,08	9,94	10,08
			20	21,21	13,19	6,75	6,49	4,59	4,94	4,89	5,05	4,99	5,08	5,11
			1	10,37	4,35	0,81	0,78	0,93	0,81	0,86	1,11	1,13	1,11	1,14
25		10	24,07	18,21	13,29	13,25	9,53	10,11	10,09	10,16	9,94	9,90	10,05	
		5	16,39	10,81	6,69	6,57	4,97	5,17	5,19	5,23	5,31	5,21	5,36	
		1	7,45	3,29	0,98	0,82	0,92	0,91	0,83	1,00	1,11	1,11	1,15	
30		10	21,38	16,36	12,19	12,17	9,05	9,79	9,55	9,62	9,55		9,55	
		5	13,65	9,06	6,07	5,87	4,61	4,87	4,70	4,89	4,79	4,94	4,83	
		1	5,34	2,71	1,08	0,97	1,05	0,99	1,00	1,12	1,17	1,15	1,21	
2		10	10	25,27	20,30	15,61	16,25	9,34	10,52	10,55	10,07	10,19	10,09	10,23
			20	17,69	12,73	8,09	7,98	4,35	5,07	4,89	4,75	4,87	4,95	4,81
			1	7,95	3,95	1,39	1,26	0,87	0,99	0,91	1,10	1,10	1,13	1,01
	25	10	21,42	17,79	14,51	14,71	9,49	10,36	10,13	9,94	9,95	10,08	10,01	
		5	14,11	10,41	7,55	7,45	4,73	5,08	5,11	5,07	5,09	4,94	5,09	
		1	5,70	3,15	1,52	1,39	0,93	0,98	0,97	1,07	1,10	1,14	1,11	
	30	10	19,37	16,19	13,40	13,55	9,53	10,13	9,95	9,81	9,93	9,98	9,88	
		5	12,25	9,17	6,76	6,76	4,54	4,89	4,78	4,86	4,83	4,89	4,79	
		1	4,37	2,57	1,27	1,21	0,87	0,99	0,96	1,04	0,99	1,08	1,03	

alternativa $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$ para os modelos normal, t-Student e logístico II. Os resultados são apresentados na Figura 2.1. Como podemos notar, o poder dos testes são similares e, como esperado, a medida em que $|\delta|$ cresce a função poder tende a 1. Simulações de poder para diferentes valores de n , p , q , ϕ e α apresentaram resultados semelhantes.

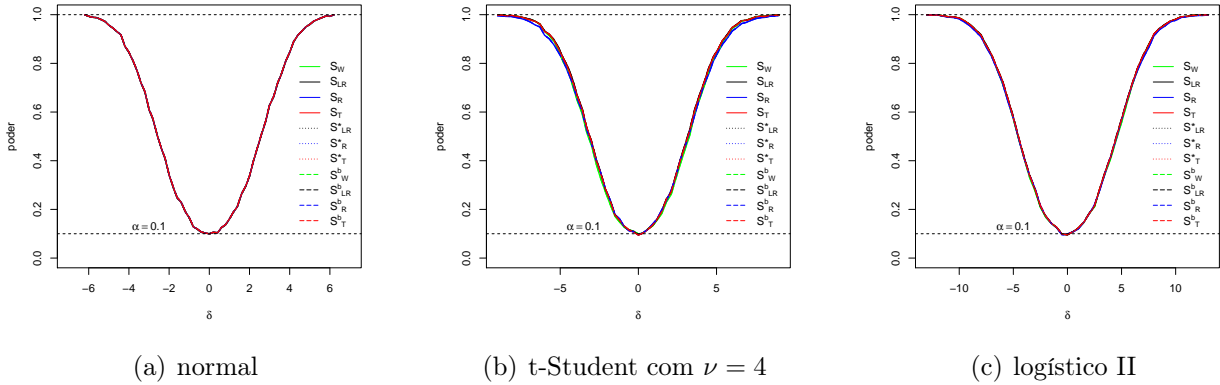


Figura 2.1: Poder dos testes para $n = 30$, $p = 4$, $q = 3$, $\phi = 3$ e $\alpha = 10\%$.

Também foram considerados estudos de simulação para avaliar a probabilidade de erro tipo I e o poder dos testes não corrigidos (S_W , S_{LR} , S_R e S_T) e dos testes corrigidos (S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b e S_T^b) nos modelos de regressão lineares log-simétricos. Notamos um comportamento análogo ao que ocorre para o tamanho e o poder dos testes nos MRLS. Portanto, omitimos os resultados desses estudos de simulação.

Os resultados dos estudos de simulação apresentados nessa seção podem ser resumidos como segue. No geral, em amostras pequenas e moderadas, os testes Wald, LR, score e gradiente são mais liberais, ou seja, rejeitam a hipótese nula (sendo esta verdadeira) com frequência maior que a indicada pelo nível nominal selecionado. De todos os testes avaliados numericamente, o teste Wald é o mais liberal. Por essa razão, não recomendamos a utilização desse teste para testar hipóteses nos MRLS e nos MRLLS quando o tamanho da amostra é pequeno ou moderado. Já os testes corrigidos pelos fatores de correção de Bartlett e tipo-Bartlett corrigem as distorções no tamanho dos testes sem perda de poder. Os testes corrigidos via bootstrap apresentaram desempenho semelhante aos testes corrigidos analiticamente porém necessitam de esforço computacional intensivo. Como não existe um fator de correção tipo-Bartlett para o teste Wald, o teste bootstrap é uma alternativa para corrigir seu comportamento liberal em amostras de tamanho pequeno e moderado. Portanto,

recomendamos os testes modificados (analiticamente ou via bootstrap) para testar hipóteses em MRLS e nos MRLLS quando o tamanho da amostra é pequeno ou moderado.

2.6 Aplicações

Para ilustrar os resultados obtidos nesse capítulo consideramos duas aplicações a dados reais. A primeira considera o conjunto de dados em [Nateghi et al. \(2012\)](#) sobre propriedades de textura em queijo cheddar. Para avaliar o efeito das variáveis regressoras, porcentagem de gordura (x_1), porcentagem de goma xantana (x_{2l}) e porcentagem de caseinato de sódio (x_{3l}), na variável resposta, coesividade (t_l), de $n = 16$ amostras de queijos cheddar, ajustamos o seguinte modelo de regressão linear log-simétrico

$$\begin{aligned} \log(t_l) = & \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \beta_4 x_{1l} x_{2l} + \beta_5 x_{1l} x_{3l} \\ & + \beta_6 x_{2l} x_{3l} + \phi \epsilon_l, \quad l = 1, \dots, 16, \end{aligned} \quad (2.10)$$

em que $\phi > 0$ é desconhecido e o mesmo para todas as observações e $\epsilon_l \sim S(0, 1)$ variáveis aleatórias independentes e identicamente distribuídas com distribuição normal ($\epsilon_l \sim N(0, 1)$), t-Student ($\epsilon_l \sim t(0, 1, \nu)$) e logística II ($\epsilon_l \sim LII(0, 1)$). O critério AIC corrigido para os modelos ajustados foram: $-135,25$ (modelo normal), $-134,87$ (modelo t_3), $-135,02$ (modelo t_4), $-135,03$ (modelo t_5) e $-134,98$ (modelo logístico II). Com os resíduos studentizados propostos por [Villegas et al. \(2013\)](#), construímos o gráfico normal de probabilidades e o envelope simulado para os modelos normal, t-Student com $\nu = 5$ e logístico II a fim de detectar especificações incorretas da distribuição do erro e a presença de observações extremas ([Figura 2.2](#)). Com base no AIC corrigido e na [Figura 2.2](#) selecionamos o modelo de regressão linear log-normal para ajustar os dados. As estimativas pontuais dos parâmetros com seus respectivos erros-padrão assintóticos em parênteses são: $\hat{\beta}_0 = -0,1321 (0,0067)$, $\hat{\beta}_1 = -0,0043 (0,0039)$, $\hat{\beta}_2 = -0,1456 (0,1712)$, $\hat{\beta}_3 = 0,0135 (0,0251)$, $\hat{\beta}_4 = -0,1864 (0,1001)$, $\hat{\beta}_5 = -0,0074 (0,0100)$, $\hat{\beta}_6 = 0,6606 (0,5007)$ e $\hat{\phi} = 0,0011 (0,0002)$.

Para o modelo selecionado, ou seja, o modelo de regressão linear log-normal, é possível determinar um teste exato para testar hipóteses sobre os coeficientes da regressão. Então, do

ponto de vista prático, não há interesse na utilização dos testes (corrigidos ou não corrigidos) que usam p -valores assintóticos ou de bootstrap. Esta aplicação visa comparar os p -valores dos diferentes testes apresentados neste capítulo com os p -valores exatos.

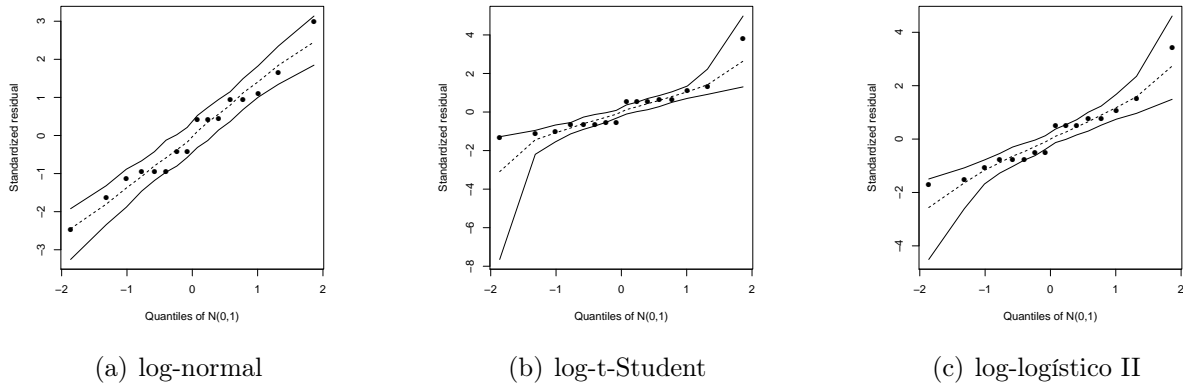


Figura 2.2: Gráfico normal de probabilidades e envelope gerado para as distribuições (a) log-normal, (b) log-t-Student e (c) log-logística II ajustadas ao modelo (2.10).

Inicialmente, testamos as hipóteses $\mathcal{H}_0 : \beta_4 = 0$, $\mathcal{H}_0 : \beta_5 = 0$ e $\mathcal{H}_0 : \beta_6 = 0$, isto é, o efeito individual das interações no modelo (2.10). Os valores das estatísticas de testes (p -valores assintóticos em parênteses) para testar $\mathcal{H}_0 : \beta_4 = 0$ foram: $S_W = 3,4632$ (p -valor 0,0628), $S_{LR} = 3,1350$ (p -valor 0,0766), $S_R = 2,8470$ (p -valor 0,0915), $S_T = 2,8469$ (p -valor 0,0915), $S_{LR}^* = 1,6654$ (p -valor 0,1969), $S_R^* = 1,7657$ (p -valor 0,1839) e $S_T^* = 1,7657$ (p -valor 0,1839). Os p -valores para todos os testes bootstrap foram 0,2036. Além disso, o p -valor exato foi de 0,196. Como podemos notar, os p -valores dos testes corrigidos analiticamente e via bootstrap são bem próximos do p -valor exato. Ao nível de significância de 5% os testes (não corrigidos, corrigidos e exato) não rejeitam a hipótese nula, porém ao nível de significância de 10% os testes Wald, LR, score e gradiente rejeitam a hipótese nula enquanto que os testes S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b , S_T^b e o teste exato não rejeitam \mathcal{H}_0 . As hipóteses $\mathcal{H}_0 : \beta_5 = 0$ e $\mathcal{H}_0 : \beta_6 = 0$ não são rejeitadas por nenhum dos testes aos níveis de significância usuais. Em seguida testamos o efeito conjunto dos fatores de interação, ou seja, testamos a hipótese $\mathcal{H}_0 : \beta_4 = \beta_5 = \beta_6 = 0$. Nenhum dos testes considerados rejeitam a hipótese nula aos níveis de significância usuais.

Como evidenciado pelos nossos estudos de simulações, os testes S_W , S_{LR} , S_R e S_T são mais liberais quando o tamanho da amostra é pequeno (nesse caso $n = 16$). Portanto, com

base nos testes corrigidos e no teste exato, decidimos pela remoção do efeito das interações do modelo (2.10) e estimamos o modelo

$$\log(t_l) = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \phi \epsilon_l, \quad l = 1, \dots, 16, \quad (2.11)$$

sendo $\hat{\beta}_0 = -0,1217(0,0022)$, $\hat{\beta}_1 = -0,0119(0,0009)$, $\hat{\beta}_2 = -0,3989(0,0438)$, $\hat{\beta}_3 = 0,0262(0,0044)$ e $\hat{\phi} = 0,0013(0,0002)$ as estimativas pontuais dos parâmetros com erros-padrão assintóticos em parênteses.

Por fim, testamos as hipóteses $\mathcal{H}_0 : \beta_1 = 0$, $\mathcal{H}_0 : \beta_2 = 0$ e $\mathcal{H}_0 : \beta_3 = 0$ que foram fortemente rejeitadas pelos testes (corrigidos, não corrigidos e exato) aos níveis de significância usuais. Notamos que, para todas as hipóteses testadas nesta aplicação, os p -valores dos testes corrigidos (analiticamente ou por bootstrap) foram bem próximos dos p -valores exatos.

Na Figura 2.3 apresentamos o gráfico normal de probabilidades e o envelope simulado para o modelo (2.11), que sugere um ajuste adequado.

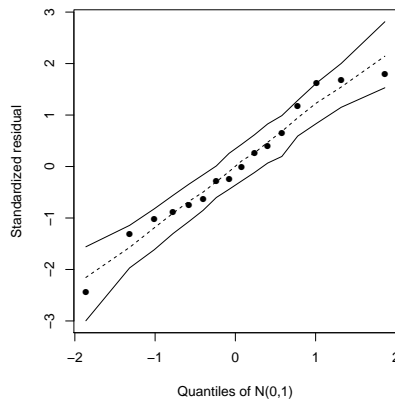


Figura 2.3: Gráfico normal de probabilidades e envelope simulado para o modelo (2.11).

Do modelo (2.11) e das estimativas obtidas, estima-se a coesividade mediana do queijo cheddar por

$$\hat{\eta} = e^{-0,1217 - 0,0119x_1 - 0,3989x_2 + 0,0262x_3}.$$

A segunda aplicação considera o conjunto de dados em [Mirhosseini e Tan \(2010\)](#) sobre

propriedades físico-químicas de emulsão em bebidas de laranja. Para avaliar o efeito das variáveis independentes, quantidade de goma arábica (x_1), quantidade de goma xantana (x_2) e quantidade de óleo de laranja (x_3), todas medidas em grama, na variável resposta, densidade de emulsão (y) dada em g/cm^3 , ajustamos o seguinte modelo de regressão

$$y_l = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \beta_4 x_{1l} x_{2l} + \beta_5 x_{1l} x_{3l} + \beta_6 x_{2l} x_{3l} + \phi \epsilon_l, \quad l = 1, \dots, 20, \quad (2.12)$$

sendo $\phi > 0$ desconhecido e o mesmo para todas as observações e $\epsilon_l \sim S(0, 1)$ erros aleatórios independentes e identicamente distribuídos com distribuição normal ($\epsilon_l \sim N(0, 1)$), t-Student ($\epsilon_l \sim t(0, 1, \nu)$) e logística II ($\epsilon_l \sim LII(0, 1)$). O AIC corrigido para os modelos considerados foram: $-150,03$ (modelo normal), $-158,68$ (modelo t_3), $-156,37$ (modelo t_4), $-154,48$ (modelo t_5) e $-151,03$ (modelo logístico II). Assim como na aplicação anterior, construímos o gráfico normal de probabilidades para os resíduos studentizados e o envelope simulado para os modelos normal, t-Student com $\nu = 3$ e logístico II (Figura 2.4). Com base no AIC corrigido e na Figura 2.4 selecionamos o modelo de regressão linear t-Student com $\nu = 3$ para ajustar os dados. As estimativas pontuais dos parâmetros com seus respectivos erros-padrão assintóticos em parênteses são: $\hat{\beta}_0 = 0,9388 (0,0220)$, $\hat{\beta}_1 = 0,0054 (0,0011)$, $\hat{\beta}_2 = 0,1563 (0,0401)$, $\hat{\beta}_3 = 0,0016 (0,0016)$, $\hat{\beta}_4 = -0,0048 (0,0015)$, $\hat{\beta}_5 = -0,0001 (0,0001)$, $\hat{\beta}_6 = -0,0070 (0,0026)$ e $\hat{\phi} = 0,0012 (0,0003)$.

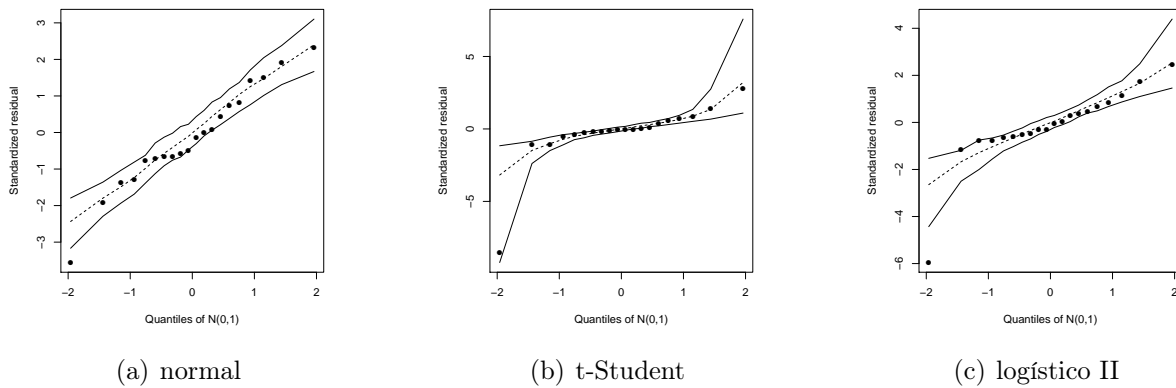


Figura 2.4: Gráfico normal de probabilidades e envelope simulado para as distribuições (a) normal, (b) t-Student e (c) logística II ajustadas ao modelo (2.12).

Inicialmente testamos o efeito individual dos fatores de interação, ou seja, testamos as hipóteses $\mathcal{H}_0 : \beta_4 = 0$, $\mathcal{H}_0 : \beta_5 = 0$ e $\mathcal{H}_0 : \beta_6 = 0$ contra as respectivas hipóteses alternativas bilaterais. Os valores das estatísticas de teste e os p -valores são apresentados na Tabela 2.8. A hipótese $\mathcal{H}_0 : \beta_4 = 0$ é rejeitada pelos testes Wald, LR e gradiente ao nível de significância de 5%, porém a decisão se altera quando utilizamos o teste escore e os testes corrigidos. A hipótese $\mathcal{H}_0 : \beta_5 = 0$ não é rejeitada por nenhum dos testes aos níveis de significância usuais. Já a hipótese $\mathcal{H}_0 : \beta_6 = 0$ é rejeitada pelos testes Wald e LR ao nível de significância de 5%, mas não é rejeitada pelos demais testes. Em seguida testamos o efeito conjunto dos fatores de interação no modelo (2.12), isto é, testamos a hipótese $\mathcal{H}_0 : \beta_4 = \beta_5 = \beta_6 = 0$. Os valores das estatísticas de testes foram: $S_W = 17,9297$ (p -valor 0,0005), $S_{LR} = 8,1531$ (p -valor 0,0430), $S_R = 2,9646$ (p -valor 0,3971), $S_T = 4,5884$ (p -valor 0,2045), $S_{LR}^* = 4,0333$ (p -valor 0,2579), $S_R^* = 2,1259$ (p -valor 0,5467) e $S_T^* = 3,1251$ (p -valor 0,3727). O p -valor dos testes S_W^b , S_{LR}^b , S_R^b e S_T^b são, respectivamente, 0,1838, 0,2280, 0,5868 e 0,3466. Neste caso, quando empregamos os testes Wald e LR a hipótese nula é rejeitada ao nível de significância de 5%, mas essa decisão é alterada quando os demais testes são utilizados.

Tabela 2.8: Valor da estatística de teste e p -valor para as hipóteses $\mathcal{H}_0 : \beta_4 = 0$, $\mathcal{H}_0 : \beta_5 = 0$ e $\mathcal{H}_0 : \beta_6 = 0$ no modelo (2.12).

estatísticas de teste	$\mathcal{H}_0 : \beta_4 = 0$		$\mathcal{H}_0 : \beta_5 = 0$		$\mathcal{H}_0 : \beta_6 = 0$	
	valor observado	p -valor	valor observado	p -valor	valor observado	p -valor
S_W	10,2240	0,0014	0,4583	0,4984	7,2474	0,0071
S_{LR}	6,5050	0,0108	0,5354	0,4644	5,2526	0,0219
S_R	3,5812	0,0584	0,6040	0,4371	3,3333	0,0679
S_T	4,1713	0,0411	0,5148	0,4731	3,5959	0,0579
S_{LR}^*	2,5065	0,1134	0,2063	0,6497	2,0239	0,1548
S_R^*	2,2753	0,1314	0,3500	0,5541	2,1023	0,1471
S_T^*	2,1510	0,1425	0,2066	0,6494	1,7897	0,1810
S_W^b		0,1054		0,6728		0,1532
S_{LR}^b		0,0930		0,6242		0,1266
S_R^b		0,1562		0,5506		0,1690
S_T^b		0,1318		0,6194		0,1650

Como evidenciado pelos nossos experimentos de simulações, os testes não corrigidos (Wald, LR, escore e gradiente) são mais liberais quando o tamanho da amostra é pequeno, nesse caso $n = 20$. Portanto, devemos ser cautelosos quanto às conclusões utilizando esses testes.

Como sugerido pelos testes corrigidos decidimos pela remoção dos efeitos de interações

do modelo (2.12) e estimamos

$$y_l = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \phi \epsilon_l, \quad (2.13)$$

sendo $\hat{\beta}_0 = 1.0198 (0.0051)$, $\hat{\beta}_1 = 0.0027 (0.0002)$, $\hat{\beta}_2 = -0.0058 (0.0059)$, $\hat{\beta}_3 = -0.0023 (0.0003)$ e $\hat{\phi} = 0.0018 (0.0004)$ as estimativas pontuais dos parâmetros (erros-padrão assintóticos em parênteses). Aos níveis de significância usuais as hipóteses $\mathcal{H}_0 : \beta_1 = 0$ e $\mathcal{H}_0 : \beta_3 = 0$ foram fortemente rejeitadas pelos testes (não corrigidos e corrigidos). Todos os testes também sugerem a exclusão da variável independente x_2 do modelo (2.13). Assim, o modelo final é

$$y_l = \beta_0 + \beta_1 x_{1l} + \beta_3 x_{3l} + \phi \epsilon_l. \quad (2.14)$$

As estimativas pontuais dos parâmetros são: $\hat{\beta}_0 = 1,0168 (0,0047)$, $\hat{\beta}_1 = 0,0027 (0,0002)$, $\hat{\beta}_3 = -0,0023 (0,0003)$ e $\hat{\phi} = 0,0018 (0,0004)$. A Figura 2.5 apresenta o gráfico normal de probabilidades dos resíduos studentizados com o envelope simulado para o modelo (2.14), que sugere um ajuste adequado.

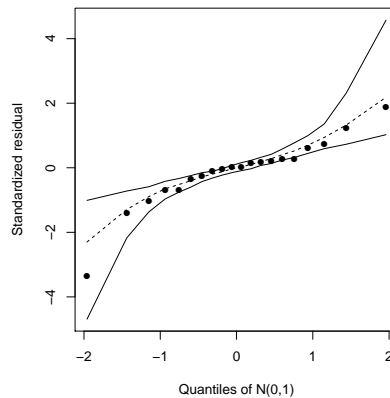


Figura 2.5: Gráfico normal de probabilidades e envelope simulado para o modelo (2.14).

2.7 Considerações finais

Neste capítulo, a partir dos resultados gerais de Vargas et al. (2013), derivamos um fator de correção tipo-Bartlett para o teste gradiente em modelos de regressão lineares simétricos e as correções de Bartlett para o teste LR e tipo-Bartlett para os testes escore e gradiente

nos modelos de regressão lineares log-simétricos. Além disso, comparamos numericamente o tamanho do teste gradiente corrigido com o tamanho dos testes Wald, LR, escore, gradiente e as versões corrigidas analiticamente e via bootstrap. Também comparamos o poder de cada teste via simulação. Além disso, apresentamos duas aplicações a conjuntos de dados reais para ilustrar os resultados obtidos.

Nossos estudos de simulação sugerem que os testes corrigidos apresentam probabilidade de erro tipo I mais próxima do nível nominal considerado do que a probabilidade do erro tipo I dos testes Wald, LR, escore e gradiente. Quanto ao poder, todos os testes apresentaram resultados similares. Além disso, os testes Wald e LR são mais liberais que os testes escore e gradiente porém, quando o número de parâmetros da regressão é fixado e o número de parâmetros a serem testados decresce, o tamanho desses testes podem apresentar consideráveis distorções. Portanto, em conjuntos de dados reais em que o tamanho da amostra é pequeno ou moderado, recomendamos a utilização dos testes corrigidos para os modelos de regressão lineares simétricos e log-simétricos.

2.8 Detalhes técnicos

Neste apêndice calculamos as quantidades A_{T1} , A_{T2} , A_{T3} , $A_{T1,\beta\phi}$, $A_{T2,\beta\phi}$, $A_{T3,\beta\phi}$ e os valores dos δ 's que determinam o teste gradiente aperfeiçoado para as hipóteses $\mathcal{H}_0 : \beta_1 = \beta_{10}$ versus $\mathcal{H}_1 : \beta_1 \neq \beta_{10}$ em modelos de regressão lineares simétricos.

2.8.1 Prova da Proposição 1

Sejam $\kappa_{rs} = E(\partial^2 \ell / \partial \beta_r \partial \beta_s)$, $\kappa_{rst} = E(\partial^3 \ell / \partial \beta_r \partial \beta_s \partial \beta_t)$, $\kappa_{rstu} = E(\partial^4 \ell / \partial \beta_r \partial \beta_s \partial \beta_t \partial \beta_u)$, $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \beta_t$, $\kappa_{rs}^{(tu)} = \partial^2 \kappa_{rs} / \partial \beta_t \partial \beta_u$, $\kappa_{\phi\phi} = E(\partial^2 \ell / \partial \phi^2)$, $\kappa_{r\phi} = E(\partial^2 \ell / \partial \beta_r \partial \phi)$, $\kappa_{r\phi}^{(s)} = \partial^2 \kappa_{r\phi} / \partial \beta_s$ e assim por diante. Os índices r , s , t e u variam de 1 a p . Para os MRLS os cumulantes são dados por

$$\begin{aligned} \kappa_{rs} &= \frac{\delta_{01000}}{\phi^2} \sum_{l=1}^n x_{lr} x_{ls}, & \kappa_{rstu} &= \frac{\delta_{00010}}{\phi^4} \sum_{l=1}^n x_{lr} x_{ls} x_{lu} x_{lt}, \\ \kappa_{rst} &= \kappa_{rs}^{(t)} = \kappa_{rst}^{(u)} = \kappa_{rs}^{(tu)} = 0, & \kappa_{\phi\phi} &= -\frac{n}{\phi^2} (\delta_{20002} - 1), \end{aligned}$$

$$\begin{aligned}
 \kappa_{\phi\phi}^{(\phi)} &= -\frac{2n}{\phi^3}(\delta_{001002} - 1), & \kappa_{\phi\phi\phi} &= -\frac{n}{\phi^3}(6\delta_{001002} + \delta_{001003} - 4), \\
 \kappa_{rs\phi} &= -\frac{1}{\phi^3}(\delta_{000101} + 2\delta_{01000}) \sum_{l=1}^n x_{lr}x_{ls}, & \kappa_{rs\phi\phi} &= -\frac{1}{\phi^4}(\delta_{000012} - 6\delta_{110001}) \sum_{l=1}^n x_{lr}x_{ls}, \\
 \kappa_{rs\phi}^{(\phi)} &= \frac{3}{\phi^4}(\delta_{000101} + 2\delta_{01000}) \sum_{l=1}^n x_{lr}x_{ls}, & \kappa_{rs}^{(\phi)} &= -\frac{2\delta_{01000}}{\phi^3} \sum_{l=1}^n x_{lr}x_{ls}, \\
 \kappa_{r\phi\phi} &= \kappa_{r\phi\phi}^{(u)} = 0;
 \end{aligned} \tag{2.15}$$

ver Ferrari e Uribe-Opazo (2001), Uribe-Opazo et al. (2008) e Uribe-Opazo (1997).

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$ um vetor de parâmetros desconhecidos de dimensão $p + 1$ em que $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ e $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. Consideramos o problema de testar a hipótese nula composta $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ contra a hipótese alternativa $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$ sendo $(\boldsymbol{\beta}_2^\top, \phi)^\top$ um vetor de parâmetros de perturbação. Definimos, $\mathcal{A} = \text{diag}\{\mathbf{0}, \mathbf{K}_{\boldsymbol{\beta}_{22}}^{-1}\}$, $\mathcal{M}_{\boldsymbol{\beta}} = \mathbf{K}_{\boldsymbol{\beta}}^{-1} - \mathcal{A}_{\boldsymbol{\beta}}$, $\mathcal{A}_{\boldsymbol{\beta}\phi} = \text{diag}\{\mathcal{A}_{\boldsymbol{\beta}}, -\kappa_{\phi\phi}^{-1}\}$ e $\mathcal{M}_{\boldsymbol{\beta}\phi} = \text{diag}\{\mathcal{M}_{\boldsymbol{\beta}}, \mathbf{0}\}$. Sejam $m^{r\phi}$ e $a^{r\phi}$ os $(r, p + 1)$ -ésimos elementos das matrizes $\mathcal{M}_{\boldsymbol{\beta}\phi}$ e $\mathcal{A}_{\boldsymbol{\beta}\phi}$, respectivamente. Analogamente, $m^{\phi\phi}$ e $a^{\phi\phi}$ representam os $(p + 1, p + 1)$ -ésimos elementos de $\mathcal{M}_{\boldsymbol{\beta}\phi}$ e $\mathcal{A}_{\boldsymbol{\beta}\phi}$. Assim, $m^{r\phi} = m^{\phi r} = m^{\phi\phi} = a^{r\phi} = a^{\phi r} = 0$ ($r = 1, \dots, p$), com $a^{\phi\phi} = -\kappa_{\phi\phi}^{-1}$.

Os coeficientes A_T 's que definem uma correção tipo-Bartlett para o teste gradiente nos MRLS quando ϕ é desconhecido são obtidos substituindo os cumulantes dados em (2.15) nas fórmulas de A_{T11} , A_{T22} e A_{T33} dadas no Teorema 1 de Vargas et al. (2013). Como a matriz informação de Fisher é bloco diagonal e $A_{T3, \boldsymbol{\beta}\phi} = 0$ podemos decompor os coeficientes A 's como $A_{T11} = A_{T1} + A_{T1, \boldsymbol{\beta}\phi}$, $A_{T22} = A_{T2} + A_{T2, \boldsymbol{\beta}\phi}$ e $A_{T33} = A_{T3}$, sendo A_{T1} , A_{T2} e A_{T3} as quantidades obtidas supondo que o parâmetro ϕ é conhecido e $A_{T1, \boldsymbol{\beta}\phi}$ e $A_{T2, \boldsymbol{\beta}\phi}$ as quantidades obtidas supondo que o parâmetro ϕ é desconhecido. Segue do Corolário 2 de Vargas et al. (2013) que

$$S_T^* = S_T [1 - (c_T + b_T S_T + a_T S_T^2)],$$

sendo a_T , b_T e c_T da ordem $O(n^{-1})$ e dados por $a_T = A_{T3}/[12q(q + 2)(q + 4)]$, $b_T = (A_{T22} - 2A_{T3})/[12q(q + 2)]$ e $c_T = (A_{T11} - A_{T22} + A_{T3})/(12q)$, com $A_{T11} = A_{T1} + A_{T1, \boldsymbol{\beta}\phi}$ e $A_{T22} = A_{T2} + A_{T2, \boldsymbol{\beta}\phi}$.

Substituindo os cumulantes dados em (2.15) na fórmula de A_{T11} no Teorema 1 de Vargas et al. (2013) o coeficiente A_{T1} se reduz a

$$A_{T1} = 6 \sum' \kappa_{jr su} m^{jr} a^{su} = \frac{6\delta_{00010}}{\phi^4} \sum' \sum_{l=1}^n x_{lr} x_{lj} x_{ls} x_{lu} m^{jr} a^{su}.$$

Invertendo a ordem dos somatórios e rearranjando os termos, temos

$$A_{T1} = 6 \sum' \kappa_{jr su} m^{jr} a^{su} = \frac{6\delta_{00010}}{\phi^4} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \left(\sum' x_{ls} a^{su} x_{lu} \right).$$

Os termos $\sum_{i,j=1}^p x_{li} a^{ij} x_{lj}$ e $\sum_{i,j=1}^p x_{li} m^{ij} x_{lj}$ representam elementos das matrizes $\phi^2 \mathbf{Z}_2 / \delta_{20000}$ e $\phi^2 (\mathbf{Z} - \mathbf{Z}_2) / \delta_{20000}$, respectivamente. Logo,

$$\begin{aligned} A_{T1} &= 6 \sum' \kappa_{jr su} m^{jr} a^{su} = \frac{6\delta_{00010}}{\phi^4} \sum_{l=1}^n \left(\frac{\phi^2}{\delta_{20000}} (z_{lu} - z_{2lu}) \right) \left(\frac{\phi^2}{\delta_{20000}} z_{2lu} \right) \\ &= \frac{6\delta_{00010}}{\delta_{20000}^2} \sum_{l=1}^n (z_{lu} - z_{2lu}) z_{2lu} \\ &= \frac{6c_0}{n} (\rho_{ZZ_2} - \rho_{Z_2 Z_2}), \end{aligned}$$

sendo $c_0 = \delta_{00010} / \delta_{20000}^2$, $\rho_{ZZ_2} = n \text{tr}(\mathbf{Z}_d \mathbf{Z}_{2d})$ e $\rho_{Z_2 Z_2} = n \text{tr}(\mathbf{Z}_{2d} \mathbf{Z}_{2d})$.

De maneira análoga obtemos $A_{T3} = 0$ e

$$\begin{aligned} A_{T2} &= -3 \sum' \kappa_{jr su} m^{jr} m^{su} = -\frac{3\delta_{00010}}{\phi^4} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \left(\sum' x_{ls} m^{su} x_{lu} \right) \\ &= -\frac{3\delta_{00010}}{\phi^4} \sum_{l=1}^n \left(\frac{\phi^2}{\delta_{20000}} (z_{lu} - z_{2lu}) \right) \left(\frac{\phi^2}{\delta_{20000}} (z_{lu} - z_{2lu}) \right) \\ &= -\frac{3\delta_{00010}}{\delta_{20000}^2} \sum_{l=1}^n (z_{lu} - z_{2lu})^2 \\ &= -\frac{3c_0}{n} (\rho_{ZZ} - 2\rho_{ZZ_2} + \rho_{Z_2 Z_2}), \end{aligned}$$

sendo $\rho_{ZZ} = n \text{tr}(\mathbf{Z}_d \mathbf{Z}_d)$.

Para derivação dos termos $A_{T1, \beta\phi}$ e $A_{T2, \beta\phi}$, já que $A_{T3, \beta\phi} = 0$, segue do Teorema 1 de Vargas et al. (2013) que

$$\begin{aligned}
 A_{T1,\beta\phi} = & \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{jr} m^{su} + \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{js} m^{ru} \\
 & + \sum' \kappa_{jrs} \kappa_{u\phi\phi} a^{\phi\phi} m^{ju} m^{rs} + 6 \sum' \kappa_{jr\phi} \kappa_{\phi\phi\phi} m^{jr} (a^{\phi\phi})^2 \\
 & + 6 \sum' \kappa_{jr\phi} \kappa_{vw\phi} m^{jr} a^{vw} a^{\phi\phi} + 6 \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{jr} a^{su} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{uvw} m^{ju} a^{vw} a^{\phi\phi} + 3 \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{ju} a^{rs} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{u\phi\phi} m^{ju} (a^{\phi\phi})^2 + 3 \sum' \kappa_{jr\phi} \kappa_{uw\phi} m^{ju} a^{rw} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{uv\phi} m^{ju} a^{sv} a^{\phi\phi} + 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} a^{rv} a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi} \kappa_{uw\phi} m^{ju} a^{sw} a^{\phi\phi} - 6 \sum' (\kappa_{jr\phi}^{(\phi)} - \kappa_{jr\phi\phi}) m^{jr} a^{\phi\phi} \\
 & - 6 \sum' \kappa_{jr\phi}^{(\phi)} m^{jr} a^{\phi\phi} - 12 \sum' \kappa_{j\phi\phi}^{(u)} m^{ju} a^{\phi\phi} - 12 \sum' \kappa_{\phi\phi}^{(\phi)} \kappa_{jr\phi} m^{jr} (a^{\phi\phi})^2 \\
 & - 12 \sum' \kappa_{kl}^{(\phi)} \kappa_{jr\phi} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) a^{\phi\phi} - 12 \sum' \kappa_{j\phi\phi} \kappa_{kl}^{(u)} (\kappa^{jk} \kappa^{lu} - a^{jk} a^{lu}) a^{\phi\phi},
 \end{aligned} \tag{2.16}$$

e

$$\begin{aligned}
 A_{T2,\beta\phi} = & - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{jr} m^{su} a^{\phi\phi} - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{js} m^{ru} a^{\phi\phi} \\
 & - \sum' \kappa_{jrs} \kappa_{u\phi\phi} m^{ju} m^{rs} a^{\phi\phi} - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{jr} m^{uv} a^{\phi\phi} \\
 & - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} m^{rv} a^{\phi\phi} - 3 \sum' \kappa_{jr\phi} \kappa_{uv\phi} m^{ju} m^{ru} a^{\phi\phi}.
 \end{aligned} \tag{2.17}$$

Substituindo os cumulantes dados em (2.15) nas expressões (2.16) e (2.17) alguns termos se anulam. Invertendo a ordem dos somatórios e rearranjando os termos obtemos

$$\begin{aligned}
 A_{T1,\beta\phi} = & \frac{6(\delta_{00101} + 2\delta_{01000})(6\delta_{01002} + \delta_{00103} - 4)}{n\phi^2(\delta_{20002} - 1)^2} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \\
 & + \frac{6(\delta_{00101} + 2\delta_{01000})^2}{n\phi^4(\delta_{20002} - 1)} \sum_{l,i=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \left(\sum' x_{iw} a^{vw} x_{iw} \right) \\
 & - \frac{36(\delta_{00101} + 2\delta_{01000})}{n\phi^2(\delta_{20002} - 1)} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) + \frac{6(\delta_{00012} - 6\delta_{11001})}{n\phi^2(\delta_{20002} - 1)} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \\
 & - \frac{24\delta_{01000}(\delta_{00101} + 2\delta_{01000})}{n\phi^4(\delta_{20002} - 1)} \sum_{l=1}^n \left(- \sum' x_{lj} k^{jk} x_{ik} \right) \left(- \sum' x_{is} k^{sr} x_{lr} \right) \\
 & + \frac{24\delta_{01000}(\delta_{00101} + 2\delta_{01000})}{n\phi^4(\delta_{20002} - 1)} \sum_{l=1}^n \left(\sum' x_{lj} a^{jk} x_{ik} \right) \left(\sum' x_{is} a^{sr} x_{lr} \right) \\
 & - \frac{24(\delta_{01002} - 1)(\delta_{00101} + 2\delta_{01000})}{n\phi^2(\delta_{20002} - 1)^2} \sum_{l=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right)
 \end{aligned}$$

e

$$A_{T_2, \beta\phi} = -\frac{3(\delta_{00101} + 2\delta_{01000})^2}{n\phi^4(\delta_{20002} - 1)} \sum_{l,i=1}^n \left(\sum' x_{lj} m^{jr} x_{lr} \right) \left(\sum' x_{iu} m^{uv} x_{iv} \right) \\ - \frac{6(\delta_{00101} + 2\delta_{01000})^2}{n\phi^4(\delta_{20002} - 1)} \sum_{l,i=1}^n \left(\sum' x_{lj} m^{ju} x_{iu} \right) \left(\sum' x_{lr} a^{rv} x_{iv} \right).$$

O termo $\sum_{i,j=1}^p x_{li} \kappa^{ij} x_{lj}$ representa o elemento (i, j) da matriz $\phi^2 \mathbf{Z} / \delta_{20000}$ e como

$$\sum' z_{ul} = \text{tr}(Z_d) = p, \quad \sum' z_{2ul} = \text{tr}(Z_2) = p - q, \quad \sum' (z_{ul} - z_{2ul}) = \text{tr}(Z_d - Z_{2d}) = q,$$

segue que

$$A_{T_1, \beta\phi} = \frac{6(\delta_{00101} + 2\delta_{01000})(6\delta_{01002} + \delta_{00103} - 4)}{n\delta_{20000}(\delta_{20002} - 1)^2} q + \frac{6(\delta_{00101} + 2\delta_{01000})^2}{n\delta_{20000}^2(\delta_{20002} - 1)} q(p - q) \\ - \frac{36(\delta_{00101} + 2\delta_{01000})}{\delta_{20000}(\delta_{20002} - 1)} q + \frac{6(\delta_{00012} - 6\delta_{11001})}{n\delta_{20000}(\delta_{20002} - 1)} q - \frac{24\delta_{01000}(\delta_{00101} + 2\delta_{01000})}{n\delta_{20000}^2(\delta_{20002} - 1)} p \\ + \frac{24\delta_{01000}(\delta_{00101} + 2\delta_{01000})}{n\delta_{20000}^2(\delta_{20002} - 1)} (p - q) - \frac{24(\delta_{01002} - 1)(\delta_{00101} + 2\delta_{01000})}{n\delta_{20000}^2(\delta_{20002} - 1)^2} q$$

e

$$A_{T_2, \beta\phi} = -\frac{3(\delta_{00101} + 2\delta_{01000})^2}{n\delta_{20000}^2(\delta_{20002} - 1)} (q^2 + 2q).$$

Fazendo,

$$m_1 = \delta_{01002} - 1, \quad m_2 = 4 - \delta_{00103} - 6\delta_{01002},$$

$$m_3 = \frac{\delta_{00101} + 2\delta_{01000}}{\delta_{20000}}, \quad m_4 = \frac{\delta_{00012} - 6\delta_{11001}}{\delta_{20000}}$$

e após alguma álgebra temos

$$A_{T_1, \beta\phi} = \frac{6q}{n} \left(-\frac{m_3^2}{m_1} (p - q) - \frac{m_2 m_3 + 2m_1 m_3}{m_1^2} - \frac{m_4}{m_1} \right)$$

e

$$A_{\Gamma 2, \beta \phi} = \frac{3 m_3^2}{n m_1} q(q+2).$$

Por fim, considerando

$$c_1 = -\frac{m_3^2}{m_1} \quad \text{e} \quad c_2 = -\frac{m_2 m_3 + 2m_1 m_3}{m_1^2} - \frac{m_4}{m_1}$$

segue o resultado.

2.8.2 Cálculo dos δ 's

Neste apêndice derivamos os valores de $\delta_{abcde} = E(g^{(1)a} g^{(2)b} g^{(3)c} g^{(4)d} z^e)$, para $a, b, c, d, e \in \{0, 1, 2, 3, 4\}$ para algumas distribuições da classe dos modelos simétricos sendo $g^{(r)} = d^r g(z)/dz^r$, $z \sim S(0, 1)$ e $g(z) = \log h(z^2)$.

Distribuição normal. Aqui,

$$g(z) = \log h(z^2) = -\log \sqrt{2\pi} - \frac{z^2}{2}$$

e conseqüentemente

$$g^{(1)} = -z, \quad g^{(2)} = -1, \quad g^{(3)} = g^{(4)} = 0.$$

Como $z \sim N(0, 1)$ então

$$E(z^r) = \begin{cases} 0, & \text{se } r \text{ ímpar} \\ \frac{r!}{2^{r/2}(r/2)!}, & \text{se } r \text{ par} \end{cases}$$

e segue que

$$\begin{aligned} \delta_{00010} &= 0, & \delta_{20000} &= 1, & \delta_{01002} &= -1, & \delta_{00103} &= 0, \\ \delta_{00101} &= 0, & \delta_{01000} &= -1, & \delta_{00012} &= 0, & \delta_{11001} &= 1, \\ \delta_{21000} &= -1, & \delta_{20002} &= 3, & \delta_{30001} &= -3, & \delta_{40002} &= 15, \\ \delta_{21002} &= -3. \end{aligned}$$

Distribuição t-Student. Nesse caso,

$$g(z) = \log h(z^2) = \log \left(\frac{\nu^{\nu/2}}{B(1/2, \nu/2)} \right) - \left(\frac{\nu+1}{2} \right) \log(\nu + z^2)$$

e conseqüentemente

$$\begin{aligned} g^{(1)} &= 2zu, & g^{(2)} &= 2u + \frac{8z^2 u^2}{\nu+1}, & g^{(3)} &= \frac{24}{\nu+1} zu^2 + \frac{64}{(\nu+1)^2} z^3 u^3, \\ g^{(4)} &= \frac{24}{\nu+1} u^2 + \frac{384}{(\nu+1)^2} z^3 u^3 + \frac{768}{(\nu+1)^3} z^4 u^4, \end{aligned}$$

sendo $u = -(\nu+1)/(2(\nu+z^2))$. Como $z \sim t(0, 1, \nu)$ então $z^2 \sim F(1, \nu)$ e

$$E(z^{2r} u^m) = (-1)^m \left(\frac{\nu+1}{2} \right)^m \frac{1}{\nu^{m-r}} \frac{B(\frac{1+2r}{2}, \frac{\nu+2(m-r)}{2})}{B(1/2, \nu/2)}.$$

Assim,

$$\begin{aligned} \delta_{00010} &= \frac{6(\nu+1)(\nu+2)}{\nu(\nu+5)(\nu+7)}, & \delta_{20000} &= \frac{\nu+1}{\nu+3}, \\ \delta_{01002} &= \frac{3-\nu}{\nu+3}, & \delta_{00103} &= \frac{6(3\nu-5)}{(\nu+3)(\nu+5)}, \\ \delta_{00101} &= \frac{6(\nu+1)}{(\nu+3)(\nu+5)}, & \delta_{01000} &= -\frac{\nu+1}{\nu+3}, \\ \delta_{00012} &= \frac{6(\nu^2-12\nu-13)}{(\nu+3)(\nu+5)(\nu+7)}, & \delta_{11001} &= \frac{(\nu+1)(\nu-1)}{(\nu+3)(\nu+5)}, \\ \delta_{21000} &= \frac{(\nu+1)^3(\nu+2)}{\nu(\nu+3)(\nu+5)(\nu+7)}, & \delta_{20002} &= \frac{3(\nu+1)}{(\nu+3)}, \\ \delta_{30001} &= -\frac{3(\nu+1)^2}{(\nu+3)(\nu+5)}, & \delta_{40002} &= \frac{15(\nu+1)^3}{(\nu+3)(\nu+5)(\nu+7)}, \\ \delta_{21002} &= \frac{3(\nu+1)^2(3-\nu)}{(\nu+3)(\nu+5)(\nu+7)}. \end{aligned}$$

Distribuição logística II. Nesse caso,

$$g(z) = \log h(z^2) = \log \left(\frac{e^z}{(1 + e^z)^2} \right)$$

e conseqüentemente

$$g^{(1)} = \frac{1 - e^z}{1 + e^z}, \quad g^{(2)} = -\frac{2e^z}{(1 + e^z)^2}, \quad g^{(3)} = \frac{2e^z(e^z - 1)}{(1 + e^z)^3}, \quad g^{(4)} = -\frac{2e^z(1 - 4e^z + e^{2z})}{(1 + e^z)^4}.$$

Fazendo $u = e^z/(1 + e^z)$ temos que $u \sim \text{Beta}(1, 1)$. Como $z = \log(u/(1 - u))$ então a função geradora de momento de z é dada por

$$M_z(t) = E(e^{tz}) = E \left[\left(\frac{u}{1 - u} \right)^t \right] = B(1 + t, 1 - t),$$

logo,

$$E \left[u^r \left(\log \left(\frac{u}{1 - u} \right) \right)^s \right] = B^{(s)}(1 + r + t, 1 - t)|_{t=0},$$

para $r = 0, 1, 2, 3, 4$ e $s = 1, 2, 3, 4$, sendo $B^{(s)}(\cdot, \cdot)$ a derivada de ordem s da função Beta.

Utilizando esses resultados obtemos

$$\begin{aligned} \delta_{00010} &= 1/15, & \delta_{20000} &= 1/3, & \delta_{01002} &\approx -0.43, & \delta_{00103} &\approx 0.6450, \\ \delta_{00101} &= 1/6, & \delta_{01000} &= -1/3, & \delta_{00012} &= 0.1140, & \delta_{11001} &= 1/6, \\ \delta_{21000} &= -1/15, & \delta_{20002} &\approx 2.43, & \delta_{30001} &= -2/3, & \delta_{40002} &\approx 1.9913, \\ \delta_{21002} &\approx -0.2193. \end{aligned}$$

Distribuição logística I. Aqui,

$$g(z) = \log h(z^2) = \log c - \log \left(\frac{e^{-z^2}}{(1 + e^{-z^2})^2} \right)$$

com $c \approx 1.4843$. Conseqüentemente

$$\begin{aligned} g^{(1)} &= 2zu, & g^{(2)} &= 2u + 2z^2(u^2 - 1), & g^{(3)} &= 6z(u^2 - 1) + 4z^3u(u^2 - 1), \\ g^{(4)} &= 6(u^2 - 1) + 24z^2u(u^2 - 1) + 4z^4(u^2 - 1)(3u^2 - 1) \end{aligned}$$

sendo $u = (e^{-z^2} - 1)/(1 + e^{-z^2})$. Como $z \sim LI(0, 1)$ então

$$E(z^{2r}u^m) = (-1)^m \frac{c}{2} \int_0^1 \left(\log \left(\frac{1+s}{1-s} \right) \right)^{\frac{2r-1}{2}} s^m ds$$

e segue que

$$\begin{aligned} \delta_{00010} &\approx -0.6669, & \delta_{20000} &\approx 1.4772, & \delta_{01002} &\approx -2.0130, & \delta_{00103} &\approx -0.5065, \\ \delta_{00101} &\approx -1.2792, & \delta_{01000} &\approx -1.4772, & \delta_{00012} &\approx 2.6593, & \delta_{11001} &\approx 2.7564, \\ \delta_{21000} &\approx -4.1538, & \delta_{20002} &\approx 4.0130, & \delta_{30001} &\approx -6.9901, & \delta_{40002} &\approx 46.7658, \\ \delta_{21002} &\approx -10.9285. \end{aligned}$$

Distribuição exponencial potência. Nesse caso,

$$g(z) = \log h(z^2) = -\frac{1}{2} |z|^{2/(1+k)}$$

e para $z \neq 0$

$$\begin{aligned} g^{(1)} &= -\frac{1}{1+k} |z|^{\frac{1-k}{1+k}} \text{sinal}(z), & g^{(2)} &= -\frac{1-k}{(1+k)^2} |z|^{-\frac{2k}{1+k}}, \\ g^{(3)} &= \frac{2k(1-k)}{(1+k)^3} |z|^{-\frac{1+3k}{1+k}} \text{sinal}(z), & g^{(4)} &= -\frac{2k(1-k)(1+3k)}{(1+k)^4} |z|^{-\frac{2(1+2k)}{1+k}}. \end{aligned}$$

Como $z \sim EP(0, 1, k)$ então

$$E(z^{2r}) = \frac{2^{r(1+k)} \Gamma((1+k)(1+2r)/2)}{\Gamma((1+k)/2)}, \quad \text{para } r > -1/2.$$

Logo, para $-1 < k < 1/3$

$$\begin{aligned} \delta_{00010} &= \frac{2^{1-2k} k(1-k) \Gamma\left(\frac{1-3k}{2}\right)}{(1+k)^4 \Gamma\left(\frac{1+k}{2}\right)}, & \delta_{20000} &= \frac{2^{1-k} \Gamma\left(\frac{3-k}{2}\right)}{(1+k)^2 \Gamma\left(\frac{1+k}{2}\right)}, \\ \delta_{01002} &= -\frac{1-k}{1+k}, & \delta_{00103} &= \frac{2k(1-k)}{(1+k)^2}, \\ \delta_{00101} &= \frac{2^{2-k} k \Gamma\left(\frac{3-k}{2}\right)}{(1+k)^3 \Gamma\left(\frac{1+k}{2}\right)}, & \delta_{01000} &= -\frac{2^{1-k} \Gamma\left(\frac{3-k}{2}\right)}{(1+k)^2 \Gamma\left(\frac{1+k}{2}\right)}, \end{aligned}$$

$$\delta_{00012} = -\frac{2^{2-k}k(1+3k)\Gamma\left(\frac{3-k}{2}\right)}{(1+k)^4\Gamma\left(\frac{1+k}{2}\right)}, \quad \delta_{11001} = \frac{2^{1-k}(1-k)\Gamma\left(\frac{3-k}{2}\right)}{(1+k)^3\Gamma\left(\frac{1+k}{2}\right)},$$

$$\delta_{21000} = -\frac{2^{1-2k}(1-k)\Gamma\left(\frac{3-3k}{2}\right)}{(1+k)^4\Gamma\left(\frac{1+k}{2}\right)}, \quad \delta_{20002} = \frac{3+k}{1+k},$$

$$\delta_{30001} = -\frac{2^{2-k}\Gamma\left(\frac{5-k}{2}\right)}{(1+k)^3\Gamma\left(\frac{1+k}{2}\right)}, \quad \delta_{40002} = \frac{2^{3-k}\Gamma\left(\frac{7-k}{2}\right)}{(1+k)^4\Gamma\left(\frac{1+k}{2}\right)},$$

$$\delta_{21002} = \frac{2^{2-k}(k-1)\Gamma\left(\frac{5-k}{2}\right)}{(1+k)^4\Gamma\left(\frac{1+k}{2}\right)}.$$

Capítulo 3

Aprimoramento de testes de hipóteses em modelos de dispersão

Resumo

Neste capítulo derivamos um fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de dispersão. Essa correção reduz o erro de aproximação da distribuição assintótica da estatística de teste da ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-3/2})$. Por essa razão, em amostras pequenas ou moderadas, espera-se que o teste gradiente que usa a estatística gradiente corrigida apresente probabilidade de erro tipo I menos distorcida. Estudos de simulação Monte Carlo foram realizados para comparar o tamanho do teste proposto com os tamanhos dos testes de Wald, razão de verossimilhanças, escore e gradiente, as versões corrigidas pelos fatores de correção de Bartlett e tipo-Bartlett dos testes da razão de verossimilhanças e escore, e testes via bootstrap. Além disso, avaliamos numericamente o poder dos testes. Os estudos de simulação sugerem que os testes corrigidos apresentam probabilidade de erro tipo I mais próxima do nível nominal sem perda de poder. Para fins ilustrativos, apresentamos duas aplicações a dados reais.

Palavras-chave: Correção de Bartlett, correção tipo-Bartlett, estatística da razão de verossimilhanças, estatística gradiente, estatística escore, modelos de dispersão, estatística Wald.

3.1 Introdução

Em geral, em amostras finitas, a distribuição exata das estatísticas de Wald, razão de verossimilhanças (LR), escore e gradiente é desconhecida. Assim, os testes são baseados em aproximações assintóticas. Sob algumas condições de regularidade, essas estatísticas são assintoticamente equivalentes e, em particular, têm a mesma distribuição limite χ^2 com erro de aproximação de ordem $\mathcal{O}(n^{-1})$ sob a hipótese nula. Além disso, também apresentam as mesmas propriedades assintóticas sob hipóteses alternativas de Pitman até primeira ordem. Porém, em amostras pequenas ou moderadas, a aproximação pela distribuição χ^2 para as distribuições dessas estatísticas pode não ser adequada, uma vez que o tamanho dos testes que usam essas estatísticas pode ser bastante distorcido.

Recentemente, [Vargas et al. \(2013\)](#) derivaram uma expansão assintótica para a função de distribuição acumulada da estatística gradiente para testar hipóteses simples ou compostas. Usando essa expansão, os autores propuseram um fator de correção tipo-Bartlett para a estatística gradiente. Essa correção melhora a aproximação da distribuição assintótica da estatística de teste pela distribuição χ^2 de referência. Similarmente, [Cordeiro e Ferrari \(1991\)](#) propuseram um fator de correção tipo-Bartlett para a estatística escore que envolve um polinômio de segundo grau na própria estatística escore, enquanto que [Lawley \(1956\)](#), generalizando os resultados derivados inicialmente por [Bartlett \(1937\)](#), propôs um fator de correção de Bartlett para a estatística LR. É importante ressaltar que não existe um fator de correção tipo-Bartlett para a estatística de Wald em generalidade.

Correções de Bartlett e tipo-Bartlett são amplamente utilizadas para melhorar a aproximação da verdadeira distribuição da estatística de teste pela distribuição qui-quadrado de referência em vários modelos paramétricos. [Lemonte et al. \(2016\)](#) derivaram fatores de correção tipo-Bartlett para a estatística escore na classe de modelos Birnbaum–Saunders não linear. [Ferrari e Uribe-Opazo \(2001\)](#) e [Uribe-Opazo et al. \(2008\)](#) derivaram, respectivamente, um fator de correção de Bartlett para a estatística LR e tipo-Bartlett para a estatística escore na classe dos modelos de regressão lineares simétricos. Nessa mesma classe de modelos, obtivemos no [Capítulo 2](#) um fator de correção tipo-Bartlett para a estatística gradiente e verificamos que esses resultados também são válidos para a classe de modelos de

regressão lineares log-simétricos. Em modelos lineares generalizados, [Cordeiro \(1983, 1987\)](#) obteve um fator de correção de Bartlett para a estatística LR, [Cribari e Ferrari \(1995\)](#) e [Cordeiro et al. \(1993\)](#) obtiveram um fator de correção tipo-Bartlett para a estatística escore, enquanto que [Vargas et al. \(2014\)](#) derivaram um fator de correção tipo-Bartlett para a estatística gradiente. A partir de estudos de simulação Monte Carlo, [Lemonte e Ferrari \(2012b\)](#) verificaram que a probabilidade de erro tipo I do teste gradiente, assim como a probabilidade de erro tipo I dos testes de Wald, LR e escore se distancia do verdadeiro nível nominal na classe dos modelos de dispersão em amostras pequenas e moderadas. Dessa forma, a obtenção de um fator de correção tipo-Bartlett para a estatística gradiente nesta classe de modelos para corrigir possíveis distorções na probabilidade de erro tipo I se torna necessária.

A classe dos modelos de dispersão, estudada em detalhes por [Jørgensen \(1997\)](#), possibilitou que uma grande variedade de conjuntos de dados não-normais pudessem ser analisados ampliando as opções para suposição da distribuição da variável resposta e flexibilizando a relação funcional entre a média e o preditor linear. Essa classe de modelos inclui distribuições da família exponencial como, por exemplo, os modelos lineares generalizados (MLG) e distribuições fora da família exponencial como, por exemplo, a distribuição von Mises, que é utilizada para modelar dados circulares ([Fisher 1993](#); [Lee 2010](#); [Mardia e Jupp 2000](#)). Alguns estudos recentes na classe dos modelos de dispersão são: [Lemonte e Ferrari \(2012b\)](#) derivaram uma expansão assintótica para a distribuição das estatísticas de Wald, LR, escore e gradiente sob uma sequência de hipóteses alternativas de Pitman e verificaram que nenhum dos quatro testes é uniformemente mais poderoso que os demais; [Simas et al. \(2012\)](#) estudaram o comportamento assintótico da função de densidade de probabilidade dos modelos de dispersão; [Simas et al. \(2011\)](#) corrigiram o viés dos estimadores de máxima verossimilhança em modelos de dispersão considerando o parâmetro de dispersão variável; e [Rocha et al. \(2010\)](#) derivaram expressões para a matriz de covariância de segunda ordem dos estimadores de máxima verossimilhança.

Neste capítulo derivamos um fator de correção tipo-Bartlett para a estatística gradiente em modelos de dispersão e, adicionalmente, revisitamos os resultados de [Cordeiro et al. \(1994\)](#) e [Ferrari et al. \(2001\)](#). Nossos resultados generalizam os obtidos por [Vargas et al.](#)

(2014), uma vez que os MLGs são um caso particular dos modelos de dispersão. Além disso, por meio de simulações Monte Carlo, avaliamos e comparamos numericamente a probabilidade de erro tipo I e o poder dos testes usuais (Wald, LR, escore e gradiente) e dos testes LR, escore e gradiente modificados. Também consideramos os testes de Wald, LR, escore e gradiente via bootstrap para efeito de comparação. Nossos estudos de simulação sugerem que a probabilidade de erro tipo I do teste gradiente corrigido é bem próxima do nível nominal considerado e que os poderes de todos os testes são similares; ou seja, o teste gradiente que usa a estatística de teste corrigida (derivada nesse capítulo) além de apresentar probabilidades de erro tipo I muito próximas do nível nominal, também possui poder similar aos demais testes.

Este capítulo está organizado como segue. Na Seção 3.2 apresentamos a classe dos modelos de dispersão e aspectos inferenciais. Na Seção 3.3 apresentamos os fatores de correção de Bartlett para a estatística LR e tipo-Bartlett para a estatística escore e derivamos um fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de dispersão. Na Seção 3.4 apresentamos estudos de simulação Monte Carlo para avaliar e comparar o tamanho e o poder de todos os testes. Na Seção 3.5 apresentamos duas aplicações a dados reais. Na Seção 3.6 apresentamos as considerações finais. Detalhes técnicos são apresentados na Seção 3.7.

3.2 Modelos de dispersão, estimação e testes

Seja y uma variável aleatória contínua com função densidade de probabilidade dada por

$$\pi(y; \theta, \phi) = \exp[\phi t(y, \theta) + a(y, \phi)], \quad (3.1)$$

em que $t(\cdot, \cdot)$ e $a(\cdot, \cdot)$ são funções conhecidas, $\theta \in \mathbb{R}$ é usualmente o parâmetro de interesse (que não necessariamente é a média de y) e $\phi > 0$ é o parâmetro de precisão. A classe de modelos definida em (3.1) é chamada de modelos de dispersão e foi estudada em detalhes por Jørgensen (1997). Algumas distribuições pertencentes a essa classe são apresentadas na Tabela 3.1. Os MLGs, obtidos de (3.1) considerando $t(y, \theta) = y\theta - b(\theta)$, formam uma subclasse

importante dos modelos de dispersão, em que $E(y) = \mu = db(\theta)/d\theta$. Outra subclasse de (3.1) é obtida tomando $a(y, \phi) = d_1(\phi) + d(y)$, sendo $d_1(\cdot)$ e $d(\cdot)$ funções conhecidas. Essa subclasse é denominada de modelos próprios de dispersão e, por exemplo, as distribuições von Mises, log-gama e simplex são membros dessa subclasse dos modelos de dispersão. Os modelos próprios de dispersão apresentam as seguintes propriedades: a distribuição da estatística $t = t(y, \theta)$ não depende do parâmetro θ , ou seja, t é uma quantidade pivotal para θ quando ϕ é conhecido; supondo que o parâmetro θ é conhecido, os modelos próprios de dispersão pertencem à família exponencial sendo t uma estatística canônica.

Tabela 3.1: Expressões de $t(y, \theta)$ e $a(y, \theta)$ para alguns modelos de dispersão.^a

modelo	$t(y, \theta)$	$a(y, \phi)$
normal	$-\frac{1}{2}(y - \theta)^2$	$-\frac{1}{2} \log(2\pi\phi^{-1})$
normal inverso	$y\theta + (-2\theta)^{1/2}$	$-\frac{1}{2}(\log(2\pi y^3 \phi^{-1}) + \phi/y)$
gama	$y\theta + \log(-\theta)$	$-\log \Gamma(\phi) + \phi \log(\phi y) - \log y$
log-gama	$y - \theta - \exp(y - \theta)$	$\phi \log \phi - \log \Gamma(\phi)$
von Mises	$\cos(y - \theta)$	$\log I_0(\phi)$

^a $I_\nu(\phi)$ é a função modificada de Bessel de primeiro tipo e ordem ν ; e $\Gamma(\cdot)$ é a função gama.

Considere y_1, \dots, y_n variáveis aleatórias independentes com y_l tendo função densidade dada por (3.1) com $\theta = \theta_l$, para $l = 1, \dots, n$. Com o objetivo de introduzir uma estrutura de regressão à classe de modelos em (3.1), assumimos que

$$g(\theta_l) = \eta_l = \mathbf{x}_l^\top \boldsymbol{\beta}, \quad l = 1, \dots, n, \quad (3.2)$$

em que $\mathbf{x}_l = (x_{l1}, \dots, x_{lp})^\top$ é um vetor de variáveis explicativas associadas à l -ésima resposta, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a ser estimado e $g(\cdot)$ é uma função contínua, inversível e duas vezes diferenciável, denominada de função de ligação. Assumimos que o parâmetro de precisão é desconhecido porém o mesmo para todas as observações e que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ é uma matriz $n \times p$ de posto completo (posto(\mathbf{X}) = p). Assumimos também que as condições de regularidade usuais para estimação de máxima verossimilhança e inferência em grandes amostras são satisfeitas (Cox e Hinkley 1974, Cap. 9). A função de log-verossimilhança para o vetor $(\boldsymbol{\beta}^\top, \phi)^\top$ é dada por

$$\ell = \ell(\boldsymbol{\beta}, \phi) = \phi \sum_{l=1}^n t(y_l, \theta_l) + \sum_{l=1}^n a(y_l, \theta_l), \quad (3.3)$$

em que $\theta_l = g^{-1}(\mathbf{x}_l^\top \boldsymbol{\beta})$. Sejam $t_l^{(i)} = t^{(i)}(y_l, \theta_l) = \partial^i t(y_l, \theta_l) / \partial \theta_l^i$ e $D_{il} = D_{il}(\theta_l, \phi) = E[t^{(i)}(y_l, \theta_l)]$ para $i = 1, 2, 3, 4$ e $l = 1, \dots, n$. Segue de condições de regularidade que $D_{1l} = 0$ e $D_{2l} = -\phi E[t^{(1)}(y_l, \theta_l)^2]$. Na Tabela 3.2 apresentamos os valores de D_{2l} , D_{3l} e D_{4l} para algumas distribuições da classe dos modelos de dispersão.

Tabela 3.2: Expressões de D_{2l} , D_{3l} e D_{4l} para alguns modelos de dispersão.

modelo	D_{2l}	D_{3l}	D_{4l}
normal	-1	0	0
normal inverso	$-(-2\theta_l)^{-3/2}$	$-3(-2\theta_l)^{-5/2}$	$-15(-2\theta_l)^{-7/2}$
gama	$-\frac{1}{\theta_l^2}$	$\frac{2}{\theta_l^3}$	$-\frac{6}{\theta_l^4}$
log-gama	-1	1	-1
von Mises	$-\frac{I_1(\phi)}{I_0(\phi)}$	0	$\frac{I_1(\phi)}{I_0(\phi)}$

O vetor escore e a matriz informação de Fisher para $\boldsymbol{\beta}$ são dados, respectivamente, por $\mathbf{U}_\beta = \phi \mathbf{X}^\top \mathbf{t}^{(1)}$ e $\mathbf{K}_\beta = \phi \mathbf{X}^\top \mathbf{W} \mathbf{X}$, em que $\mathbf{t}^{(1)} = (t_1^{(1)}, \dots, t_n^{(1)})^\top$ e $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ com $w_l = -D_{2l}(\text{d}\theta_l/\text{d}\eta_l)^2$. O estimador de máxima verossimilhança de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, obtido resolvendo a equação $\mathbf{U}_\beta = \mathbf{0}$, não pode ser determinado explicitamente, exceto para o modelo normal. Utilizando o método scoring de Fisher podemos obter o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ resolvendo iterativamente a equação

$$\mathbf{X}^{(m)\top} \mathbf{W}^{(m)} \mathbf{X}^{(m)} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^{(m)\top} \mathbf{W}^{(m)} \mathbf{y}^{*(m)}, \quad m = 0, 1, \dots,$$

em que $\mathbf{y}^{*(m)} = \mathbf{X}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{N}^{(m)} \mathbf{t}^{(1)(m)}$ com $\mathbf{N} = -\text{diag}\{D_{21}^{-1} \text{d}\eta_1/\text{d}\theta_1, \dots, D_{2n}^{-1} \text{d}\eta_n/\text{d}\theta_n\}$. O estimador $\widehat{\boldsymbol{\beta}}$ depende da distribuição assumida para variável resposta através de D_{2l} , no entanto, não depende do parâmetro ϕ . O estimador de máxima verossimilhança de ϕ , $\widehat{\phi}$, é obtido resolvendo a equação

$$\sum_{l=1}^n [t(y_l, \widehat{\theta}_l) + a^{(1)}(y_l, \widehat{\phi})] = 0, \quad (3.4)$$

em que $a^{(1)}(y_l, \hat{\phi}) = da(y_l, \theta_l)/d\phi$. É possível mostrar que $E[\partial^2 \ell(\boldsymbol{\beta}, \phi)/\partial \boldsymbol{\beta} \partial \phi] = \mathbf{0}$ e, portanto, a matriz de informação de Fisher é bloco diagonal, ou seja, os parâmetros $\boldsymbol{\beta}$ e ϕ são globalmente ortogonais. Em particular, para os modelos próprios de dispersão, a equação (3.4) se reduz a $d_1^{(1)}(\hat{\phi}) = -\sum_{l=1}^n t(y_l, \hat{\theta}_l)/n$, sendo $d_1^{(1)}(\phi) = dd_1(\phi)/d\phi$, e a matriz de informação de Fisher de $(\boldsymbol{\beta}^\top, \phi)^\top$ é dada por $\mathbf{K} = \text{diag}\{\phi \mathbf{X}^\top \mathbf{W} \mathbf{X}, -nd^2 d_1(\phi)/d\phi^2\}$.

Considere o problema de testar a hipótese nula composta $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ contra a hipótese alternativa bilateral $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$, em que o vetor de parâmetros desconhecidos $\boldsymbol{\beta}$ é particionado como $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, sendo $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ o vetor de parâmetros de interesse e $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. Aqui, $\boldsymbol{\beta}_{10}$ é um vetor especificado de constantes de dimensão q e $\boldsymbol{\beta}_2$ e ϕ são parâmetros de perturbação. A partição do vetor $\boldsymbol{\beta}$ induz às partições: $\mathbf{U}_\beta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top)^\top$ com $\mathbf{U}_{\beta_1} = \phi \mathbf{X}_1^\top \mathbf{t}^{(1)}$ e $\mathbf{U}_{\beta_2} = \phi \mathbf{X}_2^\top \mathbf{t}^{(1)}$,

$$\mathbf{K}_\beta = \begin{bmatrix} \mathbf{K}_{\beta_{11}} & \mathbf{K}_{\beta_{12}} \\ \mathbf{K}_{\beta_{21}} & \mathbf{K}_{\beta_{22}} \end{bmatrix} = \phi \begin{bmatrix} \mathbf{X}_1^\top \mathbf{W} \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{W} \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{W} \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2 \end{bmatrix},$$

em que a matriz \mathbf{X} é particionada como $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$, sendo \mathbf{X}_1 uma matriz $n \times q$ e \mathbf{X}_2 uma matriz $n \times (p - q)$. Para testar $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$, as estatísticas de Wald, LR, escore e gradiente são dadas, respectivamente, por

$$S_W = \hat{\phi}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^\top \hat{\mathbf{R}}^\top \hat{\mathbf{W}} \hat{\mathbf{R}}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}),$$

$$S_{LR} = 2[\ell(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\phi}) - \ell(\boldsymbol{\beta}_{10}, \tilde{\boldsymbol{\beta}}_2, \tilde{\phi})],$$

$$S_R = \tilde{\mathbf{s}}^\top \tilde{\mathbf{W}}^{1/2} \tilde{\mathbf{X}}_1 (\tilde{\mathbf{R}}^\top \tilde{\mathbf{W}} \tilde{\mathbf{R}})^{-1} \tilde{\mathbf{X}}_1^\top \tilde{\mathbf{W}}^{1/2} \tilde{\mathbf{s}},$$

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

em que $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{W} \mathbf{X}_1$, $\mathbf{s} = (s_1, \dots, s_n)^\top$ com $s_l = \phi^{1/2} t_l^{(1)} (-D_{2l})^{-1/2}$, $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\phi})$ é o estimador de máxima verossimilhança irrestrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$ e $(\boldsymbol{\beta}_{10}, \tilde{\boldsymbol{\beta}}_2, \tilde{\phi})$ é o estimador de máxima verossimilhança restrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$ obtido supondo a hipótese nula verdadeira. A notação “chapéu” e “til” é utilizada para indicar que as quantidades são avaliadas nos estimadores de máxima verossimilhança irrestrito e restrito de $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$,

respectivamente. Sob a hipótese nula, a distribuição limite de todas as estatísticas de teste é χ_q^2 com erro de aproximação de ordem $\mathcal{O}(n^{-1})$; veja, por exemplo, [Sen and Singer \(1993\)](#) e [Lemonte \(2016\)](#).

3.3 Estatísticas de teste aprimoradas

Daqui em diante, assumimos que $a(y, \phi) = d_1(\phi) + d(y)$ na equação (3.1). Os fatores de correção de Bartlett para a estatística LR e tipo-Bartlett para a estatística escore obtidos, respectivamente, por [Cordeiro et al. \(1994\)](#) e [Ferrari et al. \(2001\)](#) na classe dos modelos próprios de dispersão serão revisitados nessa seção. Além disso, utilizando os resultados em [Vargas et al. \(2013\)](#), derivamos um fator de correção tipo-Bartlett para a estatística gradiente de forma a diminuir distorções no tamanho do teste em amostras pequenas e moderadas. Esses resultados são novos e representam contribuições adicionais para melhorar inferências nos modelos próprios de dispersão.

Considere as matrizes: $\mathbf{Z} = \mathbf{X}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top = ((z_{lc}))$, $\mathbf{Z}_2 = \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top = ((z_{2lc}))$ quando $q < p$ e $\mathbf{Z}_2 = \mathbf{0}$ quando $q = p$, $\mathbf{Z}_d = \text{diag}\{z_{11}, \dots, z_{nn}\}$, $\mathbf{Z}_{2d} = \text{diag}\{z_{211}, \dots, z_{2nn}\}$, $\mathbf{F} = \text{diag}\{f_1, \dots, f_n\}$, $\mathbf{G} = \text{diag}\{g_1, \dots, g_n\}$, $\mathbf{E} = \text{diag}\{e_1, \dots, e_n\}$, $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$, $\mathbf{J} = \text{diag}\{j_1, \dots, j_n\}$, $\mathbf{R} = \text{diag}\{r_1, \dots, r_n\}$, $\mathbf{H} = \text{diag}\{h_1, \dots, h_n\}$, $\mathbf{S} = \text{diag}\{s_1, \dots, s_n\}$, $\mathbf{Q} = \text{diag}\{q_1, \dots, q_n\}$, em que

$$f_l = -\frac{d\theta_l}{d\eta_l} \frac{d^2\theta_l}{d\eta_l^2} D_{2l} - \left(\frac{d\theta_l}{d\eta_l}\right)^3 D_{3l}, \quad g_l = -\frac{d\theta_l}{d\eta_l} \frac{d^2\theta_l}{d\eta_l^2} D_{2l}, \quad e_l = -\left(\frac{d\theta_l}{d\eta_l}\right)^3 D'_{2l},$$

$$w_l = -\left(\frac{d\theta_l}{d\eta_l}\right)^2 D_{2l}, \quad j_l = \left(\frac{d\theta_l}{d\eta_l}\right)^4 \left(-3D_{4l} + 8D'_{3l} - 6D''_{2l} + 3\phi D_{2l}^{(2)} - 3\phi D_{2l}^2\right),$$

$$r_l = \left(\frac{d\theta_l}{d\eta_l}\right)^4 D''_{2l} + 5 \left(\frac{d\theta_l}{d\eta_l}\right)^2 \frac{d^2\theta_l}{d\eta_l^2} D'_{2l} + 2 \left[\left(\frac{d^2\theta_l}{d\eta_l^2}\right)^2 + \frac{d\theta_l}{d\eta_l} \frac{d^3\theta_l}{d\eta_l^3} \right] D_{2l},$$

$$h_l = \left(\frac{d\theta_l}{d\eta_l}\right)^2 \frac{d^2\theta_l}{d\eta_l^2} \left(2D_{3l} - 3D'_{2l}\right) + \left(\frac{d\theta_l}{d\eta_l}\right)^4 \left(-2D_{4l} + 6D'_{3l} - 5D''_{2l} + 2\phi D_{2l}^{(2)} - 2\phi D_{2l}^2\right),$$

$$s_l = \left(\frac{d\theta_l}{d\eta_l}\right)^4 D'_{3l} + 3 \left(\frac{d\theta_l}{d\eta_l}\right)^2 \frac{d^2\theta_l}{d\eta_l^2} D'_{2l} + 3 \left[\left(\frac{d^2\theta_l}{d\eta_l^2}\right)^2 + \frac{d\theta_l}{d\eta_l} \frac{d^3\theta_l}{d\eta_l^3} \right] D_{2l} + 3 \left(\frac{d\theta_l}{d\eta_l}\right)^2 \frac{d^2\theta_l}{d\eta_l^2} D_{3l},$$

$$q_l = \left(\frac{d\theta_l}{d\eta_l}\right)^4 D_{4l} + 6 \left(\frac{d\theta_l}{d\eta_l}\right)^2 \frac{d^2\theta_l}{d\eta_l^2} D_{3l} + \left[3 \left(\frac{d^2\theta_l}{d\eta_l^2}\right)^2 + 4 \frac{d\theta_l}{d\eta_l} \frac{d^3\theta_l}{d\eta_l^3} \right] D_{2l},$$

em que $D'_{il} = \partial D_{il}/\partial\theta_l$, $D''_{il} = \partial D'_{il}/\partial\theta_l$, e $D_{il}^{(j)} = E[(\partial^i t(y_l, \theta_l)/\partial\theta_l^i)^j]$ para $i, j = 1, 2, 3, 4$. Além disso, seja $\mathbf{Z}^{(2)} = \mathbf{Z} \odot \mathbf{Z}$, $\mathbf{Z}_2^{(2)} = \mathbf{Z}_2 \odot \mathbf{Z}_2$, $\mathbf{Z}^{(3)} = \mathbf{Z}^{(2)} \odot \mathbf{Z}$, etc., em que “ \odot ” denota o produto elemento a elemento de matrizes (produto de Hadamard). As matrizes $\phi^{-1}\mathbf{Z}$ e $\phi^{-1}\mathbf{Z}_2$ representam a estrutura de covariância assintótica de $\mathbf{X}\hat{\boldsymbol{\beta}}$ e $\mathbf{X}_2\tilde{\boldsymbol{\beta}}_2$, respectivamente.

Dos resultados gerais de Lawley (1956), Cordeiro et al. (1994) obtiveram um fator de correção de Bartlett para a estatística LR para testar $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ nos modelos próprios de dispersão. A estatística LR corrigida tem a forma

$$S_{\text{LR}}^* = S_{\text{LR}}(1 - a_{\text{LR}}), \quad (3.5)$$

sendo $a_{\text{LR}} = (A_{\text{LR}} + A_{\text{LR},\beta\phi})/q$,

$$\begin{aligned} A_{\text{LR}} = & -\frac{1}{3\phi} \mathbf{1}_n^\top \mathbf{G}(\mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)})(\mathbf{F} + \mathbf{G})\mathbf{1}_n + \frac{1}{4\phi} \mathbf{1}_n^\top (\mathbf{Q} + 4(\mathbf{R} - \mathbf{S}))(\mathbf{Z}_d^2 - \mathbf{Z}_{2d}^2)\mathbf{1}_n \\ & + \frac{1}{12\phi} \mathbf{1}_n^\top \mathbf{F}[2(\mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)}) + 3(\mathbf{Z}_d \mathbf{Z} \mathbf{Z}_d - \mathbf{Z}_{2d} \mathbf{Z}_2 \mathbf{Z}_{2d})]\mathbf{F}\mathbf{1}_n \\ & + \phi^{-1} \mathbf{1}_n^\top (\mathbf{E} + \mathbf{G})[(\mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)}) + (\mathbf{Z}_d \mathbf{Z} \mathbf{Z}_d - \mathbf{Z}_{2d} \mathbf{Z}_2 \mathbf{Z}_{2d})](\mathbf{E} - \mathbf{F} + \mathbf{G})\mathbf{1}_n, \end{aligned}$$

$$A_{\text{LR},\beta\phi} = \frac{q}{2n} \left[\frac{d_{(2)} + d_{(3)}}{d_{(2)}^2} - \frac{2p - q}{2d_{(2)}} \right],$$

em que a notação $(\cdot)_d$ indica que os elementos fora da diagonal da matriz são zero, $\mathbf{1}_n = (1, \dots, 1)^\top$ é um vetor n -dimensional de uns, $d_{(2)} = d_{(2)}(\phi) = \phi^2 d_1''(\phi)$ e $d_{(3)} = d_{(3)}(\phi) = \phi^3 d_1'''(\phi)$, com $d_1''(\phi) = dd_1'(\phi)/d\phi$ e $d_1'''(\phi) = dd_1''(\phi)/d\phi$. Sob a hipótese nula, a estatística S_{LR}^* tem distribuição assintótica χ_q^2 com erro de aproximação reduzido de ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-2})$.

Dos resultados gerais de Cordeiro e Ferrari (1991), Ferrari et al. (2001) obtiveram um fator de correção tipo-Bartlett para a estatística escore para testar $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ nos modelos próprios de dispersão. A estatística escore corrigida é dada por

$$S_{\text{R}}^* = S_{\text{R}}[1 - (c_{\text{R}} + b_{\text{R}}S_{\text{R}} + a_{\text{R}}S_{\text{R}}^2)], \quad (3.6)$$

em que $a_R = A_{R3}/[12q(q+2)(q+4)]$, $b_R = (A_{R22} - 2A_{R3})/[12q(q+2)]$, $c_R = (A_{R11} - A_{R22} + A_{R3})/(12q)$, $A_{R11} = A_{R1} + A_{R1,\beta\phi}$, $A_{R22} = A_{R2} + A_{R2,\beta\phi}$,

$$\begin{aligned} A_{R1} &= 3\phi^{-1}\mathbf{1}_n^\top (2\mathbf{E} + 2\mathbf{G} - \mathbf{F}) \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{Z}_{2d}(2\mathbf{E} + 2\mathbf{G} - \mathbf{F}) \mathbf{1}_n \\ &\quad - 6\phi^{-1}\mathbf{1}_n^\top (2\mathbf{E} + 2\mathbf{G} - \mathbf{F}) \mathbf{Z}_{2d}\mathbf{Z}_2(\mathbf{Z} - \mathbf{Z}_2)_d(2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) \mathbf{1}_n \\ &\quad - 6\phi^{-1}\mathbf{1}_n^\top (3\mathbf{F} - 2\mathbf{G} - 4\mathbf{E}) [\mathbf{Z}_2^{(2)} \odot (\mathbf{Z} - \mathbf{Z}_2)] (2\mathbf{E} + 2\mathbf{G} - \mathbf{F}) \mathbf{1}_n \\ &\quad - 6\phi^{-1}\mathbf{1}_n^\top \mathbf{H}(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_{2d}\mathbf{1}_n, \end{aligned}$$

$$\begin{aligned} A_{R2} &= 6\phi^{-1}\mathbf{1}_n^\top (2\mathbf{E} + 2\mathbf{G} - \mathbf{F}) \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d(2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) \mathbf{1}_n \\ &\quad - 3\phi^{-1}\mathbf{1}_n^\top (2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) [(\mathbf{Z} - \mathbf{Z}_2)_d\mathbf{Z}_2(\mathbf{Z} - \mathbf{Z}_2)_d \\ &\quad \quad + 2(\mathbf{Z} - \mathbf{Z}_2)_d^{(2)} \odot \mathbf{Z}_2] (2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) \mathbf{1}_n \\ &\quad + 3\phi^{-1}\mathbf{1}_n^\top \mathbf{J}(\mathbf{Z} - \mathbf{Z}_2)_d^{(2)}\mathbf{1}_n, \end{aligned}$$

$$\begin{aligned} A_{R3} &= \phi^{-1}\mathbf{1}_n^\top (2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) [3(\mathbf{Z} - \mathbf{Z}_2)_d(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d \\ &\quad + 2(\mathbf{Z} - \mathbf{Z}_2)^{(3)}] (2\mathbf{F} - 2\mathbf{G} - 3\mathbf{E}) \mathbf{1}_n, \end{aligned}$$

$$\begin{aligned} A_{R1,\beta\phi} &= \frac{6}{nd_{(2)}^2} [d_{(3)} - (p - q - 2)d_{(2)}] [q - \phi \text{tr}\{\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\}] \\ &\quad + \frac{6\phi}{nd_{(2)}} [2\text{tr}\{(2\mathbf{W} - \phi\mathbf{W}^{(1)})(\mathbf{Z} - \mathbf{Z}_2)\mathbf{W}^{(1)}\mathbf{Z}_2\} + \phi \text{tr}\{\mathbf{W}^{(2)}(\mathbf{Z} - \mathbf{Z}_2)\}], \end{aligned}$$

$$\begin{aligned} A_{R2,\beta\phi} &= \frac{3}{nd_{(2)}} [(-q + \phi \text{tr}\{\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\})^2 + 2(q - 2\phi \text{tr}\{\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\})] \\ &\quad + \frac{6\phi^2}{nd_{(2)}} \text{tr}\{\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\}, \end{aligned}$$

sendo $\text{tr}\{\cdot\}$ o operador traço. O fator de correção $[1 - (c_R + b_R S_R + a_R S_R^2)]$ em (3.6) reduz o erro de aproximação da distribuição assintótica da estatística S_R^* de ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-2})$.

Em um cenário geral, Vargas et al. (2013) derivaram um fator de correção tipo-Bartlett para a estatística gradiente que pode ser particularizado para cada modelo específico de interesse. Utilizando esses resultados obtivemos expressões matriciais para o fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos próprios de dispersão.

Proposição 2. *Considere o problema de testar a hipótese nula $\mathcal{H}_0 : \beta_1 = \beta_{10}$ em modelos de dispersão. A correção tipo-Bartlett para a estatística gradiente é dada por*

$$S_T^* = S_T [1 - (c_T + b_T S_T + a_T S_T^2)], \quad (3.7)$$

sendo a_T , b_T e c_T da ordem $O(n^{-1})$ e dados por $a_T = A_{T3}/[12q(q+2)(q+4)]$, $b_T = (A_{T22} - 2A_{T3})/[12q(q+2)]$ e $c_T = (A_{T11} - A_{T22} + A_{T3})/(12q)$, com $A_{T11} = A_{T1} + A_{T1,\beta\phi}$ e $A_{T22} = A_{T2} + A_{T2,\beta\phi}$,

$$\begin{aligned} A_{T1} = & 3\phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) [\mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)\mathbf{Z}_{2d} + 2\mathbf{Z}_2^{(2)} \odot (\mathbf{Z} - \mathbf{Z}_2) + \\ & + 2(\mathbf{Z} - \mathbf{Z}_2)_d \mathbf{Z}_2 \mathbf{Z}_{2d} + \mathbf{Z}_{2d}(\mathbf{Z} - \mathbf{Z}_2)(\mathbf{Z} - \mathbf{Z}_2)_d] (\mathbf{F} + 2\mathbf{G}) \mathbf{1}_n \\ & - 6\phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) [(\mathbf{Z} + \mathbf{Z}_2) \odot (\mathbf{Z}^{(2)} - \mathbf{Z}_2^{(2)}) + (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z} \mathbf{Z}_d + \mathbf{Z}_2 \mathbf{Z}_{2d}) \\ & + 2\mathbf{Z}_{2d} (\mathbf{Z} \mathbf{Z}_d - \mathbf{Z}_2 \mathbf{Z}_{2d}) + 2\mathbf{Z}_2^{(2)} \odot (\mathbf{Z} - \mathbf{Z}_2)] (\mathbf{E} + 2\mathbf{G}) \mathbf{1}_n \\ & + 12\phi^{-1} \mathbf{1}_n^\top (\mathbf{E} + 2\mathbf{G}) [\mathbf{Z}_d \mathbf{Z} \mathbf{Z}_d - \mathbf{Z}_{2d} \mathbf{Z}_2 \mathbf{Z}_{2d} + \mathbf{Z}^{(3)} - \mathbf{Z}_2^{(3)}] (\mathbf{E} + 2\mathbf{G}) \mathbf{1}_n \\ & - 6\phi^{-1} \mathbf{1}_n^\top \mathbf{S} (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z}_d + 3\mathbf{Z}_{2d}) \mathbf{1}_n + 6\phi^{-1} \mathbf{1}_n^\top \mathbf{Q} (\mathbf{Z} - \mathbf{Z}_2)_d \mathbf{Z}_{2d} \mathbf{1}_n \\ & + 12\phi^{-1} \mathbf{1}_n^\top \mathbf{R} (\mathbf{Z}_d^2 - \mathbf{Z}_{2d}^2) \mathbf{1}_n, \end{aligned}$$

$$\begin{aligned} A_{T2} = & -3\phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) \left[\frac{3}{4} (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z} - \mathbf{Z}_2) (\mathbf{Z} - \mathbf{Z}_2)_d + 2\mathbf{Z}_2 \odot (\mathbf{Z} - \mathbf{Z}_2)^{(2)} \right. \\ & \left. + \frac{1}{2} (\mathbf{Z} - \mathbf{Z}_2)^{(3)} + \mathbf{Z}_{2d} (\mathbf{Z} - \mathbf{Z}_2) (\mathbf{Z} - \mathbf{Z}_2)_d + (\mathbf{Z} - \mathbf{Z}_2)_d \mathbf{Z}_2 (\mathbf{Z} - \mathbf{Z}_2)_d \right] (\mathbf{F} + 2\mathbf{G}) \mathbf{1}_n \\ & + 6\phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) [(\mathbf{Z} - \mathbf{Z}_2) \odot (\mathbf{Z}^{(2)} - \mathbf{Z}_2^{(2)}) + (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z} \mathbf{Z}_d - \mathbf{Z}_2 \mathbf{Z}_{2d})] (\mathbf{E} + 2\mathbf{G}) \mathbf{1}_n \\ & + 3\phi^{-1} \mathbf{1}_n^\top (2\mathbf{S} - \mathbf{Q}) (\mathbf{Z} - \mathbf{Z}_2)_d^{(2)} \mathbf{1}_n, \end{aligned}$$

$$A_{T3} = \phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) \left[\frac{3}{4} (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z} - \mathbf{Z}_2) (\mathbf{Z} - \mathbf{Z}_2)_d + \frac{1}{2} (\mathbf{Z} - \mathbf{Z}_2)^{(3)} \right] (\mathbf{F} + 2\mathbf{G}) \mathbf{1}_n,$$

$$A_{T1,\beta\phi} = \frac{6q}{nd_{(2)}^2} [d_{(3)} + (2-p+q)d_{(2)}], \quad A_{T2,\beta\phi} = \frac{3q(q+2)}{nd_{(2)}}.$$

Quando ϕ é conhecido, $A_{T1,\beta\phi} = A_{T2,\beta\phi} = 0$.

Prova. Ver Seção 3.7.

O fator de correção $[1 - (c_T + b_T S_T + a_T S_T^2)]$ em (3.7) reduz o erro de aproximação da distribuição assintótica da estatística S_T^* de ordem $\mathcal{O}(n^{-1})$ para ordem $\mathcal{O}(n^{-2})$.

É interessante notar que os coeficientes A_{LR} , A_R 's e A_T 's são funções da matriz modelo \mathbf{X} através de \mathbf{Z} e \mathbf{Z}_2 , do parâmetro ϕ e das derivadas de segunda e terceira ordem da função de ligação. Os termos $A_{LR,\beta\phi}$ e $A_{T1,\beta\phi}$ dependem do número de parâmetros da regressão (p) e do número de parâmetros de interesse (q), enquanto que os termos $A_{R1,\beta\phi}$ e $A_{R2,\beta\phi}$ são funções dessas quantidades e da matriz modelo, e o termo $A_{T2,\beta\phi}$ é função do número de parâmetros de interesse. Note que essas expressões envolvem apenas operações matriciais de fácil implementação computacional em linguagens de programação com suporte para operações matriciais como, por exemplo, Ox (Doornik 2013), R (R Core Team 2013) e MAPLE (Rafter et al, 2003).

Um caso particular dos modelos de dispersão é obtido fazendo $t(y, \theta) = y\theta - b(\theta)$ em (3.1), da qual obtemos a classe dos MLGs. Nesse caso, temos que $g(db(\theta_l)/d\theta_l) = g(\mu_l) = \eta_l = \mathbf{x}_l^\top \boldsymbol{\beta}$, $D_{2l} = -V_l^{-1}$, $D_{3l} = 2V_l^{-2} dV_l/d\mu_l$, $D_{4l} = 3V_l^{-2} [d^2V_l/d\mu_l^2 - 2V_l^{-1}(dV_l/d\mu_l)^2]$, $D'_{2l} = V_l^{-2} dV_l/d\mu_l$, $D'_{3l} = 2V_l^{-2} [d^2V_l/d\mu_l^2 - 2V_l^{-1}(dV_l/d\mu_l)^2]$, $D''_{2l} = V_l^{-2} [d^2V_l/d\mu_l^2 - 2V_l^{-1}(dV_l/d\mu_l)^2]$ e $D_{2l}^{(2)} = V_l^{-2} + V_l^{-3}(dV_l/d\mu_l)^2/\phi$. Além disso,

$$f_l = \frac{1}{V_l} \frac{d\mu_l}{d\eta_l} \frac{d^2\mu_l}{d\eta_l^2} - 2 \frac{1}{V_l^2} \frac{dV_l}{d\mu_l} \left(\frac{d\mu_l}{d\eta_l} \right)^3, \quad g_l = \frac{1}{V_l^2} \frac{dV_l}{d\mu_l} \left(\frac{d\mu_l}{d\eta_l} \right)^3,$$

$$s_l = 9\lambda_{1l} - 3\lambda_{2l} - 3\lambda_{3l} - 4\lambda_{4l} + 2\lambda_{5l}, \quad r_l = 5\lambda_{1l} - 2\lambda_{2l} - 2\lambda_{3l} - 2\lambda_{4l} + \lambda_{5l},$$

$$q_l = 12\lambda_{1l} - 3\lambda_{2l} - 4\lambda_{3l} - 6\lambda_{4l} + 3\lambda_{5l}, \quad j_l = \lambda_{4l} + \lambda_{5l}, \quad h_l = \lambda_{1l} + \lambda_{5l},$$

em que

$$\lambda_{1l} = \frac{1}{V_l^2} \frac{dV_l}{d\mu_l} \left(\frac{d\mu_l}{d\eta_l} \right)^2 \frac{d^2\mu_l}{d\eta_l^2}, \quad \lambda_{2l} = \frac{1}{V_l} \left(\frac{d^2\mu_l}{d\eta_l^2} \right)^2, \quad \lambda_{3l} = \frac{1}{V_l} \frac{d\mu_l}{d\eta_l} \frac{d^3\mu_l}{d\eta_l^3},$$

$$\lambda_{4l} = \frac{1}{V_l^3} \left(\frac{dV_l}{d\mu_l} \right)^2 \left(\frac{d\mu_l}{d\eta_l} \right)^4, \quad \lambda_{5l} = \frac{1}{V_l^2} \frac{d^2V_l}{d\mu_l^2} \left(\frac{d\mu_l}{d\eta_l} \right)^4.$$

Substituindo essas quantidades em (3.5), (3.6) e (3.7) e após manipulações algébricas, obtemos os fatores de correção de Bartlett para a estatística LR dado em Cordeiro (1987), tipo-Bartlett para a estatística escore derivado por Cribari e Ferrari (1995) e tipo-Bartlett para a estatística gradiente obtido por Vargas et al. (2014).¹

Nesse ponto surge o seguinte questionamento: como se comportam os testes corrigidos por fatores de correção de Bartlett e tipo-Bartlett em amostras pequenas e moderadas na classe dos modelos de dispersão? Para responder a essa pergunta, na Seção 3.4 comparamos via simulação Monte Carlo o tamanho e o poder dos testes que usam as estatísticas S_{LR}^* , S_R^* , S_T^* , S_W , S_{LR} , S_R e S_T . Além disso, consideramos os testes bootstrap com base nas estatísticas de Wald, LR, escore e gradiente para efeito de comparação.

3.4 Estudo de simulação

Nessa seção conduzimos experimentos de simulação Monte Carlo para comparar o tamanho e o poder dos testes de Wald, LR, escore, gradiente e dos testes aperfeiçoados (S_{LR}^* , S_R^* e S_T^*) em amostras pequenas ou moderadas para a classe dos modelos de dispersão. Além disso, consideramos versões bootstrap para os testes de Wald, LR, escore e gradiente. As simulações foram realizadas utilizando a linguagem de programação Ox (Doornik 2013), que é distribuída gratuitamente para fins acadêmicos e disponível em <http://www.doornik.com>. Todas as maximizações da função de log-verossimilhança para os parâmetros do modelo foram realizadas utilizando o método quasi-Newton (BFGS) com primeiras derivadas analíticas através da sub-rotina MaxBFGS.

Calculamos a probabilidade de erro tipo I da hipótese nula $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ dos testes que utilizam as estatísticas S_W , S_{LR} , S_R , S_T , S_{LR}^* , S_R^* e S_T^* para os níveis nominais $\alpha = 10\%$, 5% e 1% ; isto é, calculamos a proporção de vezes que o valor calculado para cada estatística de teste é maior do que o quantil $1 - \alpha$ da distribuição χ_q^2 em 10000 réplicas Monte

¹Na fórmula de b_l em Vargas et al. (2014, p. 113), a quantidade λ_{3l} deve ser substituído por λ_{5l} ; veja a expressão de b em Cribari e Ferrari (1995, p. 428).

Carlo. Além disso, foram considerados diferentes valores para o número de parâmetros da regressão (p), número de parâmetros da hipótese nula (q) e diferentes tamanhos de amostra ($n = 20, 25$ e 30). Consideramos 600 réplicas bootstrap.

Inicialmente foi considerado o modelo von Mises, que é bastante utilizado na modelagem de dados circulares; ver Carta et al. (2008) e Mardia e Jupp (2000). Considere

$$\pi(y; \theta, \phi) = \frac{\exp[\phi \cos(y - \theta)]}{2\pi I_0(\phi)}, \quad y \in (-\pi, \pi), \quad (3.8)$$

em que $\theta \in (-\pi, \pi)$, $\phi > 0$ é o parâmetro de precisão e $I_\nu(\cdot)$ denota a função modificada de Bessel de primeiro tipo e ordem ν . Essa distribuição é simétrica em torno de $y = \theta$ e quanto maior o valor do parâmetro de precisão mais concentrada será a função densidade em torno de θ . O modelo dado em (3.8) é um caso particular de (3.1) quando $t(y, \theta) = \cos(y - \theta)$ e $a(y, \phi) = -\log(I_0(\phi)) - \log(2\pi)$, caracterizando um modelo próprio de dispersão. Consideramos a parte sistemática da forma

$$\tan(\theta_l/2) = \eta_l = \beta_0 + \beta_1 x_{l1} + \cdots + \beta_{p-1} x_{l,p-1},$$

em que $\theta_l = 2 \arctan(\eta_l)$, $l = 1, \dots, n$. As covariáveis $x_{l1}, \dots, x_{l,p-1}$ foram extraídas de uma distribuição $\mathcal{U}(-0,5; 0,5)$, os parâmetros da regressão, exceto os fixados na hipótese nula, são iguais a um e o parâmetro de precisão é fixado em $\phi = 4$.

As Tabelas 3.3 e 3.4 indicam que os testes de Wald (S_W) e LR (S_{LR}) são os mais liberais entre os testes considerados, isto é, rejeitam a hipótese nula com mais frequência do que o esperado com base no nível nominal selecionado. Por exemplo, para $p = 6$, $n = 25$ e $\alpha = 5\%$, as taxas de rejeição para ambos os testes foram, respectivamente, 14,06% e 11,84% ($q = 3$), 12,71% e 11,23% ($q = 2$), e 19,04% e 17,84% ($q = 1$); ver Tabela 3.4. Apesar das taxas de rejeição do teste LR serem menores do que as do teste de Wald, quando comparadas com os testes escore e gradiente esses valores são bem maiores. Considerando o exemplo anterior, as taxas de rejeição dos testes escore e gradiente foram 6,70% e 7,03%, 8,27% e 8,59%, e 15,90% e 16,42%, respectivamente.

Quando o número de parâmetros da regressão está fixado e o número de parâmetros testados decresce, as taxas de rejeição dos testes que usam as estatísticas S_W e S_{LR} tendem

a diminuir, enquanto que as taxas de rejeição dos testes que usam S_R e S_T , em geral, tendem a aumentar. Por exemplo, para $p = 4$, $n = 20$ e $\alpha = 10\%$, as taxas de rejeição dos testes de Wald e LR foram 20,70% (S_W) e 16,98% (S_{LR}) ($q = 3$), 17,13% (S_W) e 16,59% (S_{LR}) ($q = 2$), e 15,72% (S_W) e 15,43% (S_{LR}) ($q = 1$), enquanto que as taxas de rejeição dos testes que usam S_R e S_T foram 11,49% (S_R) e 11,48% (S_T) ($q = 3$), 12,92% (S_R) e 13,47% (S_T) ($q = 2$), e 14,16% (S_R) e 14,33% (S_T) ($q = 1$); ver Tabela 3.3.

Por outro lado, os testes que utilizam as estatísticas S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b e S_T^b apresentam taxas de rejeição menos distorcidas que os testes não corrigidos (S_W , S_{LR} , S_R e S_T). Enquanto que o número de parâmetros da regressão e o número de parâmetros a serem testados influenciam nas taxas de rejeição dos testes usuais, nos testes corrigidos essas quantidades têm pouca influência sobre as taxas de rejeição. Por exemplo, para $p = 6$, $n = 20$ e $\alpha = 5\%$, as taxas de rejeição foram 4,82% (S_{LR}^*), 4,79% (S_R^*) e 4,85% (S_T^*), 4,78% (S_W^b), 4,84% (S_{LR}^b), 4,91% (S_R^b) e 5,10% (S_T^b) quando $q = 3$, 4,91% (S_{LR}^*), 5,34% (S_R^*), 5,19% (S_T^*), 4,80% (S_W^b), 5,07% (S_{LR}^b), 5,12% (S_R^b) e 5,10% (S_T^b) quando $q = 2$, e 4,79% (S_{LR}^*), 5,21% (S_R^*), 5,05% (S_T^*), 4,72% (S_W^b), 4,88% (S_{LR}^b), 4,91% (S_R^b) e 4,85% (S_T^b) quando $q = 1$; veja Tabela 3.4.

No segundo experimento de simulação consideramos o modelo de dispersão log-gama que é bastante utilizado em análise de sobrevivência para caracterizar modelos de tempo de vida; veja, por exemplo, Ortega et al. (2009), Ortega et al. (2008) e Lawless (2003, Cap. 5). Temos que

$$\pi(y; \theta, \phi) = \exp[\phi \{y - \theta - \exp(y - \theta)\} + \phi \log \phi - \log \Gamma(\phi)], \quad y \in \mathbb{R}, \quad (3.9)$$

em que $\theta \in \mathbb{R}$ e $\phi > 0$. A densidade dada em (3.9) é um caso particular de (3.1) com $t(y, \theta) = y - \theta - \exp(y - \theta)$ e $a(y, \theta) = \phi \log \phi - \log \Gamma(\phi)$, caracterizando um modelo próprio de dispersão. Consideramos a parte sistemática da forma

$$g(\theta_l) = \eta_l = \theta_l = \beta_0 + \beta_1 x_{l1} + \cdots + \beta_{p-1} x_{l,p-1}.$$

As covariáveis $x_{l1}, \dots, x_{l,p-1}$ foram extraídas de uma distribuição $\mathcal{U}(0; 1)$. Os parâmetros da regressão, exceto aqueles fixados na hipótese nula, são iguais a um e o parâmetro de precisão

Tabela 3.3: Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo von Mises.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	20,70	16,98	11,49	11,48	10,38	10,25	10,18	10,00	10,34	10,50	10,57
		5	13,72	10,05	5,06	4,89	5,18	4,86	4,86	5,02	5,25	5,31	5,36
		1	5,46	2,66	0,52	0,44	0,98	0,78	0,80	1,04	1,05	1,13	1,03
	25	10	18,52	14,85	10,65	10,26	9,75	9,63	9,66	9,93	9,72	9,73	9,85
		5	11,71	8,43	4,77	4,50	4,84	4,74	4,65	4,84	4,86	4,90	4,99
		1	4,23	2,05	0,47	0,39	0,79	0,65	0,65	0,79	0,84	0,85	0,85
	30	10	16,77	14,28	10,76	10,74	10,09	9,87	9,96	10,19	10,09	10,03	10,09
		5	10,25	7,65	4,91	4,82	4,93	4,78	4,82	4,99	4,98	4,95	5,05
		1	3,29	2,00	0,83	0,72	1,10	0,96	0,97	1,06	1,19	1,09	1,16
2	20	10	17,13	16,59	12,92	13,47	9,74	10,04	9,75	9,64	9,85	10,01	9,83
		5	10,46	9,43	5,99	6,49	4,86	4,71	4,67	4,69	4,92	4,86	4,88
		1	3,49	2,63	0,94	1,08	1,01	0,93	0,90	1,04	1,14	1,12	1,16
	25	10	17,05	15,38	12,68	12,96	10,45	10,62	10,55	10,50	10,51	10,63	10,57
		5	10,51	8,96	6,24	6,41	5,30	5,13	5,25	5,36	5,42	5,22	5,45
		1	3,70	2,50	1,01	1,09	1,15	0,99	1,08	1,29	1,21	1,22	1,27
	30	10	15,02	13,97	11,64	12,11	10,05	9,97	10,17	10,03	10,06	9,97	10,17
		5	8,91	7,79	5,91	6,04	5,09	5,10	5,08	5,13	5,12	5,25	5,18
		1	2,91	2,13	1,05	1,08	1,11	1,03	1,05	1,30	1,19	1,20	1,20
1	20	10	15,72	15,43	14,16	14,33	10,12	10,53	10,40	10,09	10,19	10,21	10,26
		5	9,21	9,08	7,68	7,75	5,08	5,24	5,31	4,94	5,21	5,23	5,33
		1	2,77	2,47	1,33	1,42	0,96	0,86	0,96	0,95	1,04	1,02	1,09
	25	10	14,35	13,63	12,33	12,65	9,84	9,92	9,97	9,69	9,90	9,84	9,90
		5	8,34	7,58	6,43	6,65	4,93	5,05	4,99	4,97	4,99	5,03	5,03
		1	2,46	2,08	1,27	1,14	0,94	0,94	0,93	1,02	0,99	1,02	0,97
	30	10	13,88	13,71	12,75	13,02	10,14	10,21	10,20	9,97	10,21	10,16	10,22
		5	7,64	7,42	6,43	6,64	5,01	5,06	5,05	5,04	5,07	5,04	5,11
		1	2,16	2,00	1,24	1,32	0,83	0,93	0,89	0,93	0,97	1,02	0,98

Tabela 3.4: Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo von Mises.

q	n	$\alpha(\%)$	S_W	S_{LR}^*	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	26,37	22,68	15,20	15,77	9,90	10,29	10,29	9,77	9,92	10,09	10,06
		5	18,27	14,29	7,02	7,41	4,82	4,79	4,85	4,78	4,84	4,91	5,10
		1	7,95	4,62	0,89	1,05	1,01	0,81	0,77	1,12	1,11	1,15	1,13
	25	10	21,85	19,53	14,13	14,76	10,02	10,35	10,43	9,97	10,03	10,21	10,35
		5	14,06	11,84	6,70	7,03	4,90	4,82	4,81	4,66	4,93	4,93	4,99
		1	5,06	3,27	0,94	1,09	0,93	0,87	0,93	1,00	1,07	1,08	1,09
	30	10	20,15	17,86	13,47	14,01	10,19	10,52	10,36	10,24	10,14	10,44	10,34
		5	12,80	10,29	6,64	7,09	5,49	5,32	5,43	5,36	5,56	5,44	5,55
		1	4,86	3,33	1,03	1,12	1,11	0,94	0,93	1,18	1,24	1,16	1,17
2	20	10	24,02	20,96	16,21	16,82	9,96	10,36	10,56	10,15	10,10	9,86	10,27
		5	16,43	13,28	8,35	8,98	4,91	5,34	5,19	4,80	5,07	5,12	5,10
		1	6,96	4,41	1,35	1,51	0,89	0,85	0,92	0,93	0,96	1,02	1,07
	25	10	19,80	18,51	15,61	16,04	10,39	10,83	10,72	10,15	10,40	10,53	10,56
		5	12,71	11,23	8,27	8,59	5,24	5,25	5,26	5,13	5,27	5,18	5,22
		1	4,75	3,31	1,66	1,86	1,18	1,13	1,17	1,20	1,32	1,35	1,40
	30	10	17,83	16,23	13,81	13,96	9,99	10,26	10,10	9,93	10,03	10,08	10,01
		5	11,47	9,59	7,00	7,30	5,15	4,94	5,11	5,17	5,18	4,89	5,13
		1	4,13	2,95	1,53	1,50	1,21	1,19	1,10	1,25	1,29	1,29	1,26
1	20	10	19,04	17,84	15,90	16,42	9,59	10,14	9,93	9,67	9,69	9,75	9,77
		5	12,01	10,92	8,90	9,23	4,79	5,21	5,05	4,72	4,88	4,91	4,85
		1	4,37	3,67	1,99	2,14	0,87	0,90	0,94	0,87	1,01	0,94	1,00
	25	10	16,81	15,98	14,61	15,02	9,74	10,13	10,13	9,83	9,72	9,84	9,84
		5	10,33	9,24	7,79	8,04	4,78	5,12	4,94	4,79	4,81	4,90	4,84
		1	3,39	2,71	1,68	1,74	0,98	1,08	1,07	1,00	1,09	1,12	1,13
	30	10	15,28	14,76	13,53	13,81	10,08	10,25	10,18	10,01	10,07	10,07	10,16
		5	9,35	8,74	7,45	7,75	5,08	5,28	5,26	5,07	5,14	5,17	5,26
		1	2,81	2,35	1,49	1,63	0,88	0,91	0,92	0,72	0,95	0,97	0,97

é fixado em $\phi = 3$. Os resultados são apresentados nas Tabelas 3.5 e 3.6.

Assim como verificado para a distribuição von Mises, os testes de Wald e LR são mais liberais que os testes escore e gradiente, sendo o teste de Wald o mais liberal. Como exemplo, considere $n = 20$, $p = 4$, $q = 1$ e $\alpha = 5\%$. As taxas de rejeição da hipótese nula são 10,75% (S_W), 8,53% (S_{LR}), 6,18% (S_R) e 6,86% (S_T); ver Tabela 3.5. Além disso, quando o número de parâmetros da regressão está fixado e o número de parâmetros testados decresce, as taxas de rejeição dos testes de Wald e LR tendem a diminuir, enquanto que as taxas de rejeição dos testes escore e gradiente tendem a aumentar. Por exemplo, para $n = 25$, $p = 6$ e $\alpha = 10\%$, as taxas de rejeição dos testes que usam S_W e S_{LR} foram, respectivamente, 23,71% e 17,30% quando $q = 3$, 21,06% e 16,54% quando $q = 2$, e 17,95% e 15,57% quando $q = 1$; as taxas de rejeição dos testes que usam S_R e S_T foram, respectivamente, 11,77% e 12,70% quando $q = 3$, 12,88% e 13,81% quando $q = 2$, e 13,12% e 14,49% quando $q = 1$; ver Tabela 3.6.

Por outro lado, as taxas de rejeição dos testes que usam as estatísticas S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b e S_T^b são mais próximas do verdadeiro nível nominal do que as taxas de rejeição dos testes que usam as estatísticas S_W , S_{LR} , S_R e S_T . Por exemplo, para $n = 20$, $p = 4$, $q = 1$ e $\alpha = 5\%$, as taxas de rejeição foram 5,01% (S_{LR}^*), 5,22% (S_R^*), 5,13% (S_T^*), 5,24% (S_W^b), 5,13% (S_{LR}^b), 4,89% (S_R^b) e 5,06% (S_T^b); ver Tabela 3.5. Assim, da mesma forma que observado para o modelo von Mises, o número de parâmetros da regressão e o número de parâmetros testados influenciam nas taxas de rejeição dos testes de Wald, LR, escore e gradiente usuais, no entanto, para os testes corrigidos essas quantidades têm pouca influência nas taxas de rejeição. Como exemplo, considere $n = 25$, $p = 6$ e $\alpha = 10\%$, as taxas de rejeição foram 9,53% (S_{LR}^*), 10,32% (S_R^*), 9,77% (S_T^*), 9,96% (S_W^b), 9,55% (S_{LR}^b), 9,86% (S_R^b) e 9,57% (S_T^b) quando $q = 3$, 9,75% (S_{LR}^*), 10,80% (S_R^*), 9,98% (S_T^*), 9,96% (S_W^b), 9,73% (S_{LR}^b), 9,86% (S_R^b) e 9,76% (S_T^b) quando $q = 2$, e 10,31% (S_{LR}^*), 11,19% (S_R^*), 10,51% (S_T^*), 10,46% (S_W^b), 10,29% (S_{LR}^b), 10,12% (S_R^b) e 10,32% (S_T^b) quando $q = 1$; ver Tabela 3.6.

Também comparamos numericamente o poder dos testes. Dos estudos de simulação apresentados anteriormente, alguns testes apresentam diferentes tamanhos sob \mathcal{H}_0 . Para comparar numericamente o poder dos testes é necessário que eles apresentem o mesmo tamanho. Assim, consideramos 500000 réplicas Monte Carlo para estimar o valor crítico exato de cada teste (exceto os testes bootstrap) de acordo com o nível nominal selecionado.

Tabela 3.5: Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 4$; modelo log-gamma.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	22,71	16,27	10,40	10,77	10,28	10,22	10,19	10,19	10,25	10,35	10,38
		5	15,40	9,36	4,50	4,36	5,03	4,89	4,62	5,07	5,09	5,29	5,00
		1	6,67	2,41	0,37	0,31	0,91	0,71	0,61	1,01	1,00	0,96	1,00
	25	10	19,72	14,58	9,99	10,50	10,22	9,92	10,07	10,53	10,24	10,05	10,30
		5	12,77	8,19	4,64	4,87	5,29	4,92	5,05	5,35	5,31	5,22	5,24
		1	5,06	2,34	0,59	0,49	1,23	0,91	0,99	1,25	1,29	1,32	1,36
	30	10	18,11	14,09	10,57	10,49	10,29	10,45	10,17	10,26	10,29	10,60	10,26
		5	11,20	7,79	4,54	4,70	5,09	4,82	4,95	5,24	5,11	5,11	5,11
		1	4,06	1,69	0,64	0,61	0,93	0,91	0,80	0,93	1,00	1,12	1,03
2	20	10	20,64	16,17	11,76	12,68	10,51	10,81	10,62	10,75	10,54	10,51	10,55
		5	13,63	9,51	5,86	6,26	5,45	5,63	5,45	5,57	5,49	5,66	5,65
		1	5,66	2,77	0,66	0,70	1,05	0,76	0,87	1,05	1,10	1,07	1,14
	25	10	18,26	14,58	11,18	11,91	10,12	10,39	10,27	10,42	10,18	10,16	10,25
		5	11,33	8,18	5,25	5,63	4,98	5,08	4,94	4,99	5,06	5,13	5,05
		1	4,02	1,86	0,70	0,73	0,89	0,82	0,81	0,97	0,97	0,94	0,97
	30	10	16,11	13,45	11,12	11,60	10,31	10,51	10,35	10,19	10,38	10,33	10,37
		5	9,86	7,33	5,15	5,43	4,93	4,97	4,87	4,96	4,98	4,98	4,97
		1	3,29	1,96	0,71	0,85	1,00	0,84	0,94	1,10	1,07	0,98	1,08
1	20	10	16,37	14,28	12,15	13,11	10,22	10,72	10,45	10,46	10,25	10,00	10,25
		5	10,75	8,53	6,18	6,86	5,01	5,22	5,13	5,24	5,13	4,89	5,06
		1	3,69	2,19	1,11	1,22	1,04	0,99	1,01	1,15	1,10	1,17	1,12
	25	10	15,29	13,57	11,98	12,74	10,24	10,58	10,34	10,38	10,29	10,10	10,30
		5	9,37	7,65	6,13	6,54	5,16	5,47	5,22	5,14	5,21	5,28	5,22
		1	3,05	2,10	1,13	1,25	1,10	1,10	1,12	1,24	1,12	1,18	1,15
	30	10	13,90	12,55	11,35	11,71	9,80	10,30	9,89	9,72	9,82	9,84	9,87
		5	8,03	6,62	5,47	5,84	4,95	4,97	5,00	4,97	4,99	4,82	5,02
		1	2,46	1,65	1,06	1,17	1,06	1,01	1,02	1,06	1,13	1,09	1,12

Tabela 3.6: Taxas de rejeição (%) de $\mathcal{H}_0 : \beta_1 = \dots = \beta_q = 0$ com $p = 6$; modelo log-gama.

q	n	$\alpha(\%)$	S_W	S_{LR}	S_R	S_T	S_{LR}^*	S_R^*	S_T^*	S_W^b	S_{LR}^b	S_R^b	S_T^b
3	20	10	28,37	20,66	12,60	14,07	10,07	10,70	10,24	10,12	10,00	9,99	10,02
		5	20,21	12,49	5,81	6,45	5,05	5,34	5,02	5,13	5,04	5,06	5,10
		1	9,59	3,89	0,62	0,64	1,04	0,87	0,72	1,11	1,08	1,17	1,19
	25	10	23,71	17,30	11,77	12,70	9,53	10,32	9,77	9,96	9,55	9,86	9,57
		5	15,97	9,99	5,40	5,92	4,76	4,92	4,68	4,76	4,76	4,85	4,71
		1	6,83	2,91	0,78	0,72	0,91	0,90	0,78	0,89	0,94	1,04	0,98
	30	10	21,20	15,93	11,66	12,38	10,01	10,64	10,18	10,02	10,02	10,00	10,13
		5	13,76	9,28	5,80	6,24	5,17	5,34	5,25	5,31	5,21	5,25	5,30
		1	5,61	2,59	0,88	0,94	1,03	0,91	0,98	1,10	1,12	1,07	1,13
2	20	10	24,23	19,14	13,83	15,39	9,69	11,02	10,21	10,14	9,72	9,49	9,79
		5	16,67	11,55	6,98	7,85	4,92	5,56	5,18	4,88	4,90	4,87	5,00
		1	7,55	3,78	1,18	1,23	0,91	1,03	0,77	1,01	1,00	1,03	0,95
	25	10	21,06	16,54	12,88	13,81	9,75	10,80	9,98	9,96	9,73	9,86	9,76
		5	13,57	9,68	6,71	7,38	5,14	5,63	5,28	5,41	5,17	4,98	5,21
		1	5,87	3,17	1,27	1,38	1,06	1,14	1,02	1,18	1,14	1,09	1,13
	30	10	18,48	14,88	11,96	12,81	9,56	10,48	9,73	9,75	9,59	9,60	9,60
		5	11,42	8,45	5,98	6,46	4,76	5,08	4,90	4,97	4,79	4,65	4,87
		1	4,33	2,24	0,96	1,06	0,91	0,95	0,87	1,03	0,97	0,95	0,96
1	20	10	19,39	16,90	14,23	15,43	9,62	11,09	10,07	10,04	9,66	9,62	9,72
		5	12,73	10,11	7,53	8,58	5,01	5,76	5,33	5,25	5,04	4,97	5,12
		1	5,39	3,44	1,57	2,16	1,05	1,16	1,13	1,21	1,09	1,06	1,13
	25	10	17,95	15,57	13,12	14,49	10,31	11,19	10,51	10,46	10,29	10,12	10,32
		5	11,24	9,02	7,04	7,79	5,09	5,65	5,31	5,40	5,14	5,07	5,16
		1	4,19	2,72	1,37	1,75	1,08	1,16	1,11	1,22	1,14	1,10	1,17
	30	10	16,12	14,49	12,77	13,56	9,84	10,64	9,94	9,89	9,86	9,63	9,90
		5	9,49	7,91	6,31	6,96	4,75	5,09	4,86	4,82	4,78	4,73	4,80
		1	3,22	2,24	1,27	1,56	0,93	1,04	0,95	1,00	1,01	1,04	1,02

Para $n = 30$, $p = 3$, $q = 2$ e $\alpha = 10\%$, calculamos a taxa de rejeição da hipótese $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$ para os modelos von Mises e log-gama, considerando diferentes valores para δ . Os resultados são apresentados na Figura 3.1. Como podemos notar, o poder dos testes são similares e, como esperado, a medida que $|\delta|$ cresce, o poder tende a 1. Simulações de poder para diferentes valores de n , p , q , ϕ e α apresentaram resultados semelhantes.

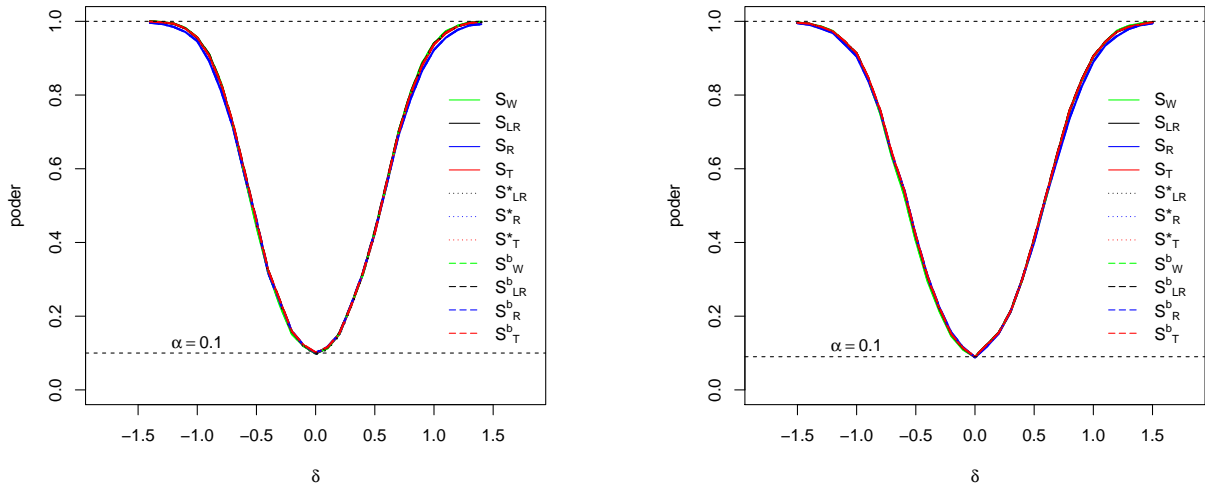


Figura 3.1: Poder dos testes para $n = 30$, $p = 4$, $q = 3$ e $\alpha = 10\%$; von Mises (esquerda) e log-gama (direita).

Em resumo, os estudos de simulação Monte Carlo indicam que, em amostras pequenas ou moderadas, a correção tipo-Bartlett derivada para a estatística gradiente corrige a distorção do tamanho do teste sem perda de poder. As correções de Bartlett para a estatística LR, tipo-Bartlett para a estatística escore e testes via bootstrap também corrigem distorções nos tamanhos dos testes. Os testes de Wald, LR, escore e gradiente usuais apresentaram comportamentos mais liberais, ou seja, rejeitam a hipótese nula com mais frequência do que o esperado com base no nível nominal selecionado. Entre todos os testes considerados, o teste de Wald é claramente o mais liberal. Como não existe um fator de correção tipo-Bartlett para a estatística de Wald em generalidade, o teste de Wald via bootstrap é uma alternativa para corrigir o seu comportamento liberal em amostras de tamanho pequeno e moderado. Por essas razões, recomendamos os testes corrigidos (analiticamente ou via bootstrap) para testar hipóteses em modelos próprios de dispersão quando o tamanho da amostra é pequeno ou moderado.

3.5 Aplicações

Nessa seção apresentamos duas aplicações a conjuntos de dados reais para ilustrar os resultados obtidos nesse capítulo. A primeira aplicação considera o conjunto de dados em [Fisher \(1993\)](#) sobre a distância percorrida por 31 pequenos caramujos azuis (*Nodilittorina unifasciata*). Assim como [Fisher e Lee \(1992\)](#), assumimos uma distribuição von Mises para a direção escolhida pelos animais. Consideramos a função de ligação

$$\tan(\theta_l/2) = \eta_l = \beta_0 + \beta_1 x_{1l}, \quad l = 1, \dots, 31,$$

em que x_{1l} é a distância percorrida (cm) pelo l -ésimo animal, e θ_l é a média direcional. A variável resposta foi transformada para o intervalo $(-\pi, \pi)$. Esse mesmo conjunto de dados foi analisado por [Lemonte e Ferrari \(2012b\)](#) para testar as hipóteses $\mathcal{H}_0 : \beta_1 = 0$ e $\mathcal{H}_0 : \beta_1 \neq \beta_{10}$, para diferentes valores de β_{10} , considerando os testes de Wald, LR, escore e gradiente usuais. Adicionalmente, consideraremos os testes corrigidos analiticamente e testes via bootstrap. As estimativas de máxima verossimilhança dos parâmetros (erro padrão assintótico em parênteses) são: $\hat{\beta}_0 = -0,3233 (0,1506)$, $\hat{\beta}_1 = -0,0129 (0,0039)$ e $\hat{\phi} = 3,2654 (0,7262)$. Os valores das estatísticas de teste (p -valores em parênteses) para testar $\mathcal{H}_0 : \beta_1 = 0$ são: $S_W = 11,0308$ (p -valor 0,0009), $S_{LR} = 9,5257$ (p -valor 0,0020), $S_R = 7,1261$ (p -valor 0,0076), $S_T = 8,2801$ (p -valor 0,0040), $S_{LR}^* = 8,7682$ (p -valor 0,0031), $S_R^* = 7,1901$ (p -valor 0,0073) e $S_T^* = 8,8840$ (p -valor 0,0029). Os p -valores dos testes que usam S_W^b , S_{LR}^b , S_R^b e S_T^b são, respectivamente, 0,0035, 0,0030, 0,0060 e 0,0015. Aos níveis de significância usuais, todos os testes rejeitam a hipótese nula.

Em seguida, testamos a hipótese $\mathcal{H}_0 : \beta_1 = \beta_{10}$ contra a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq \beta_{10}$, para diferentes valores de β_{10} . Os resultados são apresentados na Tabela 3.7 para $\beta_{10} = -0,026, -0,024, -0,022, -0,020$ e $-0,018$. Os asteriscos indicam que a hipótese nula é rejeitada aos níveis de significância de 10% (*), 5% (**) e 1% (***), respectivamente. De acordo com a Tabela 3.7, a hipótese nula $\mathcal{H}_0 : \beta_1 = -0,018$ é rejeitada por todos os testes aos níveis de significância usuais. Em todos os casos considerados, os testes que usam as estatísticas $S_R, S_T, S_R^*, S_T^*, S_{LR}^b, S_R^b$ e S_T^b sugerem a mesma conclusão a respeito da hipótese nula com base no nível nominal considerado. Por outro lado, de acordo com o

nível nominal e a hipótese considerada, os testes que usam S_W , S_{LR} , S_{LR}^* e S_W^b conduzem a conclusões distintas a respeito da hipótese nula. Por exemplo, a hipótese $\mathcal{H}_0 : \beta_1 = -0,022$ é rejeitada pelos testes que usam S_W , S_{LR} e S_W^b ao nível de significância de 5%, no entanto, a decisão se altera quando utilizamos os demais testes. Além disso, podemos notar variações consideráveis no p -valor dos testes. Por exemplo, para a hipótese nula $\mathcal{H}_0 : \beta_1 = -0.020$, o p -valor varia de 6,7% (S_W) a 14,47% (S_R^*).

Tabela 3.7: *Teste de $\mathcal{H}_0 : \beta_1 = \beta_{10}$ contra $\mathcal{H}_0 : \beta_1 \neq \beta_{10}$ (p -valor em parênteses).*

estatística	β_{10}		
	-0,026	-0,024	-0,022
S_W	11,4087(0,0007)***	8,1932(0,0042)***	5,5088(0,0189)**
S_{LR}	7,3137(0,0068)***	5,6057(0,0179)**	4,0115(0,0452)**
S_R	5,8719(0,0154)**	4,6364(0,0313)**	3,4071(0,0649)*
S_T	5,7277(0,0167)**	4,6110(0,0318)**	3,4577(0,0630)*
S_{LR}^*	6,7663(0,0093)***	5,1847(0,0228)**	3,7091(0,0541)*
S_R^*	5,7598(0,0164)**	4,4935(0,0340)**	3,2603(0,0710)*
S_T^*	5,5647(0,0183)**	4,4667(0,0346)**	3,3220(0,0684)*
S_W^b	(0,0035)***	(0,0110)**	(0,0295)**
S_{LR}^b	(0,0105)**	(0,0280)**	(0,0580)*
S_R^b	(0,0160)**	(0,0380)**	(0,0730)*
S_T^b	(0,0205)**	(0,0365)**	(0,0730)*

estatística	β_{10}	
	-0,020	-0,018
S_W	3,3554(0,0670)*	1,7331(0,1880)
S_{LR}	2,5906(0,1075)	1,4114(0,2348)
S_R	2,2511(0,1335)	1,2493(0,2637)
S_T	2,3320(0,1267)	1,3207(0,2505)
S_{LR}^*	2,3946(0,1218)	1,3043(0,2534)
S_R^*	2,1270(0,1447)	1,1669(0,2800)
S_T^*	2,2138(0,1368)	1,2370(0,2600)
S_W^b	(0,0775)*	(0,2105)
S_{LR}^b	(0,1095)	(0,2585)
S_R^b	(0,1335)	(0,2795)
S_T^b	(0,1220)	(0,2740)

A segunda aplicação considera um conjunto de dados apresentados em [Rosillo et al. \(2010\)](#) e [Rosillo \(2011\)](#) sobre o tempo de vida útil de $n = 14$ lâmpadas fluorescentes. Segundo os autores, podemos estabelecer a seguinte relação funcional:

$$\theta_l = \eta_l = \beta_0 + \beta_1 x_{1l} + \beta_2 \frac{x_{2l}}{x_{3l}} + \beta_3 \left(\frac{x_{2l}}{x_{3l}} \right)^2, \quad l = 1, \dots, 14, \quad (3.10)$$

para avaliar o efeito das variáveis regressoras dadas pelo brilho (x_1), voltagem (x_2) e voltagem nominal (x_3) na variável resposta, tempo de vida (y). Assumimos uma distribuição log-gama para o tempo de vida com parâmetro de precisão desconhecido e o mesmo para todas as observações. As estimativas de máxima verossimilhança dos parâmetros (erro padrão assintótico em parênteses) são: $\hat{\beta}_0 = 18,4791(1,7089)$, $\hat{\beta}_1 = 9,4318(4,2940)$, $\hat{\beta}_2 = -35,3273(3,6169)$, $\hat{\beta}_3 = 16,7075(1,8929)$ e $\hat{\phi} = 30,6994(11,5408)$.

As hipóteses $\mathcal{H}_0 : \beta_2 = 0$ e $\mathcal{H}_0 : \beta_3 = 0$ foram rejeitadas por todos os testes (não corrigidos e corrigidos). Em seguida, testamos a hipótese $\mathcal{H}_0 : \beta_1 = 0$. Os valores das estatísticas de teste são apresentados na Tabela 3.8. Ao nível de significância de 5%, os testes que usam S_W , S_{LR} e S_R rejeitam a hipótese nula, no entanto, essa decisão é alterada quando os testes que usam S_T , S_{LR}^* , S_R^* , S_T^* , S_W^b , S_{LR}^b , S_R^b e S_T^b são considerados. Note também que o p -valor varia de 2,81% a 9,67%.

Tabela 3.8: *Teste de $\mathcal{H}_0 : \beta_1 = 0$.*

estatística	valor observado	p -valor
S_W	4,8247	0,0281
S_{LR}	4,2598	0,0390
S_R	3,9429	0,0471
S_T	3,7532	0,0527
S_{LR}^*	3,2287	0,0724
S_R^*	3,2483	0,0715
S_T^*	3,0578	0,0803
S_W^b		0,0967
S_{LR}^b		0,0933
S_R^b		0,0783
S_T^b		0,0900

Como indicado pelos nossos experimentos de simulação, os testes de Wald, LR, escore e gradiente usuais são liberais quando o tamanho da amostra é pequeno (nesse caso $n = 14$). Portanto, devemos ser cautelosos quanto às conclusões utilizando esses testes. Como sugerido pelos testes corrigidos, decidimos pela remoção da covariável “brilho” do modelo (3.10). Assim, o modelo final é dado por

$$\theta_l = \eta_l = \beta_0 + \beta_2 \frac{x_{2l}}{x_{3l}} + \beta_3 \left(\frac{x_{2l}}{x_{3l}} \right)^2, \quad l = 1, \dots, 14.$$

As estimativas de máxima verossimilhança dos parâmetros (erro padrão assintótico em parênteses) são: $\hat{\beta}_0 = 19,3698 (1,9430)$, $\hat{\beta}_2 = -36,7263 (4,1548)$, $\hat{\beta}_3 = 17,3502 (2,1798)$ e $\hat{\phi} = 22,7321 (8,5296)$.

3.6 Considerações finais

Neste capítulo derivamos um fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de dispersão a partir dos resultados gerais apresentados em [Vargas et al. \(2013\)](#). Além disso, comparamos via simulação Monte Carlo o tamanho do teste gradiente que usa a estatística gradiente corrigida com os tamanhos dos testes de Wald, LR, escore e gradiente usuais, e as versões corrigidas pelos fatores de correção de Bartlett do teste LR, tipo-Bartlett do teste escore e testes via bootstrap. Também avaliamos numericamente os poderes dos testes. Duas aplicações a conjuntos de dados reais foram consideradas para ilustrar os resultados obtidos nesse capítulo.

Nossos experimentos de simulação mostraram que os testes corrigidos (analiticamente e via bootstrap) apresentaram probabilidades de erro tipo I mais próximas do nível nominal sem perda de poder. Além disso, os testes de Wald e LR são mais liberais do que os testes escore e gradiente. No entanto, dependendo do tamanho da amostra, do número de parâmetros da regressão e do número de parâmetros testados, os testes escore e gradiente podem apresentar consideráveis distorções na probabilidade de erro tipo I. Por essas razões, em conjuntos de dados reais em que o tamanho da amostra é pequena ou moderada, recomendamos a utilização dos testes corrigidos para a classes dos modelos de dispersão.

3.7 Detalhes técnicos: Prova da Proposição 2

Sejam $\kappa_{rs} = E(\partial^2 \ell / \partial \beta_r \partial \beta_s)$, $\kappa_{rst} = E(\partial^3 \ell / \partial \beta_r \partial \beta_s \partial \beta_t)$, $\kappa_{rstu} = E(\partial^4 \ell / \partial \beta_r \partial \beta_s \partial \beta_t \partial \beta_u)$, $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \beta_t$, $\kappa_{rs}^{(tu)} = \partial^2 \kappa_{rs} / \partial \beta_t \partial \beta_u$, $\kappa_{\phi\phi} = E(\partial^2 \ell / \partial \phi^2)$, $\kappa_{r\phi} = E(\partial^2 \ell / \partial \beta_r \partial \phi)$, $\kappa_{r\phi}^{(s)} = \partial^2 \kappa_{r\phi} / \partial \beta_s$, e assim por diante. Os índices r, s, t e u variam de 1 a p . Para os modelos próprios de dispersão, é possível mostrar que

$$\begin{aligned}
 \kappa_{rs} &= -\phi \sum_{l=1}^n w_l x_{lr} x_{ls}, & \kappa_{rstu} &= \phi \sum_{l=1}^n q_l x_{lr} x_{ls} x_{lt} x_{lu}, & \kappa_{rst}^{(u)} &= \phi \sum_{l=1}^n s_l x_{lr} x_{ls} x_{lt} x_{lu}, \\
 \kappa_{rst} &= -\phi \sum_{l=1}^n (f + 2g)_l x_{lr} x_{ls} x_{lt}, & \kappa_{rs}^{(t)} &= -\phi \sum_{l=1}^n (e + 2g)_l x_{lr} x_{ls} x_{lt}, \\
 \kappa_{rs}^{(tu)} &= \phi \sum_{l=1}^n r_l x_{lr} x_{ls} x_{lt} x_{lu}, & \kappa_{\phi\phi} &= nd_{(2)}/\phi^2, & \kappa_{\phi\phi\phi} &= \kappa_{\phi\phi}^{(\phi)} = nd_{(3)}/\phi^3, \\
 \kappa_{rs\phi} &= -\sum_{l=1}^n w_l x_{lr} x_{ls}, & \kappa_{rs}^{(\phi)} &= -\sum_{l=1}^n (w_l + \phi w_l^{(1)}) x_{lr} x_{ls}, \\
 \kappa_{rs\phi}^{(\phi)} &= -\sum_{l=1}^n w_l^{(1)} x_{lr} x_{ls}, & \kappa_{rs\phi\phi} &= \kappa_{r\phi\phi} = \kappa_{r\phi\phi}^{(s)} = 0.
 \end{aligned} \tag{3.11}$$

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$ um vetor de parâmetros desconhecidos de dimensão $p + 1$ em que $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ e $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. Consideramos o problema de testar a hipótese nula composta $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ contra a hipótese alternativa $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{10}$ sendo $(\boldsymbol{\beta}_2^\top, \phi)^\top$ um vetor de parâmetros de perturbação. Definimos, $\mathcal{A} = \text{diag}\{\mathbf{0}, \mathbf{K}_{\beta_{22}}^{-1}\}$, $\mathcal{M}_\beta = \mathbf{K}_\beta^{-1} - \mathcal{A}_\beta$, $\mathcal{A}_{\beta\phi} = \text{diag}\{\mathcal{A}_\beta, -\kappa_{\phi\phi}^{-1}\}$ e $\mathcal{M}_{\beta\phi} = \text{diag}\{\mathcal{M}_\beta, \mathbf{0}\}$. Sejam $m^{r\phi}$ e $a^{r\phi}$ os $(r, p + 1)$ -ésimos elementos das matrizes $\mathcal{M}_{\beta\phi}$ e $\mathcal{A}_{\beta\phi}$, respectivamente. Analogamente, $m^{\phi\phi}$ e $a^{\phi\phi}$ representam os $(p + 1, p + 1)$ -ésimos elementos de $\mathcal{M}_{\beta\phi}$ e $\mathcal{A}_{\beta\phi}$. Assim, $m^{r\phi} = m^{\phi r} = m^{\phi\phi} = a^{r\phi} = a^{\phi r} = 0$ ($r = 1, \dots, p$), com $a^{\phi\phi} = -\kappa_{\phi\phi}^{-1}$.

Os coeficientes A_T 's que definem uma correção tipo-Bartlett para o teste gradiente nos modelos de dispersão quando ϕ é desconhecido são obtidos substituindo os cumulantes dados em (3.11) nas fórmulas de A_{T11} , A_{T22} e A_{T33} dadas no Teorema 1 de Vargas et al. (2013). Como a matriz informação de Fisher é bloco diagonal e $A_{T3, \beta\phi} = 0$ podemos decompor os coeficientes A 's como $A_{T11} = A_{T1} + A_{T1, \beta\phi}$, $A_{T22} = A_{T2} + A_{T2, \beta\phi}$ e $A_{T33} = A_{T3}$, sendo A_{T1} , A_{T2} e A_{T3} as quantidades obtidas supondo que o parâmetro ϕ é conhecido e $A_{T1, \beta\phi}$ e $A_{T2, \beta\phi}$ as quantidades obtidas supondo que o parâmetro ϕ é desconhecido. Segue do Corolário 2 de Vargas et al. (2013) que

$$S_T^* = S_T [1 - (c_T + b_T S_T + a_T S_T^2)],$$

sendo a_T , b_T e c_T da ordem $O(n^{-1})$ e dados por $a_T = A_{T3}/[12q(q + 2)(q + 4)]$, $b_T = (A_{T22} -$

$2A_{T_3})/[12q(q+2)]$ e $c_T = (A_{T_{11}} - A_{T_{22}} + A_{T_3})/(12q)$, com $A_{T_{11}} = A_{T_1} + A_{T_{1,\beta\phi}}$ e $A_{T_{22}} = A_{T_2} + A_{T_{2,\beta\phi}}$.

Mostraremos aqui, em detalhes, a derivação das quantidades A_{T_3} e A_{T_2} . O termo A_{T_1} é obtido de forma análoga. Do Teorema 1 de Vargas et al. (2013), a quantidade A_{T_3} pode ser escrita como

$$A_{T_3} = \frac{1}{4} \sum' \kappa_{jrs} \kappa_{uvw} (3m^{jr} m^{su} m^{vw} + 2m^{ju} m^{rv} m^{sw}). \quad (3.12)$$

Substituindo os cumulantes dados em (3.11) na expressão (3.12), obtemos

$$\begin{aligned} A_{T_3} &= \frac{3\phi^2}{4} \sum' \sum (f_i + 2g_i) x_{ij} x_{ir} x_{is} \sum (f_m + 2g_m) x_{iu} x_{iv} x_{iw} m^{jr} m^{su} m^{vw} \\ &\quad + \frac{\phi^2}{2} \sum' \sum (f_i + 2g_i) x_{ij} x_{ir} x_{is} \sum (f_m + 2g_m) x_{iu} x_{iv} x_{iw} m^{ju} m^{rv} m^{sw}, \end{aligned}$$

em que os índices i e m variam de 1 a n e \sum' representa a soma sobre os índices dos parâmetros. Invertendo a ordem dos somatórios e rearranjando os termos, obtemos

$$\begin{aligned} A_{T_3} &= \frac{3\phi^2}{4} \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} m^{su} x_{mu} \right) \left(\sum' x_{mv} m^{vw} x_{mw} \right) \\ &\quad + \frac{\phi^2}{2} \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{ju} x_{mu} \right) \left(\sum' x_{ir} m^{rv} x_{mv} \right) \left(\sum' x_{is} m^{sw} x_{mw} \right). \end{aligned}$$

Por exemplo, o termo $\sum' x_{ij} m^{jr} x_{ir} = \sum_{j,r=1}^p x_{ij} m^{jr} x_{ir}$ representa o elemento (j, r) da matriz $\phi^{-1}(\mathbf{Z} - \mathbf{Z}_2)$. Logo,

$$\begin{aligned} A_{T_3} &= \frac{3}{4\phi} \sum (f_i + 2g_i)(f_m + 2g_m) (z_{il} - z_{2il})(z_{im} - z_{2im})(z_{mm} - z_{2mm}) \\ &\quad + \frac{1}{2\phi} \sum (f_i + 2g_i)(f_m + 2g_m) (z_{im} - z_{2im})(z_{im} - z_{2im})(z_{im} - z_{2im}). \end{aligned}$$

Em notação matricial,

$$A_{T_3} = \phi^{-1} \mathbf{1}_n^\top (\mathbf{F} + 2\mathbf{G}) \left[\frac{3}{4} (\mathbf{Z} - \mathbf{Z}_2)_d (\mathbf{Z} - \mathbf{Z}_2) (\mathbf{Z} - \mathbf{Z}_2)_d + \frac{1}{2} (\mathbf{Z} - \mathbf{Z}_2)^{(3)} \right] (\mathbf{F} + 2\mathbf{G}) \mathbf{1}_n,$$

em que $\mathbf{1}_n = (1, \dots, 1)^\top$ é um vetor de uns de dimensão n .

De maneira análoga, do Teorema 1 de Vargas et al. (2013), podemos escrever $A_{T_2} = A_{T_{21}} + A_{T_{22}} + A_{T_{23}}$, em que

$$A_{T21} = -3 \sum' \kappa_{jrs} \kappa_{klu} (m^{jr} m^{sk} a^{lu} + m^{jr} a^{sk} m^{lu} + 2m^{jk} m^{rl} a^{su}) \\ - \frac{3}{4} \sum' \kappa_{jrs} \kappa_{klu} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su}),$$

$$A_{T22} = 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [m^{su} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (\kappa^{sk} \kappa^{lu} - a^{sk} a^{lu})],$$

$$A_{T23} = 6 \sum' \kappa_{jrs}^{(u)} m^{jr} m^{su} - 3 \sum' \kappa_{jrsu} m^{jr} m^{su}.$$

De forma análoga ao que foi feito para a quantidade A_{T3} , obtemos

$$A_{T21} = -3\phi^2 \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} m^{sk} x_{mk} \right) \left(\sum' x_{ml} a^{lu} x_{mu} \right) \\ - 3\phi^2 \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} a^{sk} x_{mk} \right) \left(\sum' x_{ml} m^{lu} x_{mu} \right) \\ - 6\phi^2 \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jk} x_{mk} \right) \left(\sum' x_{ir} m^{rl} x_{ml} \right) \left(\sum' x_{is} a^{su} x_{mu} \right) \\ - \frac{9}{4}\phi^2 \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} m^{sk} x_{mk} \right) \left(\sum' x_{ml} m^{lu} x_{mu} \right) \\ - \frac{3}{2}\phi^2 \sum (f_i + 2g_i)(f_m + 2g_m) \left(\sum' x_{ij} m^{jk} x_{mk} \right) \left(\sum' x_{ir} m^{rl} x_{ml} \right) \left(\sum' x_{is} m^{su} x_{mu} \right),$$

$$A_{T22} = 6\phi^2 \sum (f_i + 2g_i)(f_m + g_m) \left(\sum' x_{is} m^{su} x_{mu} \right) \left(\sum' x_{ij} \kappa^{jk} x_{mk} \right) \left(\sum' x_{ml} \kappa^{lr} x_{ir} \right) \\ - 6\phi^2 \sum (f_i + 2g_i)(f_m + g_m) \left(\sum' x_{is} m^{su} x_{mu} \right) \left(\sum' x_{ij} a^{jk} x_{mk} \right) \left(\sum' x_{ml} a^{lr} x_{ir} \right) \\ + 6\phi^2 \sum (f_i + 2g_i)(f_m + g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} \kappa^{sk} x_{mk} \right) \left(\sum' x_{ml} \kappa^{lu} x_{mu} \right) \\ - 6\phi^2 \sum (f_i + 2g_i)(f_m + g_m) \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} a^{sk} x_{mk} \right) \left(\sum' x_{ml} a^{lu} x_{mu} \right),$$

$$A_{T23} = 6\phi \sum s_i \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} m^{su} x_{iu} \right) \\ - 3\phi \sum q_i \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{is} m^{su} x_{iu} \right).$$

Os termos $\sum' x_{ir} a^{rs} x_{is} = \sum_{r,s=1}^p x_{ir} a^{rs} x_{is}$ e $-\sum' x_{ir} \kappa^{rs} x_{is} = -\sum_{r,s=1}^p x_{ir} \kappa^{rs} x_{is}$ representam os elementos (r, s) das matrizes $\phi^{-1}\mathbf{Z}_2$ e $\phi^{-1}\mathbf{Z}$, respectivamente. Após manipulações algébricas, segue o resultado.

A partir do Teorema 1 de Vargas et al. (2013), podemos mostrar que o termo $A_{T3,\beta\phi} = 0$ e os termos $A_{T1,\beta\phi}$ e $A_{T2,\beta\phi}$ podem ser escritos como

$$\begin{aligned}
 A_{T1,\beta\phi} = & \sum' \kappa_{jrs}\kappa_{u\phi\phi}a^{\phi\phi}m^{jr}m^{su} + \sum' \kappa_{jrs}\kappa_{u\phi\phi}a^{\phi\phi}m^{js}m^{ru} \\
 & + \sum' \kappa_{jrs}\kappa_{u\phi\phi}a^{\phi\phi}m^{ju}m^{rs} + 6 \sum' \kappa_{jr\phi}\kappa_{\phi\phi\phi}m^{jr}(a^{\phi\phi})^2 \\
 & + 6 \sum' \kappa_{jr\phi}\kappa_{vw\phi}m^{jr}a^{vw}a^{\phi\phi} + 6 \sum' \kappa_{jrs}\kappa_{u\phi\phi}m^{jr}a^{su}a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi}\kappa_{uw\phi}m^{ju}a^{vw}a^{\phi\phi} + 3 \sum' \kappa_{jrs}\kappa_{u\phi\phi}m^{ju}a^{rs}a^{\phi\phi} \\
 & + 3 \sum' \kappa_{j\phi\phi}\kappa_{u\phi\phi}m^{ju}(a^{\phi\phi})^2 + 3 \sum' \kappa_{jr\phi}\kappa_{uw\phi}m^{ju}a^{rw}a^{\phi\phi} \\
 & + 3 \sum' \kappa_{js\phi}\kappa_{uw\phi}m^{ju}a^{sv}a^{\phi\phi} + 3 \sum' \kappa_{jr\phi}\kappa_{uw\phi}m^{ju}a^{rv}a^{\phi\phi} \\
 & + 3 \sum' \kappa_{js\phi}\kappa_{uw\phi}m^{ju}a^{sw}a^{\phi\phi} - 6 \sum' (\kappa_{jr\phi}^{(\phi)} - \kappa_{jr\phi\phi})m^{jr}a^{\phi\phi} \\
 & - 6 \sum' \kappa_{jr\phi}^{(\phi)}m^{jr}a^{\phi\phi} - 12 \sum' \kappa_{j\phi\phi}^{(u)}m^{ju}a^{\phi\phi} \\
 & - 12 \sum' \kappa_{kl}^{(\phi)}\kappa_{jr\phi}(\kappa^{jk}\kappa^{lr} - a^{jk}a^{lr})a^{\phi\phi} - 12 \sum' \kappa_{\phi\phi}^{(\phi)}\kappa_{jr\phi}m^{jr}(a^{\phi\phi})^2 \\
 & - 12 \sum' \kappa_{j\phi\phi}\kappa_{kl}^{(u)}(\kappa^{jk}\kappa^{lu} - a^{jk}a^{lu})a^{\phi\phi},
 \end{aligned} \tag{3.13}$$

$$\begin{aligned}
 A_{T2,\beta\phi} = & - \sum' \kappa_{jrs}\kappa_{u\phi\phi}m^{jr}m^{su}a^{\phi\phi} - \sum' \kappa_{jrs}\kappa_{u\phi\phi}m^{js}m^{ru}a^{\phi\phi} \\
 & - \sum' \kappa_{jrs}\kappa_{u\phi\phi}m^{ju}m^{rs}a^{\phi\phi} - 3 \sum' \kappa_{jr\phi}\kappa_{uw\phi}m^{jr}m^{uv}a^{\phi\phi} \\
 & - 3 \sum' \kappa_{jr\phi}\kappa_{uw\phi}m^{ju}m^{rv}a^{\phi\phi} - 3 \sum' \kappa_{jr\phi}\kappa_{uw\phi}m^{jv}m^{ru}a^{\phi\phi}.
 \end{aligned} \tag{3.14}$$

Substituindo os cumulantes dados em (3.11) nas expressões (3.13) e (3.14) algumas quantidades se anulam. Além disso, note que

$$\begin{aligned}
 \sum' \kappa_{jr\phi}\kappa_{\phi\phi\phi}m^{jr}(a^{\phi\phi})^2 &= -\frac{\phi d_{(3)}}{nd_{(2)}^2} \sum w_i \sum' x_{ij}m^{jr}x_{ir} \\
 &= -\frac{d_{(3)}}{nd_{(2)}^2} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} = -\frac{d_{(3)}}{nd_{(2)}^2}q,
 \end{aligned}$$

$$\begin{aligned}
 \sum' \kappa_{jr\phi}\kappa_{vw\phi}m^{jr}a^{vw}a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_i w_m \left(\sum' x_{ij}m^{jr}x_{ir} \right) \left(\sum' x_{mv}a^{vw}x_{mw} \right) \\
 &= -\frac{1}{nd_{(2)}} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} \text{tr}\{\mathbf{W}\mathbf{Z}_2\} = -\frac{q(p-q)}{nd_{(2)}},
 \end{aligned}$$

$$\begin{aligned}\sum' \kappa_{jr}^{(\phi)} m^{jr} a^{\phi\phi} &= \frac{\phi^2}{nd_{(2)}} \sum w_i^{(1)} \left(\sum' x_{ij} m^{jr} x_{ir} \right) \\ &= \frac{\phi}{nd_{(2)}} \text{tr}\{\mathbf{W}^{(1)}(\mathbf{Z} - \mathbf{Z}_2)\},\end{aligned}$$

$$\begin{aligned}\sum' \kappa_{kl}^{(\phi)} \kappa_{jr\phi} a^{\phi\phi} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) &= -\frac{\phi^2}{nd_{(2)}} \sum w_i w_m \left(\sum' x_{ij} \kappa^{jk} x_{mk} \right) \left(\sum' x_{il} \kappa^{lr} x_{mr} \right) \\ &\quad - \frac{\phi^3}{nd_{(2)}} \sum w_i^{(1)} w_m \left(\sum' x_{ij} \kappa^{jk} x_{mk} \right) \left(\sum' x_{il} \kappa^{lr} x_{mr} \right) \\ &\quad + \frac{\phi^2}{nd_{(2)}} \sum w_i w_m \left(\sum' x_{ij} a^{jk} x_{mk} \right) \left(\sum' x_{il} a^{lr} x_{mr} \right) \\ &\quad + \frac{\phi^3}{nd_{(2)}} \sum w_i^{(1)} w_m \left(\sum' x_{ij} a^{jk} x_{mk} \right) \left(\sum' x_{il} a^{lr} x_{mr} \right) \\ &= -\frac{1}{nd_{(2)}} [\text{tr}\{\mathbf{W} \mathbf{Z} \mathbf{W} \mathbf{Z}\} - \text{tr}\{\mathbf{W} \mathbf{Z}_2 \mathbf{W} \mathbf{Z}_2\}] \\ &\quad - \frac{\phi^3}{nd_{(2)}} [\text{tr}\{\mathbf{W}^{(1)} \mathbf{Z} \mathbf{W} \mathbf{Z}\} - \text{tr}\{\mathbf{W}^{(1)} \mathbf{Z}_2 \mathbf{W} \mathbf{Z}_2\}], \\ &= -\frac{q}{nd_{(2)}} - \frac{\phi^3}{nd_{(2)}} [\text{tr}\{\mathbf{W}^{(1)} \mathbf{Z} \mathbf{W} \mathbf{Z}\} - \text{tr}\{\mathbf{W}^{(1)} \mathbf{Z}_2 \mathbf{W} \mathbf{Z}_2\}],\end{aligned}$$

$$\begin{aligned}\sum' \kappa_{\phi\phi}^{(\phi)} \kappa_{jr\phi} m^{jr} (a^{\phi\phi})^2 &= -\frac{\phi d_{(3)}}{nd_{(2)}^2} \sum w_i \sum' x_{ij} m^{jr} x_{ir} \\ &= -\frac{d_{(3)}}{nd_{(2)}^2} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} = -\frac{qd_{(3)}}{nd_{(2)}^2},\end{aligned}$$

$$\begin{aligned}\sum' \kappa_{jr\phi} \kappa_{uw\phi} m^{jr} m^{uw} a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_i w_m \left(\sum' x_{ij} m^{jr} x_{ir} \right) \left(\sum' x_{mu} m^{uw} x_{mv} \right) \\ &= -\frac{1}{nd_{(2)}} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} = -\frac{q^2}{nd_{(2)}},\end{aligned}$$

$$\begin{aligned}\sum' \kappa_{jr\phi} \kappa_{uw\phi} m^{ju} m^{rv} a^{\phi\phi} &= -\frac{\phi^2}{nd_{(2)}} \sum w_i w_m \left(\sum' x_{ij} m^{ju} x_{mu} \right) \left(\sum' x_{ir} m^{rv} x_{mv} \right) \\ &= -\frac{1}{nd_{(2)}} \text{tr}\{\mathbf{W}(\mathbf{Z} - \mathbf{Z}_2) \mathbf{W}(\mathbf{Z} - \mathbf{Z}_2)\} = -\frac{q}{nd_{(2)}}.\end{aligned}$$

Substituindo essas quantidades em (3.13) e (3.14), segue o resultado.

Capítulo 4

Diferentes métodos de redução de viés de estimadores em distribuições log-simétricas

Resumo

Neste capítulo obtivemos expressões para o viés de ordem $\mathcal{O}(n^{-1})$ dos estimadores de máxima verossimilhança na classe de distribuições log-simétricas. Com base nessas expressões, propomos estimadores corrigidos pelo viés. Com esse mesmo objetivo, foram considerados estimadores modificados com base no método proposto por [Firth \(1993\)](#). Além disso, dos resultados de [Pagni et al. \(2016\)](#), obtivemos estimadores modificados com a finalidade de reduzir o viés mediano. Estudos de simulação de Monte Carlo foram realizados para comparar o desempenho dos três estimadores propostos. Para fins ilustrativos, apresentamos uma aplicação a dados reais.

Palavras-chave: Distribuições log-simétricas, estimadores de máxima verossimilhança, estimadores modificados, redução de viés, viés, viés mediano.

4.1 Introdução

Segundo [Limpert et al. \(2001\)](#), é comum a utilização da distribuição log-normal para modelar fenômenos práticos em que os dados apresentam características de assimetria e são

estritamente positivos. Recentemente, [Vanegas e Paula \(2016\)](#) estudaram uma classe de distribuições mais ampla, a classe de distribuições log-simétricas, que inclui a distribuição log-normal como caso particular. Essa classe de distribuições tem suporte no conjunto dos números reais positivos e inclui distribuições com cauda direita mais pesada e mais leve do que a da distribuição log-normal como, por exemplo, as distribuições log-t-Student, log-logística II, log-logística I, log-exponencial-potência e log-slash. Além disso, os parâmetros de escala e forma podem ser interpretados como a mediana e a assimetria (ou dispersão relativa), respectivamente. A classe de distribuições log-simétricas é flexível e possibilita a modelagem de dados estritamente positivos e assimétricos.

Em modelos estatísticos paramétricos regulares o viés do estimador de máxima verossimilhança, em geral, é de ordem $\mathcal{O}(n^{-1})$. Assim, se o tamanho da amostra (n) é pequeno ou moderado, esse viés pode ser não desprezível. Em modelos uniparamétricos, [Bartlett \(1953\)](#) obteve uma expressão para o viés de ordem $\mathcal{O}(n^{-1})$ do estimador de máxima verossimilhança. Posteriormente, supondo observações independentes mas não necessariamente identicamente distribuídas, [Cox e Snell \(1968\)](#) obtiveram uma expressão geral para o viés de ordem $\mathcal{O}(n^{-1})$ do estimador de máxima verossimilhança em modelos multiparamétricos. Esse resultado se tornou amplamente utilizado na literatura para obter expressões gerais para o viés de ordem $\mathcal{O}(n^{-1})$ e propor estimadores corrigidos pelo viés em vários modelos paramétricos: ver, por exemplo, [Cordeiro et al. \(2000\)](#), [Saha e Paul \(2005\)](#), [Lemonte et al. \(2007\)](#), [Cysneiros et al. \(2010\)](#), [Simas et al. \(2011\)](#), [Barreto-Souza e Vasconcellos \(2011\)](#). Para a classe de distribuições log-simétricas, obtemos expressões analíticas fechadas para o viés dos estimadores de máxima verossimilhança e propomos estimadores não viesados até ordem $\mathcal{O}(n^{-2})$.

Outro método de redução de viés, sugerido por [Firth \(1993\)](#), propõe uma função score modificada tal que a solução da correspondente equação de estimação fornece um estimador não viesado até ordem $\mathcal{O}(n^{-2})$. Diferentemente do método de [Cox e Snell \(1968\)](#), a função score é modificada anteriormente à obtenção das estimativas. Nesse sentido, é dito ser um método preventivo de redução de viés. Esse método tem sido estudado em modelos paramétricos em que as estimativas de máxima verossimilhança podem ser instáveis (infinitas ou pertencer à fronteira do espaço de paramétrico) como, por exemplo, em [Bull et al. \(2002\)](#), [Zorn \(2005\)](#), [Sartori \(2006\)](#), [Kosmidis e Firth \(2009, 2011\)](#) e [Kosmidis \(2014b\)](#).

Recentemente, [Pagui et al. \(2016\)](#) propuseram um método baseado em uma função escore modificada tal que a solução da correspondente equação de estimação fornece um estimador não viesado por mediano até ordem $\mathcal{O}(n^{-2})$. Um estimador é dito ser não viesado por mediana se sua mediana coincide com o estimando. Assim como o enfoque de [Firth \(1993\)](#), esse método pode ser visto como preventivo.

Nesse capítulo, obtivemos expressões analíticas para os vieses dos estimadores de máxima verossimilhança na classe de distribuições log-simétricas segundo as expressões gerais de Cox e Snell (1968) e propomos estimadores corrigidos não viesados até ordem $\mathcal{O}(n^{-2})$ para os parâmetros dessa classe de distribuições. Além disso, consideramos estimadores modificados com base nos métodos propostos por [Firth \(1993\)](#) e [Pagui et al. \(2016\)](#). Também avaliamos numericamente o desempenho dos estimadores de máxima verossimilhança e dos estimadores modificados via simulação Monte Carlo.

Este capítulo está organizado como segue. Na Seção 4.2 apresentamos a classe de distribuições log-simétricas e aspectos inferenciais. Na Seção 4.3, com base nas expressões gerais de Cox e Snell (1968), derivamos expressões analíticas para os vieses de ordem $\mathcal{O}(n^{-1})$ dos estimadores de máxima verossimilhança e propomos estimadores corrigidos pelo viés na classe de distribuições log-simétricas. Além disso, obtemos as funções escore modificadas através dos métodos de [Firth \(1993\)](#) e [Pagui et al. \(2016\)](#). Na Seção 4.4 avaliamos via simulação de Monte Carlo o desempenho dos estimadores modificados e os estimadores de máxima verossimilhança dos parâmetros. Na Seção 4.5 apresentamos uma aplicação a dados reais e na Seção 4.6 apresentamos as considerações finais. Detalhes técnicos são apresentados na Seção 4.7.

4.2 A classe de distribuições log-simétricas

Seja y uma variável aleatória simétrica contínua com parâmetro de locação $\mu \in \mathbb{R}$, parâmetro de escala $\phi > 0$ e função densidade dada por $f(y; \mu, \phi) = h[(y - \mu)^2/\phi^2]/\phi$ para $y \in \mathbb{R}$. A função $h : \mathbb{R} \rightarrow [0, \infty)$ tal que $\int_0^\infty u^{-1/2}h(u)du = 1$ é denominada função geradora de densidades e pode depender de parâmetros desconhecidos. Se y tem uma distribuição simétrica com parâmetros μ e ϕ , escrevemos $y \sim S(\mu, \phi^2)$. Considere a transformação

$t = \exp(y)$. A função densidade da variável aleatória t é dada por

$$\pi(t; \eta, \phi) = \frac{h(\tilde{t}^2)}{t\phi}, \quad t > 0,$$

sendo $\tilde{t} = \log \left[(t/\eta)^{\frac{1}{\phi}} \right]$, $\eta = \exp(\mu)$ o parâmetro de escala (mediana) e $\phi > 0$ o parâmetro de forma. Essa classe, gerada pela transformação t , foi estudada por [Vanegas e Paula \(2016\)](#) e é denominada classe de distribuições log-simétricas, cujo suporte é $(0, \infty)$. Se t tem uma distribuição log-simétrica com parâmetros η e ϕ , escrevemos $t \sim LS(\eta, \phi^2)$. Essa classe comporta distribuições com cauda direita mais leve e mais pesada do que a da distribuição log-normal. São membros dessa classe as distribuições log-normal, log-t-Student, log-logística I, log-logística II, log-exponencial-potência entre outras; ver Tabela 4.1.

Considere t_1, \dots, t_n uma amostra aleatória de tamanho n de $t \sim LS(\eta, \phi^2)$. O logaritmo da função de verossimilhança para o vetor $\boldsymbol{\theta} = (\eta, \phi)^\top$ é dado por

$$\mathcal{L}(\boldsymbol{\theta}) = -n \log(\phi) - \sum_{l=1}^n \log(t_l) + \sum_{l=1}^n g(\tilde{t}_l), \quad (4.1)$$

sendo $g(\tilde{t}_l) = \log h(\tilde{t}_l^2)$. O vetor escore de $\boldsymbol{\theta}$ é dado por $\mathbf{U}(\boldsymbol{\theta}) = (U_\eta(\boldsymbol{\theta}), U_\phi(\boldsymbol{\theta}))^\top$ sendo

$$U_\eta(\boldsymbol{\theta}) = \frac{1}{\eta\phi^2} \log \left[\prod_{l=1}^n \left(\frac{t_l}{\eta} \right)^{w(\tilde{t}_l)} \right], \quad U_\phi(\boldsymbol{\theta}) = -\frac{n}{\phi} + \sum_{l=1}^n w(\tilde{t}_l) \tilde{t}_l^2,$$

com $w_l = -2d \log h(u)/du|_{u=\tilde{t}_l^2}$ para $l = 1, \dots, n$. Os estimadores de máxima verossimilhança (EMV) de η e ϕ , denotados por $\hat{\eta}$ e $\hat{\phi}$, respectivamente, são obtidos resolvendo simultaneamente as equações $U_\eta(\boldsymbol{\theta}) = 0$ e $U_\phi(\boldsymbol{\theta}) = 0$. Para cada l , w_l pode ser interpretado como o peso induzido pela função $h(\cdot)$ da l -ésima observação nas estimativas dos parâmetros. Na Tabela 4.1 apresentamos w_l para algumas distribuições log-simétricas. Note que a distribuição log-normal assume o mesmo peso para todas as observações ($w_l = 1, l = 1, \dots, n$) na estimação dos parâmetros η e ϕ . Para as distribuições log-Cauchy, log-t-Student, log-logística II e log-exponencial potência ($0 < k < 1$), os pesos w_l são funções decrescentes em $|t_l|$. Isso significa dizer que os EMV de η e ϕ são robustos à presença de outliers no sentido de que observações com um valor grande para $|t_l|$ têm pesos menores. Nas distribuições log-

4.2. A classe de distribuições log-simétricas

logística I e log-exponencial potência ($-1 < k < 0$), os pesos w_l são funções crescentes em $|t_l|$ pois essas distribuições têm cauda diereita mais leve do que a da distribuição log-normal.

Tabela 4.1: Função geradora de densidades e expressões de w , $\alpha_{2,0}$ e $\alpha_{2,2}$ para algumas distribuições log-simétricas.^a

Modelo	$h(u), u > 0$	w	$\alpha_{2,0}$	$\alpha_{2,2}$
log-normal	$\frac{1}{\sqrt{2\pi}}e^{-u/2}$	1	-1	-1
log-t-Student	$\frac{\nu^{\nu/2}}{B(1/2, \nu/2)}(\nu + u)^{-\frac{\nu+1}{2}},$ $\nu > 0$	$\frac{\nu+1}{\nu+t^2}$	$-\frac{\nu+1}{\nu+3}$	$\frac{3-\nu}{\nu+3}$
log-logística I	$c\frac{e^{-u}}{(1+e^{-u})^2}, c \cong 1.4843$	$2 \tanh(t^2/2)$	$\cong -1, 4772$	$\cong -2, 0138$
log-logística II	$\frac{e^{-\sqrt{u}}}{(1+e^{-\sqrt{u}})^2}$	$\frac{e^{- t }-1}{ z (1+e^{- t })}$	$-\frac{1}{3}$	$\cong -0, 4299$
log-exponencial potência	$\frac{1}{C(k)} \exp\{-\frac{1}{2}u^{1/(1+k)}\},$ $-1 < k \leq 1$	$\frac{1}{(1+k)t^{2k/(1+k)}}$	$-2^{1-k} \frac{\Gamma(\frac{3-k}{2})}{(1+k)^2\Gamma(\frac{1+k}{2})}$	$-\frac{k-1}{k+1}$

^a $B(\cdot, \cdot)$ e $\Gamma(\cdot)$ são as funções beta e gamma, respectivamente, e $C(k) = \Gamma(1 + \frac{1+k}{2})2^{1+(1+k)/2}$.

A matriz informação de Fisher para $\boldsymbol{\theta}$, denotada por $\mathcal{K}(\boldsymbol{\theta})$, é bloco diagonal e é dada por

$$n^{-1}\mathcal{K}(\boldsymbol{\theta}) = \begin{bmatrix} -\alpha_{2,0}/\eta^2\phi^2 & 0 \\ 0 & (1 - \alpha_{2,2})/\phi^2 \end{bmatrix},$$

em que $\alpha_{a,b} = E(g^{(a)}(z)z^b)$ para $a, b = 0, 1, 2, 3$ com $g^{(r)} = d^r g(z)/dz^r$ e $z \sim S(0, 1)$. Portanto, η e ϕ são globalmente ortogonais e seus EMV são assintoticamente não correlacionados. Para algumas distribuições log-simétricas $\alpha_{2,0}$ e $\alpha_{2,2}$ são dados na Tabela 4.1. O método scoring de Fisher pode ser usado para estimação simultânea dos parâmetros pela solução iterativa das equações

$$\begin{aligned} \eta^{(m+1)} &= \eta^{(m)} + \frac{\eta^{(m)}}{n\alpha_{2,0}} \sum_{l=1}^n w(\tilde{t}_l^{(m)}) \log\left(\frac{t_l}{\eta^{(m)}}\right), \\ \phi^{(m+1)} &= \phi^{(m)} + \frac{\phi^{(m)}}{n(\alpha_{2,2} - 1)} \left\{ \phi^{(m)} \sum_{l=1}^n w(\tilde{t}_l^{(m)}) [\tilde{t}_l^{(m)}]^2 - n \right\}, \end{aligned}$$

em que $\tilde{t}_l^{(m)} = \log\left[(t_l/\eta^{(m)})^{1/\phi^{(m)}}\right]$. Para a distribuição log-normal, os EMV dos parâmetros

η e ϕ assumem forma fechada e podem ser utilizados como valores iniciais no processo iterativo para as demais distribuições da classe. Além disso, algumas distribuições log-simétricas podem ser obtidas como uma mistura no parâmetro de precisão da distribuição log-normal como, por exemplo, as distribuições log-t-Student e log-exponencial potência ($0 < k < 1$); ver [Andrews e Mallows \(1974\)](#) e [West \(1987\)](#), respectivamente. Portando, o algoritmo EM (Dempster et al., 1977) pode ser utilizado como um método iterativo para estimação dos parâmetros; ver [Vanegas e Paula \(2016\)](#).

Sob condições de regularidade usuais para estimação de máxima verossimilhança e inferência em grandes amostras (Cox e Hinkley, 1974, Cap. 9) temos que

$$\begin{pmatrix} \hat{\eta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_2 \left[\begin{pmatrix} \eta \\ \phi \end{pmatrix}; \begin{pmatrix} -\eta^2\phi^2/\alpha_{2,0} & 0 \\ 0 & \phi^2/(1-\alpha_{2,2}) \end{pmatrix} \right]. \quad (4.2)$$

Assim, a partir de (4.2), podemos construir intervalos de confiança assintóticos para os parâmetros η e ϕ . Para mais detalhes sobre as propriedades estatísticas e métodos de estimação clássico e bayesiano nessa classe de distribuição ver [Vanegas e Paula \(2016\)](#).

4.3 Métodos de redução de viés

Sejam y_1, \dots, y_n variáveis aleatórias independentes e identicamente distribuídas com distribuição indexada por um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ e $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança para uma única observação. Sejam $\kappa_{rs} = E(\partial^2 \ell / \partial \theta_r \partial \theta_s)$, $\kappa_{rst} = E(\partial^3 \ell / \partial \theta_r \partial \theta_s \partial \theta_t)$, $\kappa_{st,u} = E(\partial^2 \ell / \partial \theta_s \partial \theta_t \partial \ell / \partial \theta_u)$, $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \theta_t$, $\kappa_{rs}^{(tu)} = \partial^2 \kappa_{rs} / \partial \theta_t \partial \theta_u$, e assim por diante. O viés do estimador de máxima verossimilhança de θ_r , $r = 1 \dots, p$, pode ser escrito como

$$E(\hat{\theta}_r - \theta) = \frac{B_{\hat{\theta}_r}(\boldsymbol{\theta})}{n} + \mathcal{O}(n^{-2}).$$

Assim, o termo $B_{\hat{\theta}_r}(\boldsymbol{\theta})/n = \mathcal{O}(n^{-1})$ e é chamado de viés de segunda ordem do estimador $\hat{\theta}_r$. [Cox e Snell \(1968\)](#) obtiveram uma expressão geral para esse viés dada por

$$B_{\hat{\theta}_r}(\boldsymbol{\theta}) = \sum_{s,t,u} \kappa^{r,s} \kappa^{t,u} \left[\frac{1}{2} \kappa_{stu} + \kappa_{st,u} \right], \quad r = 1, \dots, p, \quad (4.3)$$

em que s, t e u variam de 1 até p e $\kappa^{r,s}$ denota o elemento (r, s) da inversa da matriz de informação de Fisher. A expressão (4.3) é bastante geral e pode ser particularizada para cada modelo específico de interesse. Para isso, é necessário obter a inversa da matriz de informação de Fisher e os cumulantes κ_{stu} e $\kappa_{st,u}$ com relação a todos os parâmetros. Em muitas situações é conveniente substituir, como consequência das identidades de Bartlett, $\frac{1}{2} \kappa_{stu} + \kappa_{st,u}$ por $\kappa_{st}^{(u)} - \frac{1}{2} \kappa_{stu}$. Consequentemente, definimos um estimador de máxima verossimilhança corrigido pelo viés por

$$\tilde{\theta}_r = \hat{\theta}_r - \frac{B_{\hat{\theta}_r}(\hat{\boldsymbol{\theta}})}{n}, \quad r = 1, \dots, p$$

sendo $B_{\hat{\theta}_r}(\hat{\boldsymbol{\theta}})$ o viés de segunda ordem de $\hat{\theta}_r$ avaliado em $\hat{\boldsymbol{\theta}}$. O estimador $\tilde{\theta}_r$ é tal que $E(\tilde{\theta}_r) = \theta_r + \mathcal{O}(n^{-2})$.

No caso particular das distribuições log-simétricas, $\boldsymbol{\theta} = (\eta, \phi)^\top$ e adotamos a seguinte notação: $\kappa_{\eta\eta} = E(\partial^2 \ell / \partial \eta^2)$, $\kappa_{\phi\phi} = E(\partial^2 \ell / \partial \phi^2)$, $\kappa_{\eta\phi} = E(\partial^2 \ell / \partial \eta \partial \phi)$, $\kappa_{\eta\phi}^{(\eta)} = \partial \kappa_{\eta\phi} / \partial \eta$, $\kappa_{\eta\phi}^{(\phi)} = \partial \kappa_{\eta\phi} / \partial \phi$, etc. Usando (4.3) e a ortogonalidade entre os parâmetros η e ϕ , podemos mostrar que o viés de segunda ordem dos EMV de η e ϕ são, respectivamente,

$$B_{\hat{\eta}}(\boldsymbol{\theta}) = \kappa^{\eta,\eta} \kappa^{\eta,\eta} \left[\kappa_{\eta\eta}^{(\eta)} - \frac{1}{2} \kappa_{\eta\eta\eta} \right] + \kappa^{\eta,\eta} \kappa^{\phi,\phi} \left[\kappa_{\eta\phi}^{(\phi)} - \frac{1}{2} \kappa_{\eta\phi\phi} \right] \quad (4.4)$$

e

$$B_{\hat{\phi}}(\boldsymbol{\theta}) = \kappa^{\phi,\phi} \kappa^{\phi,\phi} \left[\kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right] + \kappa^{\phi,\phi} \kappa^{\eta,\eta} \left[\kappa_{\phi\eta}^{(\eta)} - \frac{1}{2} \kappa_{\phi\eta\eta} \right]. \quad (4.5)$$

Após manipulações algébricas (ver detalhes técnicos na Seção 4.7) obtemos

$$B_{\hat{\eta}}(\boldsymbol{\theta}) = -\frac{\eta\phi^2}{2\alpha_{2,0}}, \quad B_{\hat{\phi}}(\boldsymbol{\theta}) = \frac{\phi}{2(\alpha_{2,2} - 1)} \left(\frac{6\alpha_{2,2} + \alpha_{3,3} - 4}{\alpha_{2,2} - 1} + \frac{\alpha_{3,1}}{\alpha_{2,0}} - 2 \right). \quad (4.6)$$

Segue de (4.6) que o viés de $\hat{\eta}$ depende da distribuição assumida para a variável aleatória t através da quantidade $\alpha_{2,0}$ e pode ser grande para valores pequenos de $\alpha_{2,0}$ e n . Além disso, o viés de $\hat{\phi}$ também depende da distribuição assumida para a variável aleatória t através das quantidades $\alpha_{2,0}$, $\alpha_{2,2}$, $\alpha_{3,3}$ e $\alpha_{3,1}$. Para algumas distribuições da classe de distribuições log-simétricas as quantidades $\alpha_{2,0}$ e $\alpha_{2,2}$ são dadas na Tabela 4.1 e as quantidades $\alpha_{3,1}$ e $\alpha_{3,3}$ são apresentadas na Tabela 4.2. Ainda da expressão (4.6), definimos os estimadores de máxima verossimilhança corrigidos pelo viés de ordem $\mathcal{O}(n^{-1})$, $\tilde{\eta}$ e $\tilde{\phi}$, como

$$\tilde{\eta} = \hat{\eta} - \frac{B_{\hat{\eta}}(\hat{\boldsymbol{\theta}})}{n}, \quad \tilde{\phi} = \hat{\phi} - \frac{B_{\hat{\phi}}(\hat{\boldsymbol{\theta}})}{n}. \quad (4.7)$$

Como $E(\tilde{\eta}) = \eta + \mathcal{O}(n^{-2})$ e $E(\tilde{\phi}) = \phi + \mathcal{O}(n^{-2})$ espera-se que, em amostras finitas, os estimadores $\tilde{\eta}$ e $\tilde{\phi}$ apresentem vieses menores do que os estimadores $\hat{\eta}$ e $\hat{\phi}$, cujos vieses são de ordem $\mathcal{O}(n^{-1})$. Se $\hat{\phi}^2 > -2n\alpha_{2,0}$ então $\tilde{\eta}$ não pertence ao espaço paramétrico $(0, \infty)$. Por exemplo, para o modelo log-normal $\alpha_{2,0} = -1$, portanto, se $\hat{\phi}^2 > 2n$ a estimativa de $\tilde{\eta}$ pode não pertencer ao espaço paramétrico de η para valores de n fixados.

Tabela 4.2: $\alpha_{3,1}$ e $\alpha_{3,3}$ para algumas distribuições log-simétricas.^a

Modelo	$\alpha_{3,1}$	$\alpha_{3,3}$
log-normal	0	0
log-t-Student	$\frac{6(\nu+1)}{(\nu+3)(\nu+5)}$	$\frac{6(3\nu-5)}{(\nu+3)(\nu+5)}$
log-logística I	-1, 2792	-0, 5089
log-logística II	$\frac{1}{6}$	0, 6449
log-exponencial potência	$\frac{k\Gamma[(3-k)/2]}{2^{k-2}(1+k)^3\Gamma[(k+1)/2]}$	$\frac{2k(1-k)}{(1+k)^2}$

^a Para a distribuição log-exponencial potência, $-1 < k < 1/3$.

Um outro método para corrigir o viés de ordem $\mathcal{O}(n^{-1})$ do EMV foi proposto por [Firth \(1993\)](#). Esse método consiste em uma modificação adequada na função escore de modo que o termo de primeira ordem na expansão assintótica do viés seja removido. Seja $U_{\theta_r}(\boldsymbol{\theta}) = \sum_{i=1}^n \partial \ell_i(\boldsymbol{\theta}) / \partial \theta_r$, para $r = 1, \dots, p$, em que $\ell_i(\boldsymbol{\theta})$ é o logaritmo da função de verossimilhança

para a i -ésima observação. Firth (1993) propôs uma função escore modificada

$$\ddot{U}_{\theta_r}(\boldsymbol{\theta}) = U_{\theta_r}(\boldsymbol{\theta}) + A_{\theta_r}(\boldsymbol{\theta}), \quad (4.8)$$

tal que a solução simultânea das equações $\ddot{U}_{\theta_r}(\boldsymbol{\theta}) = 0$, para $r = 1, \dots, p$, fornece estimadores $\ddot{\theta}_1, \dots, \ddot{\theta}_r$ não viesados até ordem $\mathcal{O}(n^{-2})$. Para isso, Firth (1993) mostrou que pode-se tomar $A_{\theta_r}(\boldsymbol{\theta}) = -\sum_{s=1}^p \kappa_{rs} B_{\hat{\theta}_s}(\boldsymbol{\theta})$ ou $A_{\theta_r}(\boldsymbol{\theta}) = -\sum_{s=1}^p U_{rs} \frac{B_{\hat{\theta}_s}(\boldsymbol{\theta})}{n}$, para $r = 1, \dots, p$, em que $U_{rs} = \sum_{i=1}^n \partial^2 \ell_i(\boldsymbol{\theta}) / \partial \theta_r \partial \theta_s$ é o elemento (r, s) da matriz de informação observada. Para uma discussão mais detalhada ver Kosmidis e Firth (2009, 2011). Segundo Firth (1993), o estimador $\ddot{\theta}_r$ possui as mesmas propriedades assintóticas do EMV de θ_r , isto é, $\ddot{\theta}_r$ é consistente para θ_r e $\ddot{\theta}_r \xrightarrow{\mathcal{D}} \mathcal{N}(\theta_r; \kappa^{r,r})$.

Para a classe de distribuições log-simétricas obtemos (ver detalhes técnicos na Seção 4.7)

$$\ddot{U}_{\eta}(\boldsymbol{\theta}) = U_{\eta}(\boldsymbol{\theta}) - \frac{1}{2\eta}, \quad \ddot{U}_{\phi}(\boldsymbol{\theta}) = U_{\phi}(\boldsymbol{\theta}) + \frac{1}{2\phi} \left(\frac{6\alpha_{2,2} + \alpha_{3,3} - 4}{\alpha_{2,2} - 1} + \frac{\alpha_{3,1}}{\alpha_{2,0}} - 2 \right). \quad (4.9)$$

Resolvendo as equações $\ddot{U}_{\eta}(\boldsymbol{\theta}) = 0$ e $\ddot{U}_{\phi}(\boldsymbol{\theta}) = 0$ obtemos estimadores $\ddot{\eta}$ e $\ddot{\phi}$ não viesados até ordem $\mathcal{O}(n^{-2})$ para η e ϕ , respectivamente. Em geral, a solução dessas equações são obtidas por métodos de otimização numérica. Os α 's são dados nas Tabelas 4.1 e 4.2.

Um estimador $\bar{\theta}$ é não viesado por mediana se a $P_{\theta}(\bar{\theta} \leq \theta) = P_{\theta}(\bar{\theta} \geq \theta)$ para todo θ , isto é, a mediana da distribuição do estimador $\bar{\theta}$ é θ qualquer que seja θ . Se a distribuição de $\bar{\theta}$ é contínua, $\bar{\theta}$ é não viesado por mediana se

$$P_{\theta}(\bar{\theta} \leq \theta) = 1/2,$$

para todo θ . Caso contrário, dizemos que $\bar{\theta}$ é viesado por mediana. Recentemente, com o objetivo de reduzir o viés mediano do EMV em amostras finitas, Pagui et al. (2016) propuseram uma função escore modificada, $\bar{U}_{\theta_r}(\boldsymbol{\theta})$, tal que $P_{\theta_r}(\bar{U}_{\theta_r}(\boldsymbol{\theta}) \leq 0) = \frac{1}{2} + \mathcal{O}(n^{-1})$, isto é, uma função escore não viesada por mediana até a segunda ordem. Os componentes de

$\bar{U}_{\theta_r}(\boldsymbol{\theta})$ são dados por

$$\bar{U}_{\theta_r}(\boldsymbol{\theta}) = U_{\theta_r}(\boldsymbol{\theta}) - \gamma_{rs}U_{\theta_s}(\boldsymbol{\theta}) - \nu_{1r} + \frac{\nu_{3r}}{6\nu_{2r}}, \quad r = 1, \dots, p,$$

em que

$$\nu_{1r} = -\frac{1}{2} \sum_{s,t,u=1}^p \kappa^{st} (\kappa_{r,st} - \gamma_{ru}\kappa_{u,st} + \kappa_{r,s,t} - \gamma_{ru}\kappa_{s,t,u}), \quad (4.10)$$

$$\nu_{2r} = \kappa_{r,r} - \sum_{s=1}^p \gamma_{rs}\kappa_{r,s}, \quad (4.11)$$

$$\nu_{3r} = \kappa_{r,r,r} - 3 \sum_{s=1}^p \gamma_{rs}\kappa_{r,r,s} + 3 \sum_{s,t=1}^p \gamma_{rs}\gamma_{rt}\kappa_{r,s,t} - \sum_{s,t,u=1}^p \gamma_{rs}\gamma_{rt}\gamma_{ru}\kappa_{s,t,u}, \quad (4.12)$$

com $\gamma_{rv} = \sum_{w=1}^p \kappa^{vw}\kappa_{r,w}$ e s, t, u, v e $w \in \{1, \dots, p\} \setminus \{r\}$. A solução simultânea $\bar{\theta}_1, \dots, \bar{\theta}_p$ das equações $\bar{U}_{\theta_r}(\boldsymbol{\theta}) = 0$, para $r = 1, \dots, p$, é tal que: (i) se $\bar{U}_{\theta_r}(\boldsymbol{\theta})$ é uma função monótona decrescente em θ_r então $P_{\boldsymbol{\theta}}(\bar{\theta}_r \leq \theta_r) = \frac{1}{2} + \mathcal{O}(n^{-1})$ e (ii) $\bar{\theta}_r \xrightarrow{\mathcal{D}} \mathcal{N}(\theta_r; \kappa^{r,r})$, ver [Pagui et al. \(2016\)](#). Segue de (i) que o estimador $\bar{\theta}_r$ é não viesado por mediana até a segunda ordem para θ_r .

Para a classe de distribuições log-simétricas obtemos (ver detalhes técnicos na Seção 4.7)

$$\bar{U}_{\eta}(\boldsymbol{\theta}) = U_{\eta}(\boldsymbol{\theta}), \quad \bar{U}_{\phi}(\boldsymbol{\theta}) = U_{\phi}(\boldsymbol{\theta}) - \nu_{1\phi} + \frac{\nu_{3\phi}}{6\nu_{2\phi}}, \quad (4.13)$$

sendo

$$\nu_{1\phi} = \frac{2\alpha_{2,0} + \alpha_{3,1}}{2\phi\alpha_{2,0}}, \quad \nu_{2\phi} = \frac{1 - \alpha_{2,2}}{\phi^2}, \quad \nu_{3\phi} = -\frac{2(3\alpha_{2,2} + \alpha_{3,3} - 1)}{\phi^3}.$$

Segue de (4.13) que $\bar{\eta} = \hat{\eta}$. Portanto, na classe de distribuições log-simétricas o EMV de η é não viesado por mediana até ordem $\mathcal{O}(n^{-2})$. Os α 's são dados nas Tabelas 4.1 e 4.2.

Os estimadores corrigidos pelo viés de [Cox e Snell \(1968\)](#) dependem dos EMV e podem apresentar algumas instabilidades como, por exemplo, não pertencer ao espaço paramétrico. Por outro lado, os estimadores obtidos pelos métodos de [Firth \(1993\)](#) e [Pagui et al. \(2016\)](#) penalizam a função score preventivamente e, por essa razão, não dependem das estimativas de máxima verossimilhança dos parâmetros. No entanto, no geral, necessitam da solução das

equações $\ddot{U}_{\theta_r}(\boldsymbol{\theta}) = 0$ e $\bar{U}_{\theta_r}(\boldsymbol{\theta}) = 0$ por meio de otimização numérica e, em espaços paramétricos com alta dimensão, essa tarefa pode apresentar um intensivo esforço computacional; ver Kosmidis (2014a).

Na seção seguinte, apresentamos estudos de simulação de Monte Carlo para comparar os estimadores obtidos pelos métodos descritos nessa seção.

4.4 Estudo de simulação

Nessa seção apresentamos estudos de simulação de Monte Carlo para avaliar o comportamento dos estimadores de máxima verossimilhança e dos estimadores propostos na Seção 4.3 para a classe de distribuições log-simétricas. As simulações foram realizadas utilizando a linguagem de programação Ox (Doornik 2013) que é distribuída gratuitamente para fins acadêmicos e disponível em <http://www.doornik.com>.

Consideramos diferentes tamanhos de amostras ($n = 20, 40$ e 80) e fixamos o número de réplicas Monte Carlo em 10000. Para cada tamanho de amostra, seja $\dot{\psi}_R$ a estimativa do parâmetro ψ obtida na R -ésima réplica. Calculamos:

$$\text{média} = \frac{1}{10000} \sum_{R=1}^{10000} \dot{\psi}_R \quad (\text{média}),$$

$$\text{vr} = \frac{\text{media}(\dot{\psi}) - \psi}{\psi} \quad (\text{viés relativo}),$$

$$\sqrt{\text{eqmr}} = \sqrt{\frac{1}{10000} \sum_{R=1}^{10000} \left(\frac{\dot{\psi}_R - \psi}{\psi} \right)^2} \quad (\text{raiz do erro quadrático médio relativo}),$$

$$\text{ps} = \frac{1}{10000} \sum_{R=1}^{10000} I_{\{\dot{\psi}_R \leq \psi\}} \quad (\text{probabilidade de subestimação}),$$

$$\text{pc}_{95} = \frac{1}{10000} \sum_{R=1}^{10000} I_{\{\psi \in (\dot{\psi}_R \mp 1,96\sqrt{\text{Var}_{\text{AS}}(\dot{\psi})})\}} \quad (\text{probabilidade de cobertura}),$$

para cada estimador de η e de ϕ . Aqui, $\text{Var}_{\text{AS}}(\dot{\psi})$ denota a variância assintótica estimada de

ψ e os intervalos de confiança assintóticos têm coeficiente de confiança assintótico de 95%. Como os vieses relativos dos estimadores $\hat{\eta} = \bar{\eta}$, $\tilde{\eta}$ e $\ddot{\eta}$ não dependem de η e o vieses relativos dos estimadores $\hat{\phi}$, $\tilde{\phi}$, $\ddot{\phi}$ e $\bar{\phi}$ não dependem de ϕ , fixamos os valores $\eta = 2$ e $\phi = 3$ em todos os experimentos de simulação. Os resultados são apresentados nas Tabelas 4.3 e 4.4 para a distribuição log-t-Student ($\nu = 3$ e $\nu = 5$) e nas Tabelas 4.5 e 4.6 para a distribuição log-normal.

Como podemos observar nas Tabelas 4.3 e 4.5, o viés relativo (em módulo) e a raiz do erro quadrático médio relativo dos estimadores $\tilde{\eta}$ e $\ddot{\eta}$ são menores do que o viés relativo (em módulo) e a raiz do erro quadrático médio relativo do estimador $\hat{\eta}$ em todos os tamanhos amostrais considerados. Por outro lado, os estimadores $\tilde{\eta}$ e $\ddot{\eta}$ tendem a subestimar o verdadeiro valor do parâmetro η . Por exemplo, para $n = 20$ as probabilidades de subestimação de $\hat{\eta}$ e $\ddot{\eta}$ são, respectivamente, 69,3% e 66,4% para a distribuição log-t-Student com $\nu = 3$ (Tabela 4.3), 67,8% e 65,7% para a distribuição log-t-Student com $\nu = 5$ (Tabela 4.3), e 64,8% e 64,1% para a distribuição log-normal (Tabela 4.5). Já o estimador $\hat{\eta}$ tem probabilidade de subestimação aproximadamente igual a 50%. Além disso, a probabilidade de cobertura dos intervalos que usam o estimador $\hat{\eta}$ é mais próxima do verdadeiro coeficiente de confiança considerado quando comparada com as probabilidades de cobertura dos intervalos que empregam os estimadores $\tilde{\eta}$ e $\ddot{\eta}$, em todos os tamanhos amostrais. Por exemplo, para $n = 40$ as probabilidades de cobertura correspondentes a $\hat{\eta}$, $\tilde{\eta}$ e $\ddot{\eta}$ são, respectivamente, 90,1%, 87,3% e 87,6% para a distribuição log-t-student com $\nu = 3$ (Tabela 4.3), 90,4%, 88,1% e 88,2% para a distribuição log-t-student com $\nu = 5$ (Tabela 4.3), e 91,2%, 89,1% e 89,1% para a distribuição log-normal (Tabela 4.5).

Quando comparamos o viés relativo, a raiz do erro quadrático médio relativo e a probabilidade de cobertura correspondentes aos estimadores $\hat{\phi}$, $\tilde{\phi}$, $\ddot{\phi}$ e $\bar{\phi}$ notamos que essas quantidades não diferem muito. No entanto, o estimador $\bar{\phi}$ apresentou, no geral, probabilidade de subestimação bem próxima de 50%. Por exemplo, para $n = 20$ a probabilidade de subestimação de $\hat{\phi}$, $\tilde{\phi}$, $\ddot{\phi}$ e $\bar{\phi}$ são, respectivamente, 57,4%, 54,0%, 53,1% e 50,3% para a distribuição log-t-Student com $\nu = 3$ (Tabela 4.4), 56,9%, 52,6%, 51,5% e 49,5% para a distribuição log-t-Student com $\nu = 5$ (Tabela 4.4), e 61,0%, 52,0%, 50,5% e 50,2% para a distribuição log-normal (Tabela 4.6).

4.4. Estudo de simulação

Tabela 4.3: Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc); modelo log-t-Student - estimação do parâmetro de escala.

n	Estimador	$\nu = 3$					$\nu = 5$				
		média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)	média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)
20	$\hat{\eta} = \bar{\eta}$	2,9	0,4	1,6	50,4	86,8	2,7	0,4	1,3	50,6	87,2
	$\tilde{\eta}$	1,7	-0,1	0,9	69,3	83,4	1,9	-0,1	0,9	67,8	83,6
	$\ddot{\eta}$	2,0	0,0	1,1	66,4	83,5	2,0	0,0	0,9	65,7	84,2
40	$\hat{\eta} = \bar{\eta}$	2,4	0,2	0,8	49,8	90,1	2,3	0,2	0,7	50,0	90,4
	$\tilde{\eta}$	2,0	0,0	0,7	62,1	87,3	2,0	0,0	0,6	61,2	88,1
	$\ddot{\eta}$	2,0	0,0	0,7	61,2	87,6	2,0	0,0	0,6	60,7	88,2
80	$\hat{\eta} = \bar{\eta}$	2,2	0,1	0,5	50,2	92,3	2,2	0,1	0,5	49,9	92,5
	$\tilde{\eta}$	2,0	0,0	0,5	58,5	90,3	2,0	0,0	0,4	58,0	90,9
	$\ddot{\eta}$	2,0	0,0	0,5	58,3	90,4	2,0	0,0	0,4	57,7	90,9

Tabela 4.4: Média das 10000 estimativas (media), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc); modelo log-t-Student - estimação do parâmetro de forma.

n	Estimador	$\nu = 3$					$\nu = 5$				
		média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)	média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)
20	$\hat{\phi}$	2,9	0,0	0,2	57,4	91,3	2,9	0,0	0,2	56,9	91,5
	$\tilde{\phi}$	3,0	0,0	0,2	54,0	92,0	3,0	0,0	0,2	52,6	92,4
	$\ddot{\phi}$	3,0	0,0	0,2	53,1	91,9	3,0	0,0	0,2	51,5	92,3
	$\bar{\phi}$	3,1	0,0	0,2	50,3	92,6	3,1	0,0	0,2	49,5	92,9
40	$\hat{\phi}$	3,0	0,0	0,2	55,6	93,2	3,0	0,0	0,1	54,8	93,4
	$\tilde{\phi}$	3,0	0,0	0,2	53,2	93,5	3,0	0,0	0,1	51,7	94,0
	$\ddot{\phi}$	3,0	0,0	0,2	52,9	93,6	3,0	0,0	0,1	51,5	93,9
	$\bar{\phi}$	3,0	0,0	0,2	50,5	93,9	3,0	0,0	0,1	49,9	94,1
80	$\hat{\phi}$	3,0	0,0	0,1	54,0	94,2	3,0	0,0	0,1	54,2	94,2
	$\tilde{\phi}$	3,0	0,0	0,1	52,3	94,5	3,0	0,0	0,1	51,9	94,4
	$\ddot{\phi}$	3,0	0,0	0,1	52,2	94,4	3,0	0,0	0,1	51,8	94,4
	$\bar{\phi}$	3,0	0,0	0,1	50,4	94,6	3,0	0,0	0,1	50,4	94,6

Tabela 4.5: Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc): model log-normal - estimação do parâmetro de escala.

n	Estimador	média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)
20	$\hat{\eta} = \bar{\eta}$	2,5	0,3	1,0	50,1	88,8
	$\tilde{\eta}$	2,0	0,0	0,8	64,8	86,0
	$\ddot{\eta}$	2,0	0,0	0,8	64,1	86,1
40	$\hat{\eta} = \bar{\eta}$	2,2	0,1	0,6	50,1	91,2
	$\tilde{\eta}$	2,0	0,0	0,5	59,3	89,1
	$\ddot{\eta}$	2,0	0,0	0,5	59,2	89,1
80	$\hat{\eta} = \bar{\eta}$	2,1	0,1	0,4	49,8	93,0
	$\tilde{\eta}$	2,0	0,0	0,4	56,3	91,7
	$\ddot{\eta}$	2,0	0,0	0,4	56,2	91,7

Tabela 4.6: Média das 10000 estimativas (média), viés relativo (vr), raiz do erro quadrático médio relativo ($\sqrt{\text{eqmr}}$), probabilidade de subestimação (ps) e probabilidade de cobertura (pc): model log-normal - estimação do parâmetro de forma.

n	Estimador	média	vr	$\sqrt{\text{eqmr}}$	ps (%)	pc (%)
20	$\widehat{\phi}$	2,9	0,0	0,2	61,0	89,8
	$\widetilde{\phi}$	3,0	0,0	0,2	52,0	91,7
	$\ddot{\phi}$	3,0	0,0	0,2	50,5	91,7
	$\overline{\phi}$	3,0	0,0	0,2	50,2	91,9
40	$\widehat{\phi}$	3,0	0,0	0,1	57,2	92,4
	$\widetilde{\phi}$	3,0	0,0	0,1	50,6	93,6
	$\ddot{\phi}$	3,0	0,0	0,1	50,2	93,6
	$\overline{\phi}$	3,0	0,0	0,1	49,6	93,7
80	$\widehat{\phi}$	3,0	0,0	0,1	55,9	93,7
	$\widetilde{\phi}$	3,0	0,0	0,1	51,2	94,2
	$\ddot{\phi}$	3,0	0,0	0,1	51,1	94,2
	$\overline{\phi}$	3,0	0,0	0,1	50,6	94,2

Nas Figuras 4.1 e 4.2, apresentamos as funções de densidades estimadas e assintóticas dos estimadores $\widehat{\eta} = \overline{\eta}$, $\widetilde{\eta}$ e $\ddot{\eta}$ para as distribuições log-normal e log-t-Student (com $\nu = 3$ e $n = 40$). Como podemos notar, estas distribuições não se aproximam da distribuição assintótica $N(\eta, \kappa^{\eta, \eta})$. No entanto, quando aumentamos o tamanho da amostra, a distribuição destes estimadores se aproximam da distribuição assintótica. Além disso, como mencionado anteriormente, os estimadores $\widetilde{\eta}$ e $\ddot{\eta}$ são viesados por mediana, enquanto que o estimador $\widehat{\eta} = \overline{\eta}$ apresenta probabilidade de subestimação próxima de 50%; ver Figuras 4.1 e 4.2. Por outro lado, as distribuições estimadas dos estimadores do parâmetro ϕ são bem próximas da distribuição assintótica $N(\phi, \kappa^{\phi, \phi})$ para $n = 40$. Além disso, a probabilidade de subestimação do estimador $\overline{\phi}$ é próxima de 50%, como podemos perceber nas Figuras 4.3 (modelo log-normal) e 4.4 (modelo log-t-Student com $\nu = 3$).

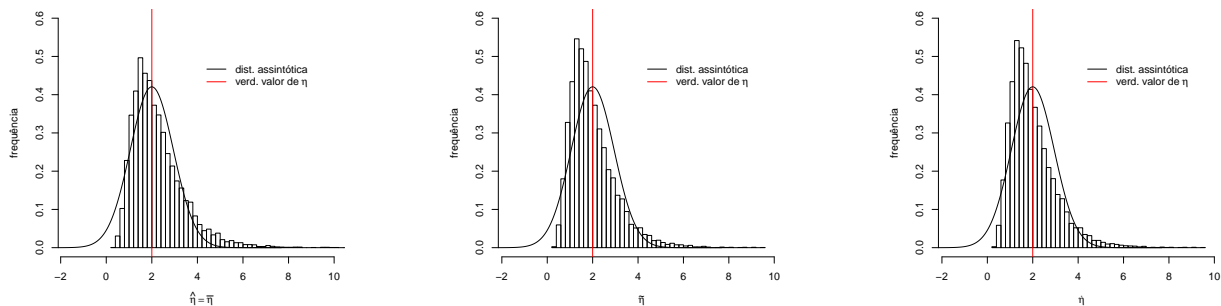


Figura 4.1: Função de densidade estimada para os estimadores de η ($n = 40$); modelo log-normal.

4.4. Estudo de simulação

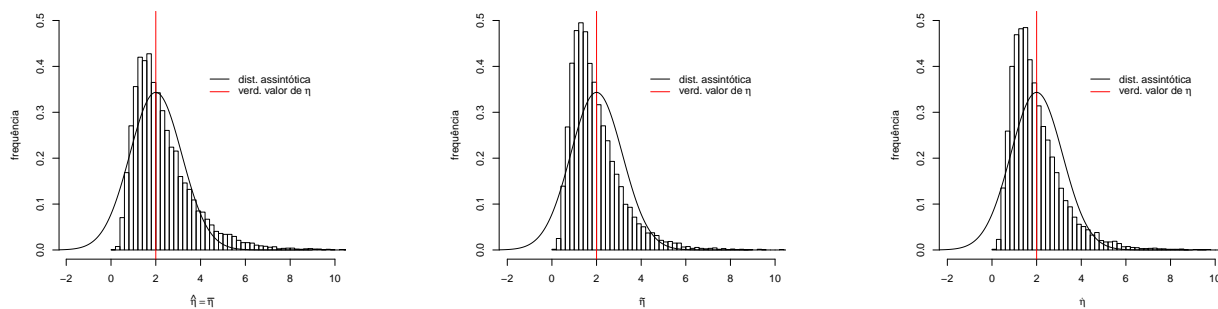


Figura 4.2: Função de densidade estimada para os estimadores de η ($n = 40$); modelo log-t-Student.

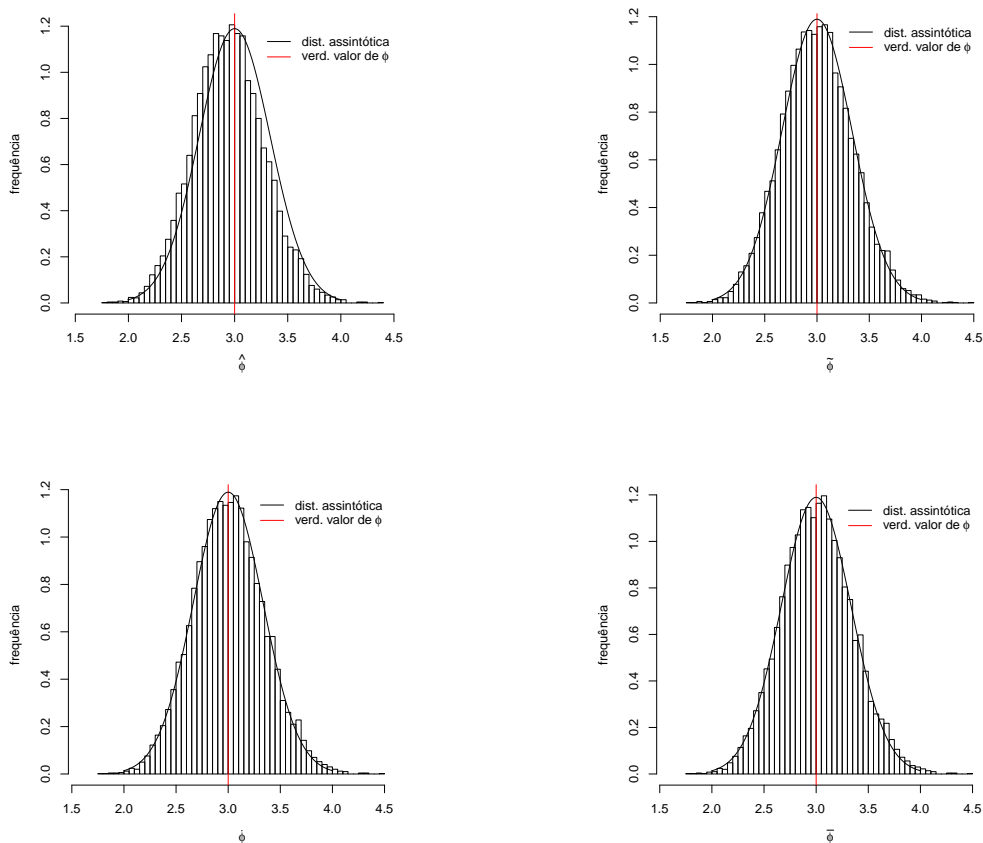


Figura 4.3: Função de densidade estimada para os estimadores de ϕ ($n = 40$); modelo log-normal.

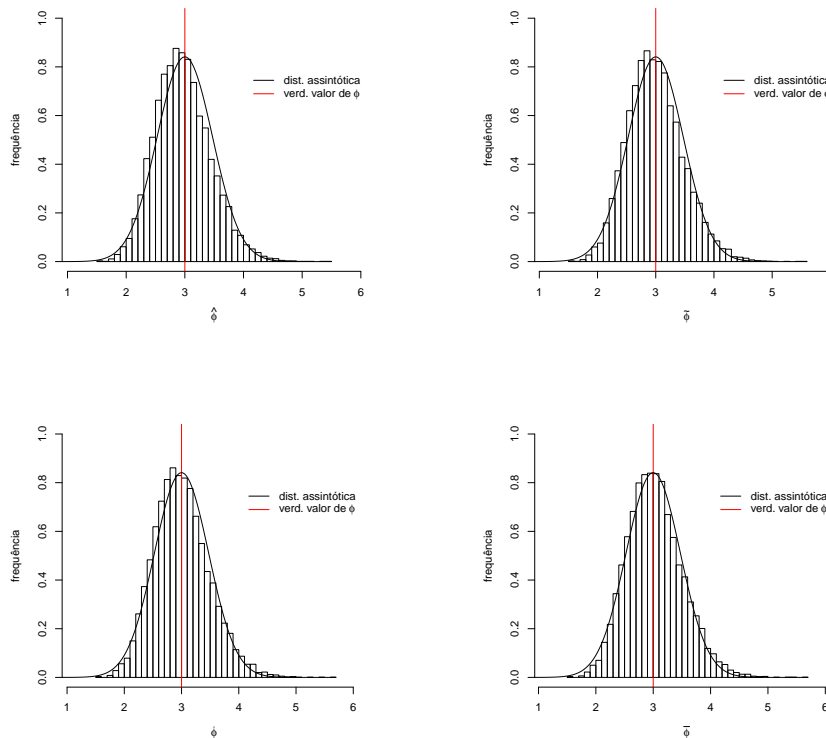


Figura 4.4: Função de densidade estimada para os estimadores de ϕ ($n = 40$); modelo *log-t-Student*.

Em resumo, de acordo com os resultados das simulações apresentados nessa seção, os estimadores $\tilde{\eta}$ e $\ddot{\eta}$ são eficazes para reduzir o viés de η porém apresentam uma alta probabilidade de subestimação e uma baixa probabilidade de cobertura. O estimador $\bar{\eta} = \hat{\eta}$ é viesado, principalmente em amostras de tamanho pequeno, mas apresenta probabilidade de subestimação em torno de 50% e probabilidade de cobertura mais próxima do coeficiente de cobertura considerado. Já os estimadores $\hat{\phi}$, $\tilde{\phi}$, $\ddot{\phi}$ e $\bar{\phi}$ apresentam pouco viés e as probabilidades de cobertura não diferem muito, porém a probabilidade de subestimação de $\bar{\phi}$ é mais próxima de 50%. Quando aumentamos o tamanho da amostra, o viés relativo e a raiz do erro quadrático médio relativo diminuem e as probabilidades de cobertura e subestimação se aproximam de 95% e 50%, respectivamente.

4.5 Aplicação

Nessa seção apresentamos uma aplicação a um conjunto de dados reais para ilustrar os resultados obtidos nesse capítulo. Para isso consideramos os dados presentes em [Mol et al.](#)

(2012) sobre a concentração do metal cádmio (t), medida em mg/kg , em 34 peixes da espécie atum albacore (*Thunnus alalunga*) capturados no Mar Mediterrâneo em 2009. Ajustamos as distribuições log-normal e log-t-Student com $\nu = 3, 5$ e 8 . O critério AIC corrigido (Burnham e Anderson, 2004) para os modelos ajustados foram: $-103,8$ (modelo log-normal), $-101,2$ (modelo $\log-t_3$) e $-102,4$ (modelo $\log-t_5$), $-103,0$ (modelo $\log-t_8$). Com base no AIC corrigido selecionamos o modelo log-normal. Nas Tabelas 4.7 e 4.8 apresentamos as estimativas pontuais (erros-padrão assintóticos em parênteses) e o intervalo de confiança assintótico (ICA) de 95% de confiança para os parâmetros η e ϕ , respectivamente.

Tabela 4.7: Estimativas pontuais e intervalares para o parâmetro η

Estimador	estimativa	ICA (95%)
$\hat{\eta} = \bar{\eta}$	0,058(0,009)	(0,042; 0,075)
$\tilde{\eta}$	0,057(0,009)	(0,041; 0,074)
$\ddot{\eta}$	0,057(0,009)	(0,041; 0,074)

Tabela 4.8: Estimativas pontuais e intervalares para o parâmetro ϕ

Estimador	estimativa	ICA (95%)
$\hat{\phi}$	0,848(0,103)	(0,648; 1,051)
$\tilde{\phi}$	0,867(0,105)	(0,666; 1,070)
$\ddot{\phi}$	0,868(0,105)	(0,667; 1,071)
$\bar{\phi}$	0,870(0,105)	(0,669; 1,073)

Na Figura 4.5 apresentamos a função densidade e a função de distribuição acumulada estimadas a partir das estimativas das Tabelas 4.7 e 4.8. Essa figura evidencia que a distribuição log-normal descreve os dados de forma satisfatória.

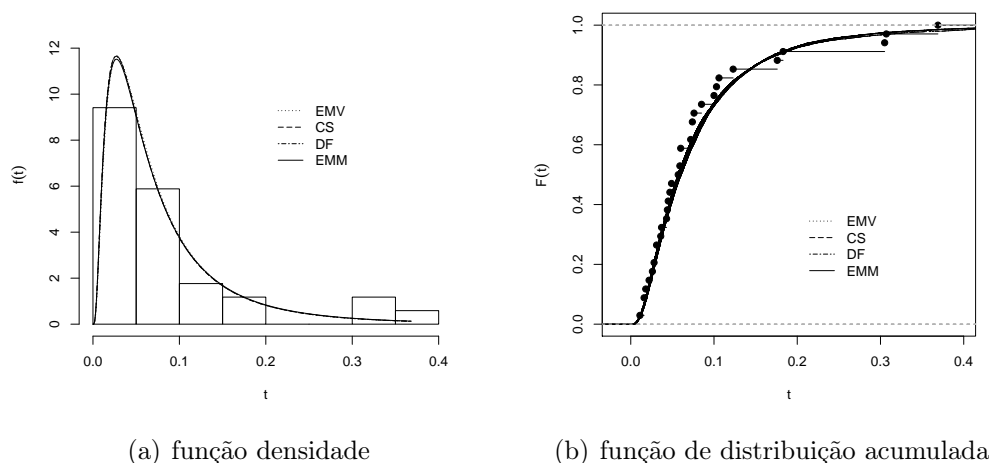


Figura 4.5: Gráficos da função densidade (a) e da função de distribuição acumulada (b) estimados.

4.6 Considerações finais

Nesse capítulo, com base nas expressões gerais de [Cox e Snell \(1968\)](#), derivamos expressões analíticas para os vieses de ordem $\mathcal{O}(n^{-1})$ dos estimadores de máxima verossimilhança e propomos estimadores corrigidos pelo viés na classe de distribuições log-simétricas. Além disso, obtivemos os vetores escore modificados pelo métodos de [Firth \(1993\)](#) e [Pagui et al. \(2016\)](#) para corrigir de forma preventiva os EMV com a finalidade de reduzir o viés e o viés mediano dos estimadores. Para a classe de distribuições consideradas mostramos que o estimador modificado de [Pagui et al. \(2016\)](#) para o parâmetro η é igual ao estimador de máxima verossimilhança. Também comparamos via simulação Monte Carlo os estimadores de máxima verossimilhança e os estimadores modificados. Uma aplicação a dados reais foi considerada para ilustrar os resultados obtidos nesse capítulo.

Nosso estudo de simulação indica que, em amostra pequenas e moderadas, os estimadores modificados reduzem o viés das estimativas exceto o estimador $\bar{\eta}$ que é igual ao estimador de máxima verossimilhança de η . Porém, as probabilidades de subestimação e de cobertura correspondentes aos estimadores $\tilde{\eta}$ e $\ddot{\eta}$ se distanciam dos verdadeiros valores considerados. Os estimadores $\tilde{\phi}$, $\ddot{\phi}$ e $\bar{\phi}$ apresentaram probabilidades de cobertura e probabilidades de subestimação próximas de 95% e 50%, respectivamente, sendo o estimador de máxima verossimilhança de ϕ o que apresentou a maior probabilidade de subestimação e a menor probabilidade de cobertura.

4.7 Detalhes técnicos

Neste apêndice obtemos as quantidades apresentadas nas expressões (4.6), (4.9) e (4.13). Para a classe de distribuições log-simétricas é possível mostrar que

$$\begin{aligned}
 \kappa_{\eta\eta} &= \frac{\alpha_{2,0}}{\phi^2\eta^2}, & \kappa_{\eta\eta}^{(\eta)} &= -\frac{2\alpha_{2,0}}{\eta^3\phi^2}, & \kappa_{\eta\eta\eta} &= -\frac{3\alpha_{2,0}}{\eta^3\phi^2}, \\
 \kappa_{\phi\phi} &= -\frac{(1-\alpha_{2,2})}{\phi^2}, & \kappa_{\phi\phi}^{(\phi)} &= \frac{2(1-\alpha_{2,2})}{\phi^3}, & \kappa_{\phi\phi\phi} &= -\frac{(6\alpha_{2,2} + \alpha_{3,3} - 4)}{\phi^3}, \\
 \kappa_{\phi\eta\eta} &= -\frac{(2\alpha_{2,0} + \alpha_{3,1})}{\eta^2\phi^3}, & \kappa_{\eta\phi} &= \kappa_{\eta\phi}^{(\phi)} = \kappa_{\phi\eta}^{(\phi)} = \kappa_{\phi\eta}^{(\eta)} = \kappa_{\eta\phi\phi} = 0.
 \end{aligned} \tag{4.14}$$

Para obtermos o viés de η e o viés de ϕ dados na expressão (4.6), basta substituir os cumulantes apresentados em (4.14) nas expressões (4.4) e (4.5). Dessa forma obtemos

$$\begin{aligned} B_{\hat{\eta}}(\boldsymbol{\theta}) &= \left(\frac{\phi^2 \eta^2}{\alpha_{2,0}} \right)^2 \left(\frac{-2\alpha_{2,0}}{\eta^3 \phi^2} + \frac{3\alpha_{2,0}}{2\eta^3 \phi^2} \right) \\ &= \left(\frac{\phi^4 \eta^4}{\alpha_{2,0}^2} \right)^2 \left(\frac{-\alpha_{2,0}}{2\eta^3 \phi^2} \right) \\ &= -\frac{\phi^2 \eta}{2\alpha_{2,0}}, \end{aligned}$$

e

$$\begin{aligned} B_{\hat{\phi}}(\boldsymbol{\theta}) &= \left(\frac{\phi^2}{(1 - \alpha_{2,2})} \right)^2 \left(\frac{2(1 - \alpha_{2,2})}{\eta^3} + \frac{(6\alpha_{2,2} + \alpha_{3,3} - 4)}{\eta^3} \right) \\ &+ \left(\frac{\phi^2}{(1 - \alpha_{2,2})} \right) \left(\frac{-\phi^2 \eta^2}{\alpha_{2,0}} \right) \left(\frac{(2\alpha_{2,0} + \alpha_{3,1})}{2\eta^2 \phi^2} \right) \\ &= -\frac{2\phi}{(\alpha_{2,2} - 1)} + \frac{\phi(6\alpha_{2,2} + \alpha_{3,3} - 4)}{2(\alpha_{2,2} - 1)^2} + \frac{\phi(2\alpha_{2,0} + \alpha_{3,1})}{2n\alpha_{2,0}(\alpha_{2,2} - 1)} \\ &= \frac{\phi}{2(\alpha_{2,2} - 1)} \left(\frac{6\alpha_{2,2} + \alpha_{3,3} - 4}{\alpha_{2,2} - 1} + \frac{\alpha_{3,1}}{\alpha_{2,0}} - 2 \right). \end{aligned}$$

Para obtermos o vetor escore modificado proposto por [Firth \(1993\)](#) para classe de distribuições log-simétricas, basta substituir as quantidades $\kappa_{\eta,\eta} = -\kappa_{\eta\eta}$ e $B_{\hat{\eta}}(\boldsymbol{\theta})$ na expressão $A_{\eta}(\boldsymbol{\theta}) = -\kappa_{\eta,\eta} B_{\hat{\eta}}(\boldsymbol{\theta})$ e $\kappa_{\phi,\phi} = -\kappa_{\phi\phi}$ e $B_{\hat{\phi}}(\boldsymbol{\theta})$ na expressão $A_{\phi}(\boldsymbol{\theta}) = -\kappa_{\phi,\phi} B_{\hat{\phi}}(\boldsymbol{\theta})$. Como isso obtemos

$$A_{\eta}(\boldsymbol{\theta}) = -\frac{1}{2\eta}, \quad A_{\phi}(\boldsymbol{\theta}) = \frac{1}{2\phi} \left(\frac{6\alpha_{2,2} + \alpha_{3,3} - 4}{\alpha_{2,2} - 1} + \frac{\alpha_{3,1}}{\alpha_{2,0}} - 2 \right).$$

O resultado segue de (4.8).

Por fim, derivamos o vetor escore modificado por mediana proposto por [Pagui et al. \(2016\)](#) para a classe de distribuições log-simétricas. Inicialmente, como $\gamma_{\eta\phi} = 0$ temos

$$\bar{U}_{\eta}(\boldsymbol{\theta}) = U_{\eta}(\boldsymbol{\theta}) - \nu_{1\eta} + \frac{\nu_{3\eta}}{6\nu_{2\eta}}. \quad (4.15)$$

Substituindo os cumulantes apresentados em (4.14) nas expressões (4.10), (4.11) e (4.12) obtemos $\nu_{1\eta} = -\frac{1}{2}\kappa^{\phi\phi}(\kappa_{\eta,\phi\phi} + \kappa_{\eta,\phi,\phi})$, $\nu_{2\eta} = \kappa_{\eta,\eta}$ e $\nu_{3\eta} = \kappa_{\eta,\eta,\eta}$. Portanto, para determinar o

vetor escore modificado dado em (4.15) basta obtermos as quantidades $\nu_{1\eta}$, $\nu_{2\eta}$ e $\nu_{3\eta}$. Segue das identidades de Bartlett e dos cumulantes dados em (4.14) que $\nu_{1\eta} = -\frac{1}{2}\kappa^{\phi\phi}\kappa_{\eta\phi\phi} = 0$, $\nu_{2\eta} = \kappa_{\eta,\eta} = -\kappa_{\eta\eta} = -\alpha_{2,0}/\phi^2\eta^2$ e $\nu_{3\eta} = \kappa_{\eta,\eta,\eta} = 2\kappa_{\eta\eta\eta} - 3\kappa_{\eta\eta}^{(\eta)} = 0$. Assim, $\bar{U}_\eta(\boldsymbol{\theta}) = U_\eta(\boldsymbol{\theta})$. Analogamente,

$$\bar{U}_\phi(\boldsymbol{\theta}) = U_\phi(\boldsymbol{\theta}) - \nu_{1\phi} + \frac{\nu_{3\phi}}{6\nu_{2\phi}},$$

sendo $\nu_{1\phi} = -\frac{1}{2}\kappa^{\eta\eta}\kappa_{\phi\eta\eta} = (2\alpha_{2,0} + \alpha_{3,1})/2\phi\alpha_{2,0}$, $\nu_{2\phi} = -\kappa_{\phi\phi} = (1 - \alpha_{2,2})/\phi^2$ e $\nu_{3\phi} = \kappa_{\phi,\phi,\phi} = 2\kappa_{\phi\phi\phi} - 3\kappa_{\phi\phi}^{(\phi)} = -2(3\alpha_{2,2} + \alpha_{3,3} - 1)/\phi^3$.

Capítulo 5

Considerações finais e pesquisas futuras

Nesta tese, derivamos uma expressão matricial para o fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de regressão lineares simétricos e verificamos que este resultado é válido para a classe dos modelos de regressão lineares log-simétricos. Além disso, derivamos um fator de correção tipo-Bartlett para a estatística gradiente na classe dos modelos de dispersão. Este fator de correção melhora a aproximação da distribuição da estatística de teste pela distribuição χ^2 de referência. Por meio de estudos de simulação Monte Carlo, comparamos a probabilidade de erro tipo I e o poder do teste gradiente corrigido com as probabilidades de erro tipo I e os poderes de outros testes propostos na literatura e de testes bootstrap. Verificamos que os testes corrigidos analiticamente e via bootstrap, incluindo o teste gradiente proposto nesta tese, apresentaram taxas de rejeição da hipótese nula mais próximas do verdadeiro nível nominal considerado sem perda de poder.

Por fim, propomos estimadores corrigidos pelo viés de [Cox e Snell \(1968\)](#) e pelo método de [Firth \(1993\)](#) com o intuito de reduzir os vieses dos estimadores dos parâmetros da classe de distribuições log-simétricas. Além disso, nesta mesma classe de distribuições, obtivemos estimadores com vies mediano reduzido utilizando o método proposto por [Pagui et al. \(2016\)](#).

Vários trabalhos de pesquisa podem ser desenvolvidos a partir dos resultados desta tese, por exemplo:

- Obter um fator de correção tipo-Bartlett para a estatística gradiente em modelos de regressão não-lineares simétricos.
- Derivar um fator de correção tipo-Bartlett para a estatística gradiente em modelos de

dispersão não-lineares.

- Investigar o viés mediano em outras classes de distribuições.
- Desenvolver pacotes no software R para os resultados obtidos nesta tese.

Esperamos ter contribuído com a pesquisa na área de teoria assintótica de alta ordem e que vários trabalhos futuros venham a ser desenvolvidos a partir dos resultados desta tese.

Referências Bibliográficas

- Andrews D.R., Mallows C.L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, **36**, 99–102. 11, 79
- Barreto-Souza, V., Vasconcellos, K.L.P. (2011). Bias and skewness in a general extreme-value regression model. *Computational Statistics & Data Analysis* **55**, 1379–1393. 75
- Barroso, L.P., Cordeiro, G.M. (2005). Bartlett corrections in heteroskedastic t regression models. *Statistics & Probability Letters* **75**, 86–96. 6
- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A* **160**, 268–282. 6, 45
- Bartlett, M.S. (1953). Aproximate confidence intervals II. *Biometrika* **40**, 306–317. 75
- Bayer, F.M., Cribari-Neto, F. (2013). Bartlett corrections in beta regression models. *Journal of Statistical Planning and Inference* **143**, 531–547. 6
- Berkane, M., Bentler, P.M. (1986). Moments of elliptically distributed random variates. *Statistics and Probability Letters* **4**, 333–335. 8
- Bull, S.B., Mark, C., Greenwood, C.M.T. (2002). A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics & Data Analysis* **39**, 57–74. 75
- Burnham, K.P., Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* **33**, 261–304.
- Carta C., Bueno J.A., Ramirez P. (2008). Statistical modelling of directional wind speeds using mixtures of von Mises distributions. *Energy Conversion and Management* **49**, 897–907. 57

- Chan, H., Chen, K., Yau, C.Y. (2014). On the Bartlett correction of empirical likelihood for gaussian long-memory time series. *Electronic Journal of Statistics* **8**, 1460–1490. [6](#)
- Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Biometrika* **74**, 265–274. [46](#)
- Cordeiro, G.M. (1987). On the corrections to the likelihood ratio statistics. *Journal of the Royal Statistical Society B* **45**, 404–413. [46](#), [56](#)
- Cordeiro, G.M., Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika* **78**, 573–582. [6](#), [13](#), [45](#), [52](#)
- Cordeiro, G.M., Ferrari, S.L.P., Paula, G.A. (1993). Improved Score Tests for Generalized Linear Models. *Journal of the Royal Statistical Society B* **55**, 661–674. [46](#)
- Cordeiro, G.M., Paula, G.A., Botter, D.A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257–274. [46](#), [51](#), [52](#)
- Cordeiro, G.M., Ferrari, S.L.P., Uribe-Opazo, M.A., Vasconcellos, K.L.P. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics & Probability Letters* **46**, 317–328. [75](#)
- Cordeiro, G.M. (2004). Corrected likelihood ratio tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation* **74**, 609–620. [6](#)
- Cordeiro, G.M., Cribari-Neto, F. (2016). *An Introduction to Bartlett Correction and Bias Reduction*. Springer, London. [2](#)
- Cribari-Neto, F., Ferrari, S.L.P. (1995). Second order asymptotics for score tests in generalised linear models. *Biometrika* **82**, 426–432. [46](#), [56](#)
- Cox, D.R., Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Serie B*, **30**, 248–275. [75](#), [79](#), [83](#), [91](#), [94](#)
- Cox, D.R., Hinkley, D.V., (1974). *Theoretical Statistics*. Chapman and Hall, London. [48](#)
- Cysneiros, A.H.M.A., Rodrigues, K.S.P., Cordeiro, G.M., Ferrari, S.L.P. (2010). Three Bartlett-type corrections for score statistics in symmetric nonlinear regression models. *Statistical Papers* **51**, 273–284. [7](#)

- Cysneiros, F.J.A., Paula, G.A., Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics and Probability Letters* **77**, 1084–1090. [5](#)
- Cysneiros, F.J.A., Cordeiro, G.M., Cysneiros, H.M.A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation* **8**, 451–461. [75](#)
- da Silva, D.N., Cordeiro, G.M. (2009). A computer program to improve LR tests for generalized linear models. *Communications in Statistics* **38**, 2184–2197. [6](#)
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1–38.
- Doornik, J.A. (2013). *Object-Oriented Matrix Programming using Ox*. Timberlake Consultants Press, London. [55](#), [56](#), [84](#)
- Fang, K., Kotz, S., Ng, K. (1990). *Symmetric Multivariate and Related Distribution*. Chapman & Hall, London. [8](#)
- Ferrari, S.L.P., Cordeiro, G.M., Cribari-Neto, F. (2001). Higher-order asymptotic refinements for score tests in proper dispersion models. *Journal of Statistical Planning and Inference* **97**, 177–190. [46](#), [51](#), [52](#)
- Ferrari, S.L.P., Uribe-Opazo, M.A. (2001). Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics* **15**, 49–67. [6](#), [12](#), [13](#), [35](#), [45](#)
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38. [74](#), [75](#), [76](#), [81](#), [82](#), [83](#), [91](#), [92](#), [94](#)
- Fisher, N. I., Lee, A.J. (1992). Regression models for an angular response. *Biometrics* **48**, 525–529. [65](#)
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press. [46](#), [65](#)
- Galea, M., Paula, G.A., Cysneiros, F.J.A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters*. **73**, 459–467. [5](#)

- Ghosh, J.K., Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *Journal of Multivariate Analysis*. **38**, 385–393.
- Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman and Hall, London. 46, 47
- Kosmidis, I., Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika* **96**, 793–804.
- Kosmidis, I., Firth, D. (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics* **4**, 1097–1112.
- Kosmidis, I., Firth, D. (2011). Multinomial logit bias reduction via the Poisson log-linear model. *Biometrika* 1–5.
- Kosmidis, I. (2014a). Bias in parametric estimation: reduction and useful side-effects. *WIRE Computational Statistics* **6**, 185–196. 84
- Kosmidis, I. (2014b). Improved estimation in cumulative link models. *Journal of the Royal Statistical Society B* **76**, 169–196. 75
- Lagos, B.M., Morettin, P.A., Barroso, L.P. (2010). Some corrections of the score test statistic for Gaussian ARMA models. *Brazilian Journal of Probability and Statistics* **24**, 434–456. 6
- Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed. Wiley, New Jersey. 58
- Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. 6, 13, 45, 52
- Lee, A. (2010). Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics*. **2**, 477–486. 46
- Lemonte, A.J., Cribari-Neto, F., Vasconcellos, K.L.P. (2007). Improved statistical inference for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics & Data Analysis* **54**, 1307–1316. 75
- Lemonte, A.J., Ferrari, S.L.P., Cribari-Neto, F. (2010). Improved likelihood inference in Birnbaum-Saunders regressions. *Computational Statistics & Data Analysis* **54**, 1307–1316. 6

- Lemonte, A.J., Ferrari, S.L.P. (2011). Small-sample corrections for score tests in Birnbaum–Saunders regressions. *Communications in Statistics – Theory and Methods* **40**, 232–243. [6](#)
- Lemonte, A.J. (2012). Local power properties of some asymptotic tests in symmetric linear regression models. *Journal of Statistical Planning and Inference* **142**, 1178–1188. [5](#)
- Lemonte, A.J., Ferrari, S.L.P. (2012a). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics* **64**, 373–381.
- Lemonte, A.J., Ferrari, S.L.P. (2012b). Local power and size properties of the LR, Wald, score and gradient tests in dispersion models. *Statistical Methodology* **9**, 537–554. [46](#), [65](#)
- Lemonte, A.J. (2016). *The Gradient Test: Another Likelihood-Based Test*. Academic Press, London. [2](#), [51](#)
- Lemonte, A.J., Cordeiro, G.M., Moreno–Arenas, G. (2016). Improved likelihood-based inference in Birnbaum–Saunders nonlinear regression models. *Applied Mathematical Modelling*. DOI:10.1016/j.apm.2016.04.007. [45](#)
- Limpert, E., Stahel, W.A., Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *BioScience* **51**, 4656–4681. [74](#)
- Liu, S.Z. (2000). On local influence for elliptical linear models. *Statistical Papers* **41**, 211–224. [5](#)
- Mardia, K.V., Jupp, P.E. (2000). *Directional Statistics*. Wiley. [46](#), [57](#)
- Melo, T.F.N., Ferrari, S.L.P., Cribari-Neto, F. (2009). Improved testing inference in mixed linear models. *Computational Statistics & Data Analysis* **53**, 2573–2582. [6](#)
- Mirhosseini, H., Tan, C.P. (2010). Discrimination of orange beverage emulsions with different formulations using multivariate analysis. *Journal of the Science of Food and Agriculture* **90**, 1308–1316. [30](#)
- Mol, S., Ozden, O., Karakulak, S. (2012). Levels of selected metals in albacore (*thunnus alalunga*, Bonnaterre, 1788) from the eastern Mediterranean. *Journal of Aquatic Food Product Technology* **21**, 111–117. [89](#)
- Nateghi, L., Roohinejad, S., Totosaus, A., Mirhosseini, H., Shuhaimi, M., Meimandipour, A., Omidzadeh, A., Manap, M.Y.A (2012). Optimization of textural properties and formulation of

- reduced fat Cheddar cheeses containing fat replacers. *Journal of Food, Agriculture & Environment* **10**, 46–54. 28
- Ortega, E.M.M., Paula, G.A. Bolfarine, H. (2008). Deviance residuals in generalised log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation* **78**, 747–764. 58
- Ortega, E.M.M., Gancho, V.G., Paula, G.A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis* **15**, 79–106. 58
- Pagui, K.E.C., Salvan, A., Sartori, N. (2016). Median bias reduction of maximum likelihood estimates. *arXiv:1604.04768v2*. 74, 76, 82, 83, 91, 92, 94
- Paula, G.A. Cysneiros, F.J.A. (2009). Systematic risk estimation in symmetric models. *Applied Economics Letters* **16**, 217–221. 5
- Rafter, J., Abell, M.L., Braselton, J.P. (2003). *Statistics With Maple*. Elsevier Science & Technology, London. 55
- Rao, B.L.S.P. (1990). Remarks on univariate symmetric distributions. *Statistics and Probability Letters* **10**, 307–315. 8
- Rao, C.R. (2005). Score test: historical review and recent developments. In: Balakrishnan, N., Kannan, N., Nagaraja, H.N. (Eds.) *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*. Birkhauser, Boston. 1
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. 55
- Rocha, A.V., Simas, A.B., Cordeiro, G.M. (2010). Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models. *Statistics and Probability Letters* **80**, 718–725. 46
- Rosillo, F.G. Martín, N. Egido, M.A. (2010). Comparison of conventional and accelerated lifetime testing of fluorescent lamps. *Lighting Research and Technology* **42**, 243–259. 66
- Rosillo, F.G. (2011). Predicción Del Tempo de Vida de Lámparas Fluorescentes Aplicación a Usos Fotovoltaicos. Tese de Doutorado, Instituto de Energía Solar, Universidad Politécnica de Madrid, Madrid. 66

- Saha, K., Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics* **61**, 179–185. 75
- Sartori, N. (2006). Bias prevention of maximum likelihood estimator for scalar skew normal and skew t distributions. *Journal Statistical Planning and Inference* **136**, 4259–4275. 75
- Sen, P.K., Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman and Hall/CRC, Boca Raton. 51
- Silva-Junior, A.H.M., da Silva, D.N., Ferrari, S.L.P. (2014). mdscore: an R package to compute improved score tests in generalized linear models. *Journal of Statistical Software* **61**, 1–16. 6
- Simas, A.B., Rocha, A.V., Barreto-Souza, W. (2011). Bias-corrected estimators for dispersion models with dispersion covariates. *Journal of Statistical Planning and Inference* **141**, 3063–3074. 46, 75
- Simas, A.B.; Cordeiro, G.M.; Nadarajah, S. (2012). Asymptotic tail properties of the distributions in the class of dispersion models. *Theory of Probability and its Applications* **56**, 703–710. 46
- Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* **34**, 206–215.
- Uribe-Opazo, M.A. (1997). Aperfeiçoamento de Testes Estatísticos em Várias Famílias de Distribuições. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. 35
- Uribe-Opazo, M.A., Ferrari, S.L.P., Cordeiro, G.M. (2008). Improved score tests in symmetric linear regression models. *Communications in Statistics–Theory and Methods* **37**, 261–276. 7, 10, 12, 13, 35, 45
- Vanegas, L.H., Paula, G.A. (2015a). A semiparametric approach for joint modeling of median and skewness. *test* **24**, 110–135. 5, 17
- Vanegas, L.H., Paula, G.A. (2015b). ssym: fitting semi-parametric log-symmetric regression models. R package version 1.5.2. <http://CRAN.R-project.org/package=ssym>. 11
- Vanegas, L.H., Paula, G.A. (2016). Log-symmetric distributions: statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics* **30**, 196–220. 5, 16, 17, 75, 77, 79

Vargas, T.M., Ferrari, S.L.P., Lemonte, A.J. (2013). Gradient statistic: higher order asymptotics and Bartlett-type correction. *Electronic Journal of Statistics* **7**, 43–61. [2](#), [6](#), [12](#), [14](#), [33](#), [35](#), [36](#), [45](#), [51](#), [54](#), [68](#), [69](#), [70](#), [72](#)

Vargas, T.M., Ferrari, S.L.P., Lemonte, A.J. (2014). Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis* **74**, 110–124. [6](#), [46](#), [56](#)

Villegas, C., Paula, G.A., Cysneiros, F.J.A., Galea, M. (2013). Influence diagnostics in generalized symmetric linear models. *Computational Statistics & Data Analysis* **59**, 161–170. [5](#)

West M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646–648. [11](#), [79](#)

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis* **13**, 157–170.

[75](#)