APERFEIÇOAMENTO DE TESTES DA
RAZÃO DE VEROSSIMILHANÇA EM
MODELOS ESPECIAIS COM ALGUMAS
APLICAÇÕES E EXTENSÕES

Elisete da Conceição Quintaneiro Aubin

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

Área de Concentração: Estatística

Orientador: Prof. Dr. Gauss Moutinho Cordeiro

– São Paulo, agosto de 1995 –

**APERFEIÇOAMENTO DE TESTES DA RAZÃO
DE VEROSSIMILHANÇA EM MODELOS
ESPECIAIS COM ALGUMAS APLICAÇÕES
E EXTENSÕES**

Este exemplar corresponde à redação final
da tese devidamente corrigida e defendida por
Elisete da Conceição Quintaneiro Aubin
e aprovada pela Comissão Julgadora.

São Paulo, 11 de agosto de 1995

Banca Examinadora:

- Prof. Dr. Gauss Moutinho Cordeiro (Orientador) - UFPE
- Prof. Dr. Pedro Alberto Morettin - IME-USP
- Prof. Dr. Ruben Klein - LNCC-CNPq
- Prof. Dr. Enrico Colosimo - UFMG
- Profa. Dra. Clarice Garcia Borges Demétrio - ESALQ-USP

Aos meus filhos

Fabio e Aline

AGRADECIMENTOS

O término de uma tese faz-nos refletir sobre tudo que passamos para alcançar este ideal. Se por um lado amargamos sentimentos como insegurança, desânimo, incerteza e cansaço, por outro, desfrutamos da amizade, carinho, confiança, solidariedade e apoio oferecidos por várias pessoas com quem convivemos.

- Primeiramente, quero expressar minha imensa gratidão ao meu orientador, **Professor Gauss Moutinho Cordeiro**, por prontamente aceitar-me como sua orientanda. Quero agradecer-lhe pela forma dedicada com que me guiou para que fossem vencidas as muitas dificuldades deste trabalho e pela forma pacienciosa, agradável e muito rápida com que sempre me atendeu. A sua reconhecida competência e o seu entusiasmo no trabalho acadêmico e na coordenação do nosso grupo de pesquisas, aliados ao seu profundo conhecimento científico, contribuíram de uma forma marcante para o desenvolvimento de minhas atividades de pesquisa, presentes e futuras.

Quero agradecer, ainda,

- à Professora **Silvia Lopes de Paula Ferrari**, colega, brilhante acadêmica, pelo seu grande apoio e pelo trabalho em conjunto na obtenção dos resultados do Capítulo 4,
- à Professora **Denise Aparecida Botter**, pela atenção em revisar a versão final deste trabalho, apresentando inúmeras sugestões,
- ao amigo Professor **Francisco Cribari-Neto**, por sua co-autoria nos resultados do Capítulo

4 e pela força e incentivo constantes que sempre me transmitiu, apesar da distância,

- aos Professores **Pedro Alberto Morettin** e **Pedro Luiz Valls Pereira**, por suas contribuições pessoais ligadas às áreas de séries temporais e econometria,
- à amiga Professora **Mariane Streibel**, por sua amizade sincera e transparente e por todo seu apoio, amparo e alegria nos momentos críticos da minha vida profissional, ajudando-me a superar obstáculos imensos,
- ao aluno de doutorado **Miguel Angel Uribe Opazo**, por sua valiosa ajuda na verificação dos resultados dos Capítulos 4 e 5 e aos alunos de iniciação científica **Alberto Pereira dos Santos** e **Ricardo Alexandre Melo da Silva**, pela valiosa ajuda na elaboração dos programas computacionais de simulação,
- ao **Waltinho**, funcionário da CPG, pelo excelente trabalho de digitação e pela paciência ao aguentar as inúmeras alterações na fase final deste trabalho,
- finalmente, à minha família, **Luiz, Fabio e Aline**, por sua amorosa compreensão e paciência e pelo sacrifício de vários fins de semana perdidos e, também, à **minha mãe**, por substituir-me inúmeras vezes nas minhas obrigações familiares.

O apoio de todas estas pessoas e de outros professores do Instituto de Matemática e Estatística da Universidade de São Paulo foi primordial para que este trabalho fosse desenvolvido e concluído com sucesso.

RESUMO

O enfoque principal deste trabalho é o aperfeiçoamento da estatística do teste da razão de verossimilhança, através do fator de correção de Bartlett, em alguns modelos especiais. Consideramos os modelos normais lineares heterocedásticos com estrutura de variância-covariância que depende de um conjunto de parâmetros desconhecidos. As fórmulas das correções são dadas em notação matricial e têm vantagens do ponto de vista numérico, podendo ser implementadas em sistemas de computação algébrica, para obter expressões em forma fechada, quando aplicadas a casos especiais. Estudamos os modelos uniparamétricos e os biparamétricos pertencentes à família exponencial. As expressões das correções de Bartlett, que obtemos, são bastante simples, envolvendo somente operações triviais com certas funções e suas derivadas. Resultados de estudos de simulação são apresentados para alguns dos casos considerados, mostrando a importância prática das fórmulas obtidas. Um tópico final abordado, diz respeito ao aperfeiçoamento de outras estatísticas, baseado na metodologia desenvolvida por Cordeiro e Ferrari (1991). Primeiramente, obtemos o aperfeiçoamento de uma estatística expressa pelo quociente entre duas variáveis aleatórias independentes, ambas convergindo para distribuições qui-quadrados centrais, até ordem n^{-1} . Finalizando, apresentamos um resultado mais geral, que constitui uma extensão às correções de tipo-Bartlett, obtendo o aperfeiçoamento de estatísticas-testes quaisquer com distribuição assintótica de primeira ordem arbitrária, conhecendo-se a expansão assintótica da distribuição da estatística original, até ordem n^{-1} ou $n^{-1/2}$. Ilustramos a utilização desta técnica através de algumas aplicações.

ABSTRACT

The improvement of the likelihood ratio test statistic through Bartlett corrections factors has been considered for special models. We study the heteroscedastic normal linear models and models belonging to the exponential family. The formulae developed for the former models are given in matrix notation leading to numerical advantages. They can be easily implemented in algebraic computing systems to obtain closed-form expressions in special cases. Simple formulae are presented for the Bartlett corrections related with uniparametric and biparametric exponential family models. The importance of the corrections when handling small samples is confirmed by Monte Carlo simulations conducted for several special cases. The final issue concerns on the Bartlett-type corrections developed by Cordeiro and Ferrari (1991). The improvement of the ratio of two independent random variables, both converging to central chi-squared distributions, up to order n^{-1} , is obtained. An extension to the generalized Bartlett corrections is presented by providing the asymptotic expansion, up to order n^{-1} or $n^{-1/2}$, for the distribution of the original statistic with an arbitrary first-order approximating distribution. The technic is illustrated with some applications.

CONTEÚDO

1. UMA REVISÃO DAS CORREÇÕES DE BARTLETT	1
1.1 Introdução	1
1.2 Uma revisão da literatura	5
1.3 Cálculo das correções	9
1.3.1 Método de Lawley	9
1.3.2 Método de McCullagh e Cox	13
1.3.3 Caso da hipótese nula simples	14
1.3.4 Correção da estatística escore	15
1.4 Método matricial	18
1.5 Correções de Bartlett generalizadas	20
2. APERFEIÇOAMENTO DO TESTE DA RAZÃO DE VEROSSIMILHANÇA PARA O MODELO NORMAL LINEAR COM ESTRUTURA DE COVARIÂNCIA PARA OS ERROS	23
2.1 Introdução	23
2.2 Modelo normal linear	24
2.3 Correção de Bartlett para os modelos normais lineares	28
2.3.1 Obtenção da quantidade ε_{p+q}	28
2.3.2 Comentários sobre a quantidade ε_{p+q}	32
2.4 Alguns casos especiais	33

CONTEÚDO

2.4.1 Dois parâmetros unidimensionais ortogonais	37
2.4.2 Parâmetro γ unidimensional	39
3. VIÉS EM MODELOS DE REGRESSÃO NORMAL LINEAR COM MATRIZ DE COVARIÂNCIA DESCONHECIDA	42
3.1 Introdução	42
3.2 Viés da estimativa do parâmetro γ	44
3.3 Ilustrações	49
3.4 Alguns estudos de simulação	53
3.4.1 Modelo heterocedástico multiplicativo	53
3.4.2 Modelo auto-regressivo de ordem 1	54
3.4.3 Modelo de médias móveis de ordem 1	57
4. CORREÇÕES DE BARTLETT PARA MODELOS NA FAMÍLIA EXPONENCIAL UNIPARAMÉTRICA	61
4.1 Introdução	61
4.2 Derivação da correção de Bartlett	63
4.3 Alguns casos especiais	67
4.4 Análise gráfica	85
4.5 Família exponencial natural	91
4.6 Alguns estudos de simulação	95
4.6.1 Distribuição geométrica	96
4.6.2 Distribuição binomial	97
4.6.3 Distribuição de Poisson truncada	99
4.6.4 Distribuição em série logarítmica	100
4.6.5 Distribuição binomial negativa	102

CONTEÚDO

4.6.6 Distribuição qui-quadrado	104
4.6.7 Distribuição de Poisson	106
5. CORREÇÕES DE BARTLETT EM MODELOS BIPARAMÉTRICOS DA FAMÍLIA	
EXPONENCIAL	108
5.1 Introdução	108
5.2 Obtenção da correção para parâmetros ortogonais	110
5.3 Dois casos especiais	114
5.4 Aplicações	117
5.4.1 Distribuição normal $N(\theta, \phi)$	118
5.4.2 Distribuição normal inversa $NI(\theta, \phi)$	119
5.4.3 Distribuição gama $G(\theta, \phi)$	120
5.4.4 Distribuição log-gama $LG(\theta, \phi)$	124
5.4.5 Distribuição gama inversa $GI(\theta, \phi)$	125
5.5 Construção de parâmetros globalmente ortogonais	127
5.6 Um estudo de simulação	139
6. CORREÇÕES DE BARTLETT PARA MODELOS NORMAIS	
HETEROCEDÁSTICOS	144
6.1 Introdução	144
6.2 Obtenção da quantidade ε_{p+q}	146
6.3 Alguns casos especiais	150
6.4 Aperfeiçoamento de testes da razão de verossimilhança	156
6.4.1 Teste de homocedasticidade	156
6.4.2 Teste de homogeneidade de médias	159
6.4.3 Teste sobre os vetores de parâmetros β e γ	161

CONTEÚDO

6.4.4	Teste sobre o vetor de parâmetros β	162
6.4.5	Teste sobre o vetor de parâmetros γ	164
6.4.6	Teste sobre um subconjunto do vetor $(\beta^\top, \gamma^\top)^\top$	165
6.4.7	Teste sobre subconjuntos de β e de γ	166
6.5	Um estudo de simulação	169
7.	UMA EXTENSÃO ÀS CORREÇÕES DE BARTLETT	172
7.1	Introdução	172
7.2	Expansão assintótica da distribuição do quociente de variáveis assintóticas qui-quadrados	173
7.3	Aperfeiçoamento do quociente entre qui-quadrados assintóticos	177
7.4	Uma extensão das correções	181
7.5	Algumas aplicações	183
CONCLUSÃO		188
APÊNDICE 1		191
APÊNDICE 2		194
APÊNDICE 3		207
APÊNDICE 4		215
REFERÊNCIAS BIBLIOGRÁFICAS		224

CAPÍTULO 1

UMA REVISÃO DAS CORREÇÕES DE BARTLETT

1.1 INTRODUÇÃO

Testes em grandes amostras são usados com bastante frequência para solucionar problemas estatísticos, pois testes exatos nem sempre existem, tornando necessária uma aproximação simples a fim de alcançarmos objetivos práticos. Estes testes são denominados na literatura, em geral, por *testes assintóticos de primeira ordem*, isto é, baseiam-se em valores críticos obtidos de uma distribuição limite conhecida. Surge, então, um problema natural de verificar se a aproximação de primeira ordem é adequada para a distribuição nula da estatística em consideração.

Os testes em grandes amostras mais conhecidos, cuja distribuição de referência é qui-quadrado, são o da razão de verossimilhança, o escore e o teste de Wald, cujas estatísticas

denotamos ao longo deste trabalho por ω , S_R e W , respectivamente. As estatísticas ω , S_R e W são equivalentes em amostras grandes, no sentido de terem a mesma distribuição assintótica e, em problemas regulares, convergem, segundo a hipótese nula, para a distribuição χ_s^2 , onde s é o número de restrições impostas pela hipótese nula. Propriedades de otimalidade assintótica destes testes podem ser encontradas em Cox e Hinkley (1974, Seção 9.3) e Engle (1984, Seção 6), por exemplo. Entretanto, a aproximação de primeira ordem, em amostras pequenas, pode não ser satisfatória, conduzindo a taxas de rejeição bastante distorcidas. Pesquisas recentes mostram como as estatísticas da razão de verossimilhança, escore e Wald podem ser modificadas de modo que as suas distribuições, segundo a hipótese nula, sejam mais próximas à χ^2 de referência do que aquelas das estatísticas não-modificadas. A idéia de aperfeiçoamento de testes foi iniciada por Bartlett (1937) para a estatística da razão de verossimilhança. O princípio básico é construir uma estatística da razão de verossimilhança modificada, digamos ω^* , obtida multiplicando-se a estatística ω por um fator escalar, que tem distribuição nula mais próxima da χ^2 de referência do que aquela de ω . O argumento utilizado por Bartlett (1937) é muito simples. Suponha, que sob uma hipótese nula composta (ou simples) H , o valor esperado da estatística da razão de verossimilhança é $E(\omega) = s\{1 + b + O(n^{-2})\}$, onde b , de ordem n^{-1} , ou é uma constante ou pode ser estimada, consistentemente, segundo H , s representa a diferença das dimensões dos espaços dos parâmetros segundo as hipóteses alternativa A e nula H e n é o tamanho da amostra. Pode ser verificado, facilmente, que a estatística modificada $\omega^* = (1 + b)^{-1}\omega$ tem valor esperado s , exceto por termos de ordem n^{-2} . A grande importância desta modificação da estatística da razão de verossimilhança vem do fato que, sob a hipótese nula, $\Pr(\omega \leq x) = \Pr(\chi_s^2 \leq x) + O(n^{-1})$, enquanto $\Pr(\omega^* \leq x) = \Pr(\chi_s^2 \leq x) + O(n^{-2})$. Esta última equação mostra que a simples correção do primeiro momento da estatística ω corrige a sua distribuição, exceto por termos de ordem n^{-2} (vide Lawley, 1956; Cordeiro, 1987). Denotando $c = 1 + b$, a estatística corrigida é, em geral, escrita como $\omega^* = c^{-1}\omega$, sendo $c = s^{-1}E(\omega)$ com $E(\omega)$ avaliado até ordem n^{-1} sob a hipótese nula. Deve-se destacar que o fator de correção c não depende do valor da estatística da razão de verossimilhança, mas pode depender de

parâmetros desconhecidos e, neste caso, eles devem ser substituídos por suas estimativas de máxima verossimilhança sob H . O fator de correção $c = 1 + b$, que aperfeiçoa a estatística da razão de verossimilhança passou a ser amplamente conhecido na literatura como *correção de Bartlett*. Bartlett mostrou que para o teste de homogeneidade de variâncias os três primeiros momentos de ω^* concordam com aqueles correspondentes da distribuição χ_s^2 , com erro de ordem $n^{-3/2}$. O próprio Bartlett (1938, 1947, 1954) aplicou esta técnica para melhorar o teste da razão de verossimilhança em vários problemas de análise multivariada. A obtenção da correção de Bartlett depende do modelo adotado para os dados. Não é fácil obter a correção em alguns problemas, já que esta necessita do cálculo de cumulantes conjuntos de derivadas da log-verossimilhança total (vide Seção 1.3). Assim, grande parte deste trabalho é destinada à obtenção de expressões em forma fechada para estes cumulantes, em modelos especiais. Encontramos, em alguns casos, fórmulas matriciais para as correções de Bartlett em generalidade, já que aplicam-se a testes de hipóteses nulas simples e compostas sobre parâmetros da regressão e a problemas com parâmetros uni e multidimensionais.

No Capítulo 2 obtemos uma fórmula geral em notação matricial para o fator de correção de Bartlett para a estatística da razão de verossimilhança na classe dos modelos normais lineares, onde os erros têm estruturas de covariância não-escalares, que dependem de um conjunto de parâmetros desconhecidos. Este resultado estende aquele apresentado por Cordeiro (1993b), que considera o modelo heterocedástico multiplicativo e o modelo variável dependente. As expressões resultantes são bastante atraentes sob o ponto de vista numérico e simbólico, pois requerem somente operações simples com matrizes e vetores. Um aspecto de igual importância a ser explorado nestes modelos são os vieses dos estimadores de máxima verossimilhança dos parâmetros em amostras de tamanho pequeno ou moderado. Assim, o objetivo principal do Capítulo 3 é encontrar uma fórmula geral para os vieses assintóticos destes estimadores e, a partir dela, conseguir expressões em forma fechada para os vieses de segunda ordem, em termos dos parâmetros, em uma variedade de modelos importantes. Uma comprovação da importância

prática da fórmula de correção dos vieses é mostrada através de alguns estudos de simulação.

No Capítulo 4 nossa atenção está voltada para as distribuições uniparamétricas na família exponencial. Esta família, além de possuir diversas propriedades matemáticas úteis, goza de vasta aplicação prática. Obtemos, para estes modelos, uma fórmula simples para a correção de Bartlett, que é algebricamente mais interessante do que aquelas gerais desenvolvidas por Lawley (1956) e McCullagh e Cox (1986). Usamos este resultado para encontrar expressões simples das correções de Bartlett em muitos casos especiais. Em outros, quando o cálculo da correção envolve funções matemáticas complicadas, fornecemos algumas aproximações que simplificam bastante suas expressões e são úteis na prática. Apresentamos alguns estudos de simulação que comprovam a importância dos nossos cálculos. Sob o mesmo enfoque, estudamos, no Capítulo 5, as correções de Bartlett para as distribuições biparamétricas na família exponencial, quando os parâmetros são ortogonais. Quando os parâmetros não são ortogonais é possível encontrar uma reparametrização que forneça parâmetros ortogonais. A construção de parâmetros ortogonais, utilizando o método de Cox e Reid (1987) é apresentada, bem como, aplicações a modelos particulares.

Em muitas situações práticas, é comum supormos, nos modelos normais lineares, que os erros aleatórios são não-correlacionados. Assim, no Capítulo 6, desenvolvemos a correção de Bartlett para testes da razão de verossimilhança nos modelos normais heterocedásticos, sendo as variâncias dependentes, através de uma função arbitrária, de um conjunto de parâmetros desconhecidos e as covariâncias nulas. Esta classe de modelos é uma generalização daquela estudada por Cordeiro (1993b), referente aos modelos normais heterocedásticos multiplicativos. A abordagem é a mesma utilizada por Cordeiro (1993b).

Finalizando, no Capítulo 7, abordamos duas extensões das correções tipo Bartlett. Tratamos do quociente de variáveis aleatórias independentes que convergem para distribuições qui-quadrados, e das variáveis aleatórias que convergem para uma distribuição de referência

arbitrária.

Para melhor entendimento do trabalho de pesquisa que desenvolvemos aqui, apresentamos, neste capítulo, alguns dos resultados que foram significativos para o crescente avanço das pesquisas sobre correções de Bartlett nos últimos anos, bem como, as pesquisas mais recentes nesta área.

Na Seção 1.2 fazemos uma revisão da literatura, mostrando, de forma sucinta o desenvolvimento histórico das pesquisas sobre correções de Bartlett. É claro que não é possível citarmos e comentarmos sobre todos os trabalhos desenvolvidos nesta área. Fornecemos, então, referências que complementam as informações dadas. Na Seção 1.3 é apresentado o cálculo das correções para a estatística da razão de verossimilhança, segundo o método de Lawley (1956) (Seção 1.3.1), o método de McCullagh e Cox (1986) (Seção 1.3.2) e para o caso da hipótese nula simples (Seção 1.3.3). Apresentamos, também, o fator de correção para a estatística escore, denominado correção tipo-Bartlett (Seção 1.3.4). A Seção 1.4 revisa o trabalho de Cordeiro (1993a) para calcular a correção de Bartlett através de fórmula matricial. Finalmente, na Seção 1.5, é revisada uma generalização das correções de Bartlett para situações mais amplas (Cordeiro e Ferrari, 1991), onde as estatísticas dos testes são equivalentes, assintoticamente, no sentido que convergem para a distribuição qui-quadrado quando n cresce.

1.2 UMA REVISÃO DA LITERATURA

A fórmula da correção de Bartlett $c = 1 + b = s^{-1}E(\omega)$, onde $E(\omega)$ é avaliado até ordem n^{-1} , mais frequentemente citada, é a fórmula original apresentada no artigo clássico de Lawley (1956). Em problemas regulares, para o teste de uma hipótese nula composta qualquer, Lawley (1956) obtém uma fórmula geral para b , em termos do que é hoje denominado de *tensores* ou *cumulantes da log-verossimilhança*, que são valores esperados de produtos de derivadas do logaritmo da função de verossimilhança. Através de um cálculo algébrico extremamente

complicado, Lawley conclui, ainda, que todos os momentos de ω^* concordam com aqueles correspondentes da distribuição de referência χ_s^2 , com erro de aproximação de ordem n^{-2} . Este resultado tem grande importância, uma vez que mostra que a simples correção do primeiro momento de ω permite obter uma estatística aperfeiçoada baseada em ω^* , cujos momentos (segundo H) concordam, até ordem n^{-1} , com aqueles da distribuição qui-quadrado de referência.

Na prática, o uso da fórmula de Lawley tem o inconveniente do cálculo de um grande número de tensores e produtos de tensores em testes envolvendo três ou mais parâmetros. Este inconveniente explica, em parte, a inexistência de qualquer publicação utilizando a fórmula de Lawley, em modelos com vários parâmetros, durante quase três décadas, de 1956 a 1983. Williams (1976) obtém correções de Bartlett em tabelas de contingência multidimensionais com estimativas de máxima verossimilhança em forma fechada, expandindo ω em série de Taylor, em vez de aplicar a equação geral de Lawley (1956) diretamente. Cordeiro (1983) foi o primeiro pesquisador a utilizar esta equação em modelos com um número arbitrário de parâmetros, após ter determinado a propriedade de invariância dos cumulantes da log-verossimilhança, segundo permutação de parâmetros, nos modelos lineares generalizados (Nelder e Wedderburn, 1972, McCullagh e Nelder, 1989).

Uma contribuição bastante significativa no aperfeiçoamento do teste da razão de verossimilhança foi dada por Hayakawa (1977), ao obter a expansão assintótica da distribuição nula de ω , até ordem n^{-1} , para o teste de uma hipótese nula composta. Hayakawa demonstra que a função de distribuição acumulada de ω , segundo H e até ordem n^{-1} , tem a forma

$$F_\omega(x) = G_s(x) + a_2 G_{s+4}(x) - (2a_2 - a_1) G_{s+2}(s) + (a_2 - a_1) G_s(x) + O(n^{-2}), \quad (1.1)$$

onde $G_m(x)$ é a função de distribuição acumulada da variável χ^2 com m graus de liberdade e a_1 e a_2 são funções extremamente complicadas (de ordem n^{-1}) de cumulantes das quatro primeiras derivadas da função de log-verossimilhança. Hayakawa conclui que, se e somente se, $a_2 = 0$, a estatística da razão de verossimilhança ω pode ser melhorada pelo fator de correção

de Bartlett $c = 1 + b$, mencionado anteriormente, tal que $\omega^* = c^{-1}\omega$ tenha distribuição χ_s^2 , até ordem n^{-1} sob a hipótese nula. Assim, a expansão (1.1) estaria em desacordo com o resultado de Lawley, isto é, ω^* tendo todos os momentos concordando com aqueles correspondentes da distribuição χ_s^2 , até ordem n^{-1} , a menos que $a_2 \equiv 0$. O conflito, aparentemente existente, entre os resultados de Lawley (1956) e Hayakawa (1977), só foi desfeito cerca de 10 anos depois, quando Harris (1986) e Cordeiro (1987), através de metodologias distintas, demonstraram que, em problemas regulares, o termo a_2 é identicamente nulo. Assim sendo, (1.1) reduz-se a

$$F_\omega(x) = G_s(x) + a_1 \{G_{s+2}(x) - G_s(x)\} + O(n^{-2}).$$

Usando a relação de recorrência $g_{s+2}(x) = xg_s(x)/s$, onde $g_m(x)$ é a função densidade de probabilidade da variável χ^2 com m graus de liberdade, segue que

$$f_\omega(x) = g_s(x) \left\{ 1 + a_1 \left(\frac{x}{s} - 1 \right) \right\} + O(n^{-2}), \quad (1.2)$$

onde $f_\omega(x) = dF_\omega(x)/dx$ é a função densidade da estatística da razão de verossimilhança ω . Utilizando (1.2) não é difícil mostrar que a densidade da estatística modificada ω^* , com ω^* tendo a forma $\omega^* = \omega/\{1 + b\}$ ou $\omega^* = \omega\{1 - b\}$, é $f_{\omega^*}(x) = g_s(x)$, até termos de ordem n^{-1} , o que comprova que a estatística da razão de verossimilhança modificada pela correção de Bartlett tem, sob a hipótese nula, distribuição idêntica à χ_s^2 , exceto por termos $O(n^{-2})$, como estabelecido, primeiramente, por Lawley. Um ponto importante que devemos destacar é que a afirmação anterior só é verdadeira se certas condições de regularidade estão satisfeitas. Detalhes destas condições podem ser encontradas em Chandra e Ghosh (1979), Chandra (1985) e Jensen (1993).

Na década de 80, mais precisamente nos últimos 12 anos, inúmeros artigos têm sido publicados apresentando correções de Bartlett em problemas específicos e para modelos amplos, bem como, descrevendo métodos alternativos de cálculo das correções de Bartlett. Cordeiro (1983, 1985, 1987) obtém fórmulas fechadas para estas correções em modelos lineares generalizados e extensões. Em particular, o resultado de Williams (1976) pode ser deduzido como um

caso particular de seus resultados. Barndorff-Nielsen e Cox (1984a) apresentam um método indireto de cálculo das correções de Bartlett em modelos paramétricos gerais, a partir de uma simples relação entre a correção e as constantes normalizadoras da distribuição do estimador de máxima verossimilhança, condicional a uma estatística ancilar exata ou aproximada. Também, Barndorff-Nielsen e Cox (1984b) investigam a distribuição de ω com relação a vários tipos de censura e regras de parada nos processos Browniano e de Poisson. Correções para modelos de seleção de covariáveis, quando a estatística da razão de verossimilhança tem forma fechada, foram apresentadas por Porteous (1985). Møller (1986) desenvolve correções para testes em modelos multivariados com matrizes de covariâncias estruturais.

Uma interpretação para a correção de Bartlett em termos de combinações invariantes de cumulantes das duas primeiras derivadas da função de log-verossimilhança é feita por McCullagh e Cox (1986) (vide Seção 1.3.2), que fornecem, também, interpretações geométricas para alguns modelos específicos. Barndorff-Nielsen e Blaesild (1986) desenvolvem um método alternativo para a obtenção das correções de Bartlett, visando à simplificação dos cálculos numéricos, em situações onde os cumulantes das derivadas da função de log-verossimilhança são obtidos facilmente sob determinada parametrização, enquanto as hipóteses a serem testadas são todas lineares sob a mesma ou sob outra parametrização. Uma forma para o valor esperado da estatística ω , $E(\omega)$, até ordem n^{-1} , invariante em relação à permutação de parâmetros, foi desenvolvida por Ross (1987) para modelos exponenciais. Cordeiro e Paula (1989) encontram uma expressão para $E(\omega)$ nos modelos não-lineares da família exponencial. Attfield (1991) e Cordeiro (1993b) mostram como corrigir os testes da razão de verossimilhança em modelos normais heterocedásticos. Cordeiro (1993a) também apresenta uma fórmula matricial para calcular correções de Bartlett em geral (vide Seção 1.4). Esta fórmula é bastante útil quando usada em conjunto com pacotes de manipulação algébrica, tais como REDUCE (Rayna, 1987), MAPLE (Abell e Braselton, 1994), MATHEMATICA (Wolfram, 1991) ou GAUSS (Gruntz e Monagan, 1993). Ele considera, ainda, o caso da parametrização ortogonal entre um parâmetro

uni-dimensional de interesse e um vetor de parâmetros de perturbação. Correções para a classe dos modelos de dispersão propostos por Jørgensen (1987) são obtidas por Cordeiro, Paula e Botter (1994). Em particular, todos os resultados de Cordeiro (1983, 1985, 1987) e Cordeiro e Paula (1989), seguem deste artigo como casos especiais. Em modelos de regressão logística, testes da razão de verossimilhança aperfeiçoados são obtidos por McCullagh (1985) e Moulton, Weissfeld e Laurent (1993).

Inúmeros outros trabalhos poderiam ser mencionados, pois pesquisas sobre correções de Bartlett têm apresentado um grande desenvolvimento nos últimos anos. Uma revisão da literatura especializada em correções é apresentada por Cribari-Neto e Cordeiro (1995) e Botter (1993, Capítulo 1). Também, excelentes revisões bibliográficas sobre correções de Bartlett para a estatística da razão de verossimilhança são encontradas nos artigos de Reid (1988), Cordeiro e Paula (1989) e Cordeiro, Paula e Botter (1994).

1.3 CÁLCULO DAS CORREÇÕES

Apresentamos, nesta seção, o cálculo das correções de Bartlett segundo duas metodologias distintas: o método devido a Lawley (1956) e o de McCullagh e Cox (1986), este último para o caso de uma hipótese nula simples. Uma reconciliação dos resultados de McCullagh e Cox (1986) e Lawley (1956) é feita, no sentido de mostrar que os dois resultados são equivalentes. Apresentamos, ainda, o teste da razão de verossimilhança e o teste escore no caso em que a log-verossimilhança depende apenas de um único parâmetro.

1.3.1 MÉTODO DE LAWLEY

Consideramos $Y_1, \dots, Y_n$ n variáveis aleatórias independentes e $y = (y_1, \dots, y_n)^T$ o vetor de valores observados destas variáveis. Seja $\ell(\beta)$ o logaritmo da função de verossimilhança

total, dado o vetor observado y , que depende de um vetor de parâmetros desconhecidos $\beta = (\beta_1^\top, \beta_2^\top)^\top$ com r componentes, sendo $\beta_1 = (\beta_1, \dots, \beta_s)^\top$ o vetor de parâmetros de interesse e $\beta_2 = (\beta_{s+1}, \dots, \beta_r)^\top$ o vetor de dimensão $(r-s)$ dos parâmetros de perturbação. Admitimos que a função $\ell(\beta)$ é regular com respeito às derivadas em relação aos componentes de β até quarta ordem. Nosso objetivo é testar a hipótese nula composta $H : \beta_1 = \beta_1^{(0)}$ contra $A : \beta_1 \neq \beta_1^{(0)}$, onde $\beta_1^{(0)}$ é um vetor especificado de dimensão s .

Sejam $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ e $\tilde{\beta}_2$ as estimativas de máxima verossimilhança irrestrita de β e restrita de β_2 a H , respectivamente. A estatística da razão de verossimilhança para o teste de H é expressa por

$$\omega = -2 \log \lambda ,$$

onde λ é o quociente entre os supremos das verossimilhanças, condicional à hipótese nula H e incondicional, ou seja, sob o espaço paramétrico total. Então, escrevemos

$$\omega = 2\{\ell(\hat{\beta}_1, \hat{\beta}_2) - \ell(\beta_1^{(0)}, \tilde{\beta}_2)\} . \quad (1.3)$$

Em amostras grandes, satisfeitas as condições de regularidade, que mencionamos na seção anterior, a distribuição de ω , bem como, das estatísticas escore S_R e de Wald W é, aproximadamente, χ_s^2 , isto é, χ^2 com s graus de liberdade.

Introduzimos, agora, a notação tradicionalmente adotada nos desenvolvimentos algébricos da correção de Bartlett. Para as derivadas da função de log-verossimilhança total em relação aos componentes do vetor de parâmetros β , denotamos $U_i = \partial \ell(\beta) / \partial \beta_i$, $U_{ij} = \partial^2 \ell(\beta) / \partial \beta_i \partial \beta_j$, etc. Consideramos a notação tensorial para os cumulantes conjuntos de derivadas da log-verossimilhança (Lawley, 1956; Hayakawa, 1977; McCullagh, 1987):

$$\begin{aligned} \kappa_{ij} &= E(U_{ij}) , \quad \kappa_{ijk} = E(U_{ijk}) , \quad \kappa_{ijkl} = E(U_{ijkl}) , \\ \kappa_{i,j} &= E(U_i U_j) , \quad \kappa_{i,jk} = E(U_i U_{jk}) , \quad \kappa_{i,jkl} = E(U_i U_{jkl}) , \\ \kappa_{ij,kl} &= E(U_{ij} U_{kl}) - \kappa_{ij} \kappa_{kl} , \quad \kappa_{i,j,kl} = E(U_i U_{jkl}) - \kappa_{i,j} \kappa_{kl} \quad \text{e} \end{aligned}$$

$$\kappa_{i,j,k,l} = E(U_i U_j U_k U_l) - \kappa_{i,j} \kappa_{k,l} - \kappa_{i,k} \kappa_{j,l} - \kappa_{i,l} \kappa_{j,k} .$$

As derivadas dos cumulantes em relação aos componentes do vetor β são denotadas por

$$\kappa_{ij}^{(k)} = \frac{\partial \kappa_{ij}}{\partial \beta_k} , \quad \kappa_{ij}^{(kl)} = \frac{\partial^2 \kappa_{ij}}{\partial \beta_k \partial \beta_l} , \quad \kappa_{ijk}^{(l)} = \frac{\partial \kappa_{ijk}}{\partial \beta_l} ,$$

etc. Todos os κ 's referem-se a um total sobre a amostra e são, em geral, de ordem n . Os sufixos na notação acima variam pelos inteiros $1, 2, \dots, r$. A matriz de informação total de Fisher para β tem elementos $\kappa_{i,j} = -\kappa_{ij}$, e a correspondente inversa tem elementos $\kappa^{i,j} = -\kappa^{ij}$.

Os cumulantes satisfazem certas relações (Lawley, 1956), denominadas *identidades de Bartlett*, que frequentemente facilitam seus cálculos, tais como, $\kappa_{i,j} = -\kappa_{ij}$, $\kappa_{ij}^{(k)} + \kappa_{i,j,k} + \kappa_{i,j,k} + \kappa_{j,ik} = 0$, $\kappa_{ijk}^{(l)} = \kappa_{ijkl} + \kappa_{l,ijk}$, $\kappa_{i,j,k} = -\kappa_{ijk} - \sum_{(3)} \kappa_{i,j,k}$, $\kappa_{i,j,k,l} = -3\kappa_{ijkl} + 2 \sum_{(4)} \kappa_{ij}^{(l)} - \sum_{(6)} \kappa_{ij}^{(kl)} + \sum_{(3)} \kappa_{ij,k,l}$ e $\kappa_{i,j,k,l} = \kappa_{ijkl} - 2\kappa_{ikl}^{(j)} + \kappa_{kl}^{(ij)} - \kappa_{ij,k,l}$, onde $\sum_{(k)}$ está representando o somatório tomado sobre todas as k permutações de índices, por exemplo, $\sum_{(3)} \kappa_{i,j,k} = \kappa_{i,j,k} + \kappa_{j,ik} + \kappa_{k,ij}$.

A estatística da razão de verossimilhança ω , em (1.3), pode ser reescrita por

$$\omega = 2\{[\ell(\hat{\beta}_1, \hat{\beta}_2) - \ell(\beta_1^{(0)}, \beta_2)] - [\ell(\beta_1^{(0)}, \tilde{\beta}_2) - \ell(\beta_1^{(0)}, \beta_2)]\} , \quad (1.4)$$

onde $\beta^\top = (\beta_1^{(0)\top}, \beta_2^\top)$ é o vetor de parâmetros verdadeiro, segundo a hipótese nula H . Lawley (1956) demonstrou que

$$2E\{\ell(\hat{\beta}_1, \hat{\beta}_2) - \ell(\beta_1^{(0)}, \beta_2)\} = r + \varepsilon_r + O(n^{-2}) , \quad (1.5)$$

onde o termo ε_r , de ordem n^{-1} , é obtido por

$$\varepsilon_r = \Sigma'(\ell_{ijkl} - \ell_{ijklmn}) , \quad (1.6)$$

sendo que Σ' representa o somatório sobre todos os componentes do vetor β , isto é, os índices i, j, k, l, m e n variam sobre os r parâmetros, e os ℓ 's têm expressões

$$\begin{aligned} \ell_{ijkl} &= \kappa^{ij} \kappa^{kl} \left\{ \frac{1}{4} \kappa_{ijkl} - \kappa_{ij}^{(l)} + \kappa_{ij}^{(kl)} \right\} , \\ \ell_{ijklmn} &= \kappa^{ij} \kappa^{kl} \kappa^{mn} \left\{ \kappa_{ikm} \left(\frac{\kappa_{jln}}{6} - \kappa_{jn}^{(l)} \right) + \kappa_{ikl} \left(\frac{\kappa_{jmn}}{4} - \kappa_{jn}^{(m)} \right) + \kappa_{ik}^{(m)} \kappa_{jn}^{(l)} + \kappa_{ik}^{(l)} \kappa_{jn}^{(m)} \right\} . \end{aligned} \quad (1.7)$$

Segundo resultados de Lawley (1956), o valor esperado de $2\{\ell(\beta_1^{(0)}, \tilde{\beta}_2) - \ell(\beta_1^{(0)}, \beta_2)\}$ segue expressão análoga àquela de $2E\{\ell(\hat{\beta}) - \ell(\beta)\}$ em (1.5), ou seja,

$$2E\{\ell(\beta_1^{(0)}, \tilde{\beta}_2) - \ell(\beta_1^{(0)}, \beta_2)\} = r - s + \varepsilon_{r-s} + O(n^{-2}), \quad (1.8)$$

com ε_{r-s} obtido a partir de (1.6), observando, agora, que o somatório Σ' naquela fórmula estende-se somente sobre os componentes do vetor β_2 , isto é, sobre $r - s$ parâmetros, já que o vetor β_1 está fixo em $\beta_1^{(0)}$. Logo, de (1.5) e (1.8) segue que, sob a hipótese nula H , o valor esperado da estatística da razão de verossimilhança ω em (1.4) é expresso por

$$E(\omega) = s + \varepsilon_r - \varepsilon_{r-s} + O(n^{-2}). \quad (1.9)$$

Podemos melhorar a aproximação da estatística ω pela distribuição χ_s^2 trabalhando com $\omega^* = c^{-1}\omega$, ao invés de ω , onde a correção de Bartlett é determinada por $c = s^{-1}E(\omega)$, sendo que $E(\omega)$ é avaliado até ordem n^{-1} , ou seja, é dado pela soma das três primeiras parcelas do lado direito de (1.9). Assim,

$$c = 1 + \frac{\varepsilon_r - \varepsilon_{r-s}}{s}. \quad (1.10)$$

A estatística corrigida ω^* tem distribuição χ_s^2 até ordem n^{-1} , sob H . Em outras palavras, o teste da razão de verossimilhança aperfeiçoado compara a estatística modificada $\omega^* = c^{-1}\omega$ com a distribuição χ_s^2 de referência. A dificuldade do aperfeiçoamento reside no cálculo de ε_r e ε_{r-s} a partir das equações (1.6) e (1.7). No caso da correção de Bartlett depender de parâmetros desconhecidos, estes devem ser substituídos pelas suas respectivas estimativas de máxima verossimilhança segundo H , mas isto não altera a ordem da aproximação obtida, ou seja, neste caso, a estatística resultante $\omega^* = \tilde{c}^{-1}\omega$ ainda tem distribuição χ_s^2 até ordem n^{-1} sob H (Lawley, 1956). No caso em que o interesse é o teste da hipótese nula simples $H : \beta = \beta^{(0)}$, o fator de correção de Bartlett c em (1.10), simplifica-se em

$$c = 1 + \frac{\varepsilon_r}{r}, \quad (1.11)$$

onde ε_r é calculado através de (1.6).

1.3.2 MÉTODO DE McCULLAGH E COX

McCullagh e Cox (1986) apresentam uma fórmula alternativa para a quantidade ε_r , envolvida no cálculo da correção de Bartlett no teste de uma hipótese nula (simples ou composta) sem fazer referência às equações (1.6) e (1.7) de Lawley. O objetivo nesta seção é mostrar a equivalência entre os resultados de McCullagh e Cox (1986) e Lawley (1956). Apresentamos a demonstração no caso uniparamétrico ($r = 1$), mas esta pode ser estendida ao caso multiparamétrico segundo a mesma linha de raciocínio. Este desenvolvimento foi apresentado por Botter (1993).

A expressão de ε_r no caso uniparamétrico, segundo Lawley (1956), é dada por (1.6) e tem a forma

$$\varepsilon_1 = \frac{1}{\kappa_{\beta,\beta}^2} \left(\frac{1}{4} \kappa_{\beta\beta\beta\beta} - \kappa_{\beta\beta\beta}^{(\beta)} + \kappa_{\beta\beta}^{(\beta\beta)} \right) - \frac{1}{\kappa_{\beta,\beta}^3} \left(\frac{5}{12} \kappa_{\beta\beta\beta}^2 - 2\kappa_{\beta\beta}^{(\beta)} \kappa_{\beta\beta\beta} + 2\kappa_{\beta\beta}^{(\beta)^2} \right). \quad (1.12)$$

Por outro lado, a fórmula de McCullagh e Cox (1986) para a correção $1 + b(\beta)$ da estatística da razão de verossimilhança, no teste $H : \beta = \beta^{(0)}$, é expressa por

$$b(\beta) = \frac{1}{12} (5\delta_3^2 - 3\delta_4) + \frac{1}{4} \frac{(\alpha_{02}\alpha_2 - \alpha_{11}^2)}{\alpha_2^3} - \frac{1}{2} \frac{(\alpha_{21}\alpha_2 - \alpha_{11}\alpha_3)}{\alpha_2^3}, \quad (1.13)$$

onde

$$\delta_3^2 = \frac{(\kappa_{\beta\beta\beta} + 3\kappa_{\beta,\beta\beta})^2}{\kappa_{\beta,\beta}^3}, \quad \delta_4 = -\frac{(\kappa_{\beta\beta\beta\beta} + 4\kappa_{\beta,\beta\beta\beta} + 3\kappa_{\beta\beta,\beta\beta} + 6\kappa_{\beta,\beta,\beta\beta})}{\kappa_{\beta,\beta}^2},$$

$$\alpha_2 = \kappa_{\beta,\beta}, \quad \alpha_3 = -\kappa_{\beta\beta\beta} - 3\kappa_{\beta,\beta\beta},$$

$$\alpha_{02} = \kappa_{\beta\beta,\beta\beta}, \quad \alpha_{11} = \kappa_{\beta,\beta\beta} \quad \text{e} \quad \alpha_{21} = \kappa_{\beta,\beta,\beta\beta}.$$

A vantagem de $b(\beta)$, em (1.13), sobre ε_1 , em (1.12), é que os termos individuais de $b(\beta)$ têm interpretação estatística. Definimos as seguintes variáveis aleatórias: $R_1 = \kappa_{\beta,\beta}^{-1/2} U$ e $R_2 = (U' - \kappa_{\beta\beta} - \gamma_{21}R_1)/\kappa_{\beta,\beta}$, onde $U = d\ell(\beta)/d\beta$ é a função escore, $U' = d^2\ell(\beta)/d\beta^2$ e $\gamma_{21} = \text{Cov}(U' - \kappa_{\beta\beta}, R_1) = \kappa_{\beta,\beta\beta}/\kappa_{\beta,\beta}^{1/2}$. Então, o primeiro termo em (1.13) corresponde a uma

medida da não-normalidade ou não-normalidade inversa da função escore U . O segundo termo é um quarto do quadrado da curvatura de Efron (1975), enquanto o terceiro é, simplesmente, $-\frac{1}{2}\text{Cov}(R_1^2, R_2)$.

Mostramos, a seguir, que a fórmula (1.13) é equivalente à (1.12). Das identidades de Bartlett, introduzidas na Seção 1.3.1, vêm

$$\kappa_{\beta,\beta\beta} = \kappa_{\beta\beta}^{(\beta)} - \kappa_{\beta\beta\beta} , \quad (1.14)$$

$$\kappa_{\beta,\beta\beta\beta} = \kappa_{\beta\beta\beta}^{(\beta)} - \kappa_{\beta\beta\beta\beta} \quad (1.15)$$

e

$$\kappa_{\beta\beta,\beta\beta} + \kappa_{\beta,\beta,\beta\beta} = \kappa_{\beta\beta}^{(\beta\beta)} + \kappa_{\beta\beta\beta\beta} - 2\kappa_{\beta\beta\beta}^{(\beta)} . \quad (1.16)$$

Substituindo as expressões para δ_3^2 , δ_4 , α_2 , α_3 , α_{02} , α_{11} e α_{21} em (1.13), chegamos a

$$b(\beta) = \frac{1}{12\kappa_{\beta,\beta}^3} \left[5\kappa_{\beta\beta\beta}^2 + 24\kappa_{\beta,\beta\beta}^2 + 24\kappa_{\beta,\beta\beta}\kappa_{\beta\beta\beta} + 3\kappa_{\beta,\beta}(\kappa_{\beta\beta\beta\beta} + 4\kappa_{\beta,\beta\beta\beta} + 4\kappa_{\beta\beta,\beta\beta} + 4\kappa_{\beta,\beta,\beta\beta}) \right] . \quad (1.17)$$

Substituindo, agora, (1.14) – (1.16) em (1.17) e rearranjando termos, obtemos

$$b(\beta) = \frac{1}{\kappa_{\beta,\beta}^3} \left(\frac{5}{12}\kappa_{\beta\beta\beta}^2 - 2\kappa_{\beta\beta}^{(\beta)}\kappa_{\beta\beta\beta} + 2\kappa_{\beta\beta}^{(\beta)^2} \right) + \frac{1}{\kappa_{\beta,\beta}^2} \left(\frac{1}{4}\kappa_{\beta\beta\beta\beta} - \kappa_{\beta\beta\beta\beta}^{(\beta)} + \kappa_{\beta\beta}^{(\beta\beta)} \right) ,$$

que coincide com a fórmula (1.12) de Lawley.

1.3.3 CASO DA HIPÓTESE NULA SIMPLES

Um caso importante diz respeito à hipótese nula simples, onde admitimos que a função de log-verossimilhança $\ell = \ell(\theta)$ depende apenas de um único parâmetro θ . Neste caso, a estatística da razão de verossimilhança tem a forma $\omega = 2[\ell(\hat{\theta}) - \ell(\theta^{(0)})]$, onde $\theta^{(0)}$ é o valor de θ especificado na hipótese nula, e o seu valor esperado é dado por $E(\omega) = 1 + \varepsilon_1 + O(n^{-2})$, com $\varepsilon_1 = \ell_{00000} - \ell_{0000000}$ calculado por (1.12). A estatística da razão de verossimilhança modificada pela

correção $1+\varepsilon_1$, isto é, $\omega^* = (1+\varepsilon_1)^{-1}\omega$, tem distribuição χ_1^2 até ordem n^{-1} . No Capítulo 4, o caso da hipótese nula simples para a família exponencial uniparamétrica de distribuições, é estudado exaustivamente e exibimos a correção de Bartlett para inúmeras distribuições pertencentes a esta família.

1.3.4 CORREÇÃO DA ESTATÍSTICA ESCORE

Os testes escore têm sido bastante utilizados em várias áreas de estatística e, como já mencionamos no início deste capítulo, em problemas regulares, a estatística escore (S_R), tal como a estatística da razão de verossimilhança (ω) e a de Wald (W), tem distribuição assintótica χ_s^2 sob a hipótese nula. Em algumas situações, o teste escore tem grande vantagem computacional sobre os demais. A estatística escore, entretanto, não admite um fator de correção simples, que independe da estatística original, como a da razão de verossimilhança. Este fato tem sido frequentemente citado na literatura (Harris, 1985, 1987; Cox, 1988; Barndorff-Nielsen, 1990). Harris (1985), seguindo as idéias de Hayakawa (1977), desenvolve uma expansão assintótica até ordem n^{-1} para a função distribuição da estatística escore. Dada a impossibilidade de se obter um fator de correção simples para S_R , Harris (1985) propõe uma modificação do valor crítico obtido da distribuição χ^2 de referência para um nível de significância nominal fixado, em vez de modificar a estatística do teste, baseada em uma expansão, até ordem n^{-1} , para os quantis de S_R . Cordeiro e Ferrari (1991) mostram que é sempre possível encontrar uma estatística escore aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} sob a hipótese nula. A estatística escore modificada S_R^* é dada por S_R multiplicada por um fator de correção que envolve um polinômio de segundo grau na própria estatística e cujos coeficientes são, em geral, funções de parâmetros desconhecidos que devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula. Este fator tem sido, freqüentemente, denominado na literatura por *correção tipo-Bartlett*, já que não é a correção de Bartlett clássica, mas uma extensão desta (vide Seção 1.5).

Mostramos, agora, como determinar a correção tipo-Bartlett para a estatística escore, sob o enfoque de Cordeiro e Ferrari (1991), uma vez que este é mais geral no sentido de permitir parâmetros de perturbação. Todo desenvolvimento aqui é feito utilizando a notação adotada na Seção 1.3.1. Nosso interesse é o teste das hipóteses $H : \beta_1 = \beta_1^{(0)}$ contra $A : \beta_1 \neq \beta_1^{(0)}$, onde $\beta = (\beta_1^\top, \beta_2^\top)^\top$ está definido na Seção 1.3.1. Seja

$$U(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = (U_1^\top(\beta_1, \beta_2), U_2^\top(\beta_1, \beta_2))^\top$$

a função escore partida da mesma forma que β . Supomos, também, que a matriz de informação total de Fisher $K(\beta) = E\{U(\beta)U(\beta)^\top\}$ e sua inversa estão divididas segundo β , ou seja, a partição $\beta = (\beta_1^\top, \beta_2^\top)^\top$ conduz às partições

$$K = K(\beta) = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \quad \text{e} \quad K^{-1} = K^{-1}(\beta) = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}, \quad (1.18)$$

para a matriz de informação e sua inversa, respectivamente.

A estatística escore S_R para testar $H : \beta_1 = \beta_1^{(0)}$ versus $A : \beta_1 \neq \beta_1^{(0)}$ é dada por

$$S_R = \tilde{U}_1^\top \tilde{K}^{11} \tilde{U}_1,$$

onde $K^{11} = K^{11}(\beta_1, \beta_2)$ é a inversa da matriz de covariância assintótica de $\hat{\beta}_1$, obtida de (1.18), e a notação com til ($\sim$), como antes, é utilizada para indicar funções avaliadas nas estimativas de máxima verossimilhança, segundo a hipótese nula, neste caso, em $\tilde{\beta} = (\beta_1^{(0)\top}, \tilde{\beta}_2^\top)^\top$. Definimos as matrizes

$$A = \begin{pmatrix} 0 & 0 \\ 0 & K_{22}^{-1} \end{pmatrix} \quad \text{e} \quad M = K^{-1} - A,$$

onde K_{22}^{-1} representa a matriz de covariância assintótica de $\tilde{\beta}_2$. Sejam F_{S_R} a função distribuição da estatística S_R , G_m a função distribuição da χ_m^2 e f_{S_R} e g_m suas respectivas funções densidades. Uma expansão assintótica para a distribuição nula da estatística escore S_R é dada por Harris (1985) como

$$\begin{aligned} F_{S_R}(x) = & G_s(x) + \frac{1}{24} \{ (A_2 - A_1 - A_3)G_s(x) + (3A_3 - 2A_2 + A_1)G_{s+2}(x) \\ & + (A_2 - 3A_3)G_{s+4}(x) + A_3G_{s+6}(x) \}, \end{aligned} \quad (1.19)$$

sendo os A_i 's funções de ordem n^{-1} dadas por

$$\begin{aligned}
 A_1 = & 3\Sigma'(\kappa_{ijk} + 2\kappa_{i,jk})(\kappa_{rst} + 2\kappa_{r,s,t})a_{ij}a_{st}m_{kr} \\
 & - 6\Sigma'(\kappa_{ijk} + 2\kappa_{i,jk})\kappa_{r,s,t}a_{ij}a_{kr}m_{st} \\
 & + 6\Sigma'(\kappa_{i,jk} - \kappa_{i,j,k})(\kappa_{rst} + 2\kappa_{r,s,t})a_{js}a_{kt}m_{ir} \\
 & - 6\Sigma'(\kappa_{i,j,k,r} + \kappa_{i,j,kr})a_{kr}m_{ij} ,
 \end{aligned} \tag{1.20}$$

$$\begin{aligned}
 A_2 = & -3\Sigma'\kappa_{i,j,k}\kappa_{r,s,t}a_{kr}m_{ij}m_{st} + 6\Sigma'(\kappa_{ijk} + 2\kappa_{i,jk}\kappa_{r,s,t}a_{ij}m_{kr}m_{st} \\
 & - 6\Sigma'\kappa_{i,j,k}\kappa_{r,s,t}a_{kt}m_{ir}m_{js} + 3\Sigma'\kappa_{i,j,k,r}m_{ij}m_{kr} ,
 \end{aligned} \tag{1.21}$$

$$A_3 = 3\Sigma'\kappa_{i,j,k}\kappa_{r,s,t}m_{ij}m_{kr}m_{st} + 2\Sigma'\kappa_{i,j,k}\kappa_{r,s,t}m_{ir}m_{js}m_{kt} , \tag{1.22}$$

com a_{ij} e m_{ij} sendo os elementos (i, j) das matrizes A e M , respectivamente.

Observamos que as expressões em (1.20) - (1.22) parecem ser funções bastante complicadas dos cumulantes κ 's da log-verossimilhança total. Entretanto, expressões em forma fechada podem ser obtidas para modelos especiais. Ferrari (1991, Capítulos 3 a 5) obtém expressões relativamente simples para os A 's para testes em diversas classes de modelos. Harris (1985) mostra, também, que os três primeiros cumulantes da estatística escore são dados por

$$\begin{aligned}
 \kappa_1(S_R) &= s + \frac{A_1}{12} + o(n^{-1}) , \\
 \kappa_2(S_R) &= 2s + \frac{A_1 + A_2}{3} + o(n^{-1}) , \\
 \kappa_3(S_R) &= 8s + 2(A_1 + A_2 + A_3) + o(n^{-1}) .
 \end{aligned}$$

Como sabemos, $\kappa_1(\chi_s^2) = s$, $\kappa_2(\chi_s^2) = 2s$ e $\kappa_3(\chi_s^2) = 8s$. Assim, determinando A_1 , A_2 e A_3 , podemos usar as expressões acima para encontrarmos os três primeiros cumulantes da estatística escore até ordem n^{-1} e compará-los com os cumulantes de uma variável aleatória χ_s^2 , que é a base de nossa aproximação de primeira ordem. Deve-se destacar que este resultado implica

que não existe uma transformação escalar baseada na estatística do teste que corrige todos cumulantes até uma certa ordem de precisão, como acontece com a correção de Bartlett para a estatística da razão de verossimilhança.

A correção a ser aplicada diretamente à estatística do teste escore é obtida por Cordeiro e Ferrari (1991). Eles mostram que a estatística modificada

$$S_R^* = S_R \{1 - (c + bS_R + aS_R^2)\} , \quad (1.23)$$

onde o fator multiplicativo entre chaves é um tipo de correção de Bartlett (função da própria estatística não modificada S_R) e os coeficientes a , b e c , de ordem n^{-1} , sendo dados por

$$\begin{aligned} a &= \frac{A_3}{12s(s+2)(s+4)} , \\ b &= \frac{A_2 - 2A_3}{12s(s+2)} , \\ c &= \frac{A_1 - A_2 + A_3}{12s} , \end{aligned} \quad (1.24)$$

é distribuída como χ_s^2 quando termos de ordem menor que n^{-1} são desprezados. Quando os A 's envolvem parâmetros desconhecidos devem ser substituídos por suas estimativas de máxima verossimilhança sob a hipótese nula H , sem que isto afete a ordem de aproximação da correção.

Cordeiro e Ferrari (1991) obtêm, também, um resultado mais geral, que consiste em aperfeiçoar qualquer estatística que é assintoticamente distribuída como χ_s^2 (vide Seção 1.5).

1.4 MÉTODO MATRICIAL

Apresentamos, nesta seção, uma fórmula matricial para o cálculo do termo ε_r da correção de Bartlett. Esta fórmula é deduzida por Cordeiro (1993a), sendo bastante útil nas aplicações, pois facilita o cálculo numérico e algébrico da correção de Bartlett em situações gerais. Ela torna-se atraente quando usada em conjunção com pacotes de computação algébrica, tais como

REDUCE, MAPLE ou MATHEMATICA ou com linguagens de álgebra linear numérica, por exemplo, GAUSS ou APL.

Definimos as matrizes quadradas de ordem r , $A^{(tu)}$, $P^{(t)}$, e $Q^{(u)}$, cujos elementos genéricos (r, s) são dados explicitamente entre chaves,

$$A^{(tu)} = \left\{ \frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \quad P^{(t)} = \{ \kappa_{rst} \} \quad \text{e} \quad Q^{(u)} = \{ \kappa_{su}^{(r)} \},$$

para $t, u = 1, \dots, r$. Notamos que $P^{(t)}$ é simétrica, mas as matrizes $A^{(tu)}$ e $Q^{(u)}$ não são. Consideramos, ainda, as seguintes matrizes quadradas de ordem r definidas a partir das anteriores, onde K^{-1} é a inversa da matriz de informação: L , M_1 , M_2 e M_3 são matrizes com os correspondentes (r, s) -ésimos elementos dados por $\text{tr}(K^{-1}A^{(rs)})$, $\text{tr}(K^{-1}P^{(r)}K^{-1}P^{(s)})$, $\text{tr}(K^{-1}P^{(r)}K^{-1}Q^{(s)\top})$ e $\text{tr}(K^{-1}Q^{(r)}K^{-1}Q^{(s)})$, respectivamente; N_1 , N_2 e N_3 são matrizes expressas pelos seus respectivos elementos genéricos (r, s) entre chaves, $N_1 = \{ \text{tr}(P^{(r)}K^{-1})\text{tr}(P^{(s)}K^{-1}) \}$, $N_2 = \{ \text{tr}(P^{(r)}K^{-1})\text{tr}(Q^{(s)}K^{-1}) \}$ e $N_3 = \{ \text{tr}(Q^{(r)}K^{-1})\text{tr}(Q^{(s)}K^{-1}) \}$.

A partir dessas definições, Cordeiro (1993a) mostra que a quantidade ε_r , envolvida no cálculo da correção de Bartlett, pode ser expressa, matricialmente, por

$$\varepsilon_r = \text{tr}[K^{-1}(L - M - N)], \quad (1.25)$$

onde $M = -\frac{1}{6}M_1 + M_2 - M_3$ e $N = -\frac{1}{4}N_1 + N_2 - N_3$. Muito embora a fórmula (1.6) pareça mais fácil de ser usada do que a expressão matricial (1.25), esta última é mais adequada para cálculo algébrico ou numérico da correção de Bartlett em modelos paramétricos gerais, pois envolve apenas operações com matrizes quadradas de ordem r , evitando a manipulação com tensores de ordem superior a 2. O cálculo de ε_r em (1.25) exige $(r+1)^2$ matrizes quadradas de ordem r : K^{-1} , r matrizes $P^{(t)}$, r matrizes $Q^{(u)}$ e r^2 matrizes $A^{(tu)}$. As outras matrizes envolvidas são reproduzidas a partir destas matrizes através de simples operações matriciais. Em muitas aplicações da fórmula acima, as matrizes requeridas podem ser dadas em forma fechada, explorando estruturas especiais dos cumulantes envolvidos. Esta é a situação, por exemplo, dos

casos especiais discutidos no Capítulo 2. Assim, conhecendo-se os cumulantes κ 's, a equação (1.25) descreve uma maneira fácil de se obterem as correções de Bartlett na prática. Esta abordagem é, também, considerada por Ferrari e Cordeiro (1994), que expressam a correção tipo-Bartlett num teste de uma hipótese nula composta, para a estatística escore, através de fórmulas matriciais simples. Suas fórmulas, envolvendo somente operações elementares entre matrizes e vetores, apresentam vantagens computacionais em relação às equações algébricas (1.20) - (1.22), provenientes da expansão assintótica de Harris (1985), para a distribuição da estatística escore sob a hipótese nula.

1.5 CORREÇÕES DE BARTLETT GENERALIZADAS

O problema de desenvolver correções similares à correção de Bartlett para outras estatísticas, equivalentes à da razão de máxima verossimilhança, no sentido de terem distribuição assintótica qui-quadrado, foi sugerido por Cox (1988) e resolvido, três anos depois, em generalidade, por Cordeiro e Ferrari (1991) e por Chandra e Mukerjee (1991) e Taniguchi (1991) para certos casos especiais. Para que correções sejam obtidas através da fórmula de Cordeiro e Ferrari (1991) é necessário conhecer a expansão assintótica da distribuição da estatística não modificada, sob a hipótese nula, até ordem n^{-1} , e que certas suposições de regularidade sejam satisfeitas. Uma vez que as expansões assintóticas das distribuições das estatísticas da razão de verossimilhança e escore, sob hipóteses nulas simples e compostas, e da estatística de Wald, sob hipóteses nulas simples, são conhecidas até esta ordem (Hayakawa, 1977; Hayakawa e Puri, 1985; Harris, 1985), fatores de correção para estas estatísticas podem ser calculados através da fórmula proposta. O fator de correção de Bartlett clássico c que aperfeiçoa a estatística da razão de verossimilhança é um caso particular desta fórmula. Em vista disto, este fator, que melhora a aproximação das distribuições das estatísticas assintoticamente qui-quadrado sob a hipótese nula, passa a ser interpretado como um *fator de correção de Bartlett generalizado*, para o qual apresentamos, a seguir, uma breve descrição de seu desenvolvimento.

Seja S uma estatística qualquer cuja distribuição assintótica é χ_q^2 . Supondo condições gerais de regularidade (vide Chandra, 1985 e Chandra e Goshi, 1979), Chandra (1985) demonstra que a distribuição acumulada de S até ordem n^{-1} pode ser escrita como

$$F_S(x) = P(S \leq x) = G_q(x) + \sum_{i=0}^k a_i G_{q+2i}(x), \quad (1.26)$$

onde os a_i 's são quantidades de ordem n^{-1} , expressas como funções complicadas de tensores e, em geral, dependendo de parâmetros desconhecidos. Aqui usamos a mesma notação da Seção 1.2, em que G_m e g_m representam as funções distribuição e densidade da χ_m^2 , respectivamente. Para que a função F_S em (1.26) seja uma função distribuição até essa ordem é necessário que a condição $\sum_{i=0}^k a_i = 0$ seja satisfeita. A equação (1.26) mostra que a distribuição, até ordem n^{-1} , de qualquer estatística que tenha distribuição assintótica qui-quadrado, é dada por uma combinação linear de qui-quadrados com graus de liberdade $q, q+2, \dots, q+2k$. A fórmula (1.1) é um caso especial de (1.26). Para as estatísticas escore e de Wald temos $k = 3$, enquanto que para a da razão de verossimilhança temos $k = 1$.

A partir da equação (1.26) e usando as fórmulas de recorrência

$$G_{m+2}(x) = G_m(x) - \frac{2x}{m} g_m(x), \quad g_{m+2}(x) = \frac{x}{m} g_m(x), \quad (1.27)$$

entre as funções de distribuição acumuladas e função densidade qui-quadrados com m e $m+2$ graus de liberdade, Cordeiro e Ferrari (1991) mostram que a estatística modificada

$$S^* = S \left\{ 1 - 2 \sum_{i=1}^k \left(\sum_{j=i}^k a_j \right) \mu_i'^{-1} S^{i-1} \right\} \quad (1.28)$$

tem distribuição nula igual à χ_q^2 , exceto por termos de ordem n^{-2} , onde $\mu_i' = 2^i \Gamma\left(\frac{q}{2} + i\right) / \Gamma\left(\frac{q}{2}\right)$ é o i -ésimo momento ordinário da distribuição χ_q^2 e $\Gamma(\cdot)$ é a função gama. A fórmula (1.28) é bastante geral e pode ser usada para melhorar muitos testes importantes em estatística e econometria. O termo entre chaves em (1.28) é denominado de *fator de correção de Bartlett generalizado*, porque envolve a própria estatística não-modificada e, portanto, não é uma correção de Bartlett tradicional. A diferença conceitual básica entre a correção em (1.28) e

a clássica correção de Bartlett é que a primeira envolve a própria estatística não-modificada e parâmetros desconhecidos, enquanto que a segunda depende somente de parâmetros desconhecidos. É claro que, quando $k = 1$, a correção em (1.28) reduz-se à correção de Bartlett tradicional. Segundo Ferrari (1991), é importante notarmos que para as estatísticas cujas distribuições assintóticas, até ordem n^{-1} , são conhecidas, os coeficientes do polinômio em (1.28) são construídos diretamente dos coeficientes envolvidos na própria distribuição assintótica.

Conforme já ressaltado por Cordeiro e Ferrari (1991), o resultado em (1.28) constitui uma ferramenta poderosa para o aperfeiçoamento de amplas classes de estatísticas. Em especial, Cordeiro e Ferrari (1991) discutem o aperfeiçoamento do teste escore via (1.28), onde os a_i 's são dados por Harris (1986). Ferrari (1991) mostra, ainda, como a estatística da razão de verossimilhança aperfeiçoada $\omega^* = \omega(1 - b)$, equivalente a $\omega^* = \omega/(1 + b)$, pode ser obtida a partir de (1.28). Segundo Ferrari (1991), outras aplicações interessantes podem incluir a estatística de Wald e sua versão modificada proposta por Hayakawa e Puri (1985).

Pelo exposto aqui, é possível prevermos que, nos próximos anos, muitos trabalhos deverão surgir visando ao aperfeiçoamento de estatísticas com distribuição assintótica de primeira ordem χ^2 , através do cálculo de fatores de correções Bartlett generalizado em (1.28), para modelos especiais.

CAPÍTULO 2

APERFEIÇOAMENTO DO TESTE DA RAZÃO DE VEROSSIMILHANÇA PARA O MODELO NORMAL LINEAR COM ESTRUTURA DE COVARIÂNCIA PARA OS ERROS

2.1 INTRODUÇÃO

Recentemente, Cordeiro (1993b) obtém uma expressão matricial geral que possibilita o cálculo do fator de correção de Bartlett para a estatística da razão de verossimilhança associada a alguns testes de hipóteses em dois modelos de regressão normal heterocedásticos: o modelo multiplicativo e o modelo variável dependente.

O objetivo deste capítulo é estender o trabalho desenvolvido por Cordeiro (1993b) a uma classe de modelos mais ampla, ou seja, obter uma fórmula geral em notação matricial para o

fator de correção de Bartlett associado à estatística da razão de verossimilhança na classe dos modelos normais lineares, onde os erros têm estrutura de covariância não-escalar, que depende de um conjunto de parâmetros desconhecidos. No Capítulo 6 generalizamos, também, o artigo de Cordeiro (1993b) para modelos normais heterocedásticos com função de variância arbitrária. Entretanto, esse resultado é mais restritivo do que aquele desenvolvido neste capítulo. Os modelos que estudamos aqui têm dois conjuntos distintos de parâmetros, um associado à média das observações e, o outro, relativo à estrutura de variância-covariância dos erros. No modelo que estamos considerando, pelo fato de os dois conjuntos de parâmetros desconhecidos serem globalmente ortogonais (Cox e Reid, 1987), ocorrem algumas simplificações nos cálculos, pois muitos dos cumulantes mistos são nulos.

Na Seção 2.2 introduzimos a estrutura básica para os modelos normais lineares tratados aqui e a notação padrão adotada para os cumulantes conjuntos de derivadas da função de log-verossimilhança total. O valor esperado da estatística da razão de verossimilhança, até ordem n^{-1} , baseado no trabalho de Lawley (1956), é deduzido na Seção 2.3. As expressões resultantes apresentam vantagens sob o ponto de vista numérico e simbólico, pois requerem apenas operações simples com matrizes e vetores. Encerramos este capítulo apresentando, na Seção 2.4, alguns casos especiais, que servem como uma verificação das fórmulas obtidas.

2.2 MODELO NORMAL LINEAR

Vamos considerar o modelo normal linear onde a matriz de covariância é não-escalar e depende de um conjunto de parâmetros desconhecidos que podem ser eficientemente estimados por máxima verossimilhança. O modelo adotado é descrito, matricialmente, por

$$Y = X\beta + U, \quad (2.1)$$

onde Y é um vetor $n \times 1$ de variáveis aleatórias, X é uma matriz modelo de regressores conhecidos e fixados, β é um vetor $p \times 1$ de parâmetros de regressão desconhecidos e U é um vetor $n \times 1$

de erros não observados. Para efeito de estimação por máxima verossimilhança e testes de hipóteses supomos que o vetor aleatório de erros U segue uma distribuição normal multivariada com vetor média zero e matriz de covariância V^{-1} $n \times n$ não-singular. A classe de modelos (2.1) inclui muitos dos modelos importantes discutidos na literatura como, por exemplo, os modelos com erros ARMA e o modelo de regressão heterocedástico multiplicativo.

Por conveniência trabalhamos com a matriz de precisão V $n \times n$, sendo, portanto, a inversa da matriz de covariância. Os elementos de $V = V(\gamma)$ são funções contínuas de um vetor paramétrico γ q -dimensional, desconhecido. O espaço paramétrico para β é o espaço Euclidiano p -dimensional, enquanto que o espaço paramétrico para γ é um conjunto Γ num espaço Euclidiano q -dimensional. Assumimos ainda que p e q são pequenos comparados com n e que $X^T V X$ é positiva definida para todo γ em Γ .

O modelo (2.1) com as suposições acima permite-nos afirmar que o vetor de variáveis aleatórias $Y = (Y_1, \dots, Y_n)^T$ segue uma distribuição normal n -variada com vetor média $\mu = X\beta$ e matriz de covariância $V^{-1} = V^{-1}(\gamma)$. Seja $\theta = (\beta^T, \gamma^T)^T$ o vetor $(p + q)$ -dimensional de parâmetros desconhecidos e $\ell = \ell(\theta)$ o logaritmo da função de verossimilhança total para θ , dado o vetor de observações $y = (y_1, \dots, y_n)^T$, onde y_l denota o valor observado da variável aleatória Y_l . Para o modelo (2.1) temos

$$\ell = -\frac{n}{2} \log(2\pi) + \frac{1}{2} \log |V| - \frac{1}{2} (y - X\beta)^T V (y - X\beta). \quad (2.2)$$

Certamente, em amostras grandes, ℓ deve satisfazer algumas condições de regularidade, conforme já mencionamos no Capítulo 1. Uma discussão detalhada das condições de regularidade para o modelo que estamos considerando aqui é apresentada por Rothenberg (1984). Assumimos, ainda, que a função ℓ seja regular (Cox e Hinkley, 1974; Capítulo 9) com respeito às derivadas em relação a todos os componentes de β e γ , até e incluindo as de quarta ordem.

Um caso especial importante da função (2.2) é o logaritmo da função de verossimilhança para modelos ARMA gerais. Expressões para V ou V^{-1} para estes modelos podem ser encontradas,

por exemplo, em Shaman (1973), Galbraith e Galbraith (1974) e Ljung e Box (1979). Outro caso especial corresponde ao modelo heterocedástico multiplicativo estudado por Cordeiro (1993b), em que V^{-1} é uma matriz de covariância diagonal com elementos diagonais $\sigma_1^2, \dots, \sigma_n^2$, onde $\sigma_l^2 = \text{Var}(Y_l) = \exp(\omega_l^\top \gamma)$, $\omega_l^\top$ sendo um vetor $1 \times q$ de constantes conhecidas. Este modelo é bastante útil em diversas áreas, incluindo engenharia, economia e ciências físicas e biológicas. Podemos generalizar este modelo considerando uma variância $\sigma_l^2 = H(\omega_l^\top \gamma)$, onde $H(\cdot)$ é uma função contínua diferenciável até terceira ordem. Para este caso, aperfeiçoamentos da estatística de máxima verossimilhança para alguns testes de hipóteses são apresentados no Capítulo 6, usando a metodologia desenvolvida por Cordeiro (1993b). Outra situação especial é aquela na qual a variância depende das covariáveis na resposta esperada como, por exemplo, $\sigma_l^2 = \sigma^2 V(x_l^\top \beta)$ para $l = 1, \dots, n$, onde $V(\cdot)$ é conhecida, monótona e assumimos ser duas vezes diferenciável e $x_l^\top = (x_{l1}, \dots, x_{lp})$ é um vetor de valores fixados de p variáveis auxiliares para a l -ésima observação. Esta formulação leva a um modelo heterocedástico variável dependente, para o qual Cordeiro (1993b) desenvolveu o aperfeiçoamento de alguns testes da razão de verossimilhança.

Adotamos a notação tensorial para os cumulantes conjuntos de derivadas da função de log-verossimilhança total (vide, por exemplo, McCullagh, 1984, 1987):

$$\begin{aligned} \kappa_{ij} &= E \left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right], \quad \kappa_{i,j} = E \left[\left(\frac{\partial \ell}{\partial \theta_i} \right) \left(\frac{\partial \ell}{\partial \theta_j} \right) \right], \\ \kappa_{ijl} &= E \left[\frac{\partial^3 \ell}{\partial \theta_i \partial \theta_j \partial \theta_l} \right], \quad \kappa_{ij,l} = E \left[\left(\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right) \left(\frac{\partial \ell}{\partial \theta_l} \right) \right], \\ \kappa_{ijlm} &= E \left[\frac{\partial^4 \ell}{\partial \theta_i \partial \theta_j \partial \theta_l \partial \theta_m} \right], \quad \kappa_{ij,lm} = E \left[\left(\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right) \left(\frac{\partial^2 \ell}{\partial \theta_l \partial \theta_m} \right) \right] - \kappa_{ij} \kappa_{lm}, \\ \kappa_{i,j,lm} &= E \left[\left(\frac{\partial \ell}{\partial \theta_i} \right) \left(\frac{\partial \ell}{\partial \theta_j} \right) \left(\frac{\partial^2 \ell}{\partial \theta_l \partial \theta_m} \right) \right] - \kappa_{i,j} \kappa_{lm} \\ \text{e} \quad \kappa_{i,j,l,m} &= E \left[\left(\frac{\partial \ell}{\partial \theta_i} \right) \left(\frac{\partial \ell}{\partial \theta_j} \right) \left(\frac{\partial \ell}{\partial \theta_l} \right) \left(\frac{\partial \ell}{\partial \theta_m} \right) \right] - \kappa_{i,j} \kappa_{l,m} - \kappa_{i,l} \kappa_{j,m} - \kappa_{i,m} \kappa_{j,l}. \end{aligned}$$

Usamos, também, a seguinte notação para as derivadas de cumulantes:

$$\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}, \quad \kappa_{ij}^{(lm)} = \frac{\partial^2 \kappa_{ij}}{\partial \theta_l \partial \theta_m},$$

e assim por diante, onde os componentes de θ podem representar qualquer componente de β e γ , ou seja, $i, j, l, m = 1, \dots, p + q$. A matriz de informação total de Fisher K , de ordem $p + q$, tem elementos $\kappa_{i,j} = -\kappa_{j,i}$, sendo K^{-1} a sua inversa, com elementos $\kappa^{i,j} = -\kappa^{j,i}$. Todos os κ 's referem-se a um total sobre a amostra e são, em geral, de ordem n . Denotamos o somatório sobre todos os componentes do vetor de parâmetros $\theta = (\beta^\top, \gamma^\top)^\top$ por Σ' e sobre os dados amostrais por Σ .

Diferenciando (2.2) e calculando valores esperados podemos encontrar os cumulantes conjuntos de derivadas de $\ell(\theta)$ em relação aos componentes de β e γ . É fácil mostrar que $E(-\partial^2 \ell / \partial \beta \partial \gamma) = 0$, isto é, os vetores paramétricos β e γ são globalmente ortogonais (Cox e Reid, 1987). Assim, a partição $\theta^\top = (\beta^\top, \gamma^\top)$ induz uma correspondente matriz de informação bloco-diagonal $K = \text{diag}\{K_{\beta,\beta}, K_{\gamma,\gamma}\}$ com submatrizes $K_{\beta,\beta} = E(-\partial^2 \ell / \partial \beta^2)$, a informação para β , e $K_{\gamma,\gamma} = E(-\partial^2 \ell / \partial \gamma^2)$, a informação para γ .

Sejam $\hat{\beta}$ e $\hat{\gamma}$ as estimativas de máxima verossimilhança de β e γ , respectivamente, e seja $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$ a estimativa de máxima verossimilhança do vetor completo θ . Assumimos que a estimativa $\hat{\theta}$ converge para o verdadeiro parâmetro θ quando o tamanho da amostra n cresce. Segundo as condições gerais de regularidade, os estimadores de máxima verossimilhança de β e γ , $\hat{\beta}$ e $\hat{\gamma}$ digamos, têm assintoticamente distribuição normal. Esta propriedade aliada à estrutura bloco diagonal da matriz de informação K , permite-nos afirmar que as estimativas $\hat{\beta}$ e $\hat{\gamma}$ são assintoticamente independentes. As equações de máxima verossimilhança para $\hat{\beta}$ e $\hat{\gamma}$ podem ser escritas por

$$\hat{\beta} = (X^\top \hat{V} X)^{-1} X^\top \hat{V} y$$

e

$$\text{tr}(\hat{V} \hat{V}_R) = (y - X \hat{\beta})^\top \hat{V}_R (y - X \hat{\beta}), \quad (2.3)$$

para $R = 1, \dots, q$, onde $\hat{V} = V(\hat{\gamma})$ e $\hat{V}_R = V_R(\hat{\gamma})$, sendo $V_R = \partial V / \partial \gamma_R$. As estimativas $\hat{\beta}$ e $\hat{\gamma}$ podem ser calculadas numericamente por qualquer algoritmo iterativo, como por exemplo,

o método de Newton-Raphson ou o método escore de Fisher. Sabemos que estas estimativas podem ser viesadas quando o tamanho n da amostra ou a informação total de Fisher é pequena. Expressões para os vieses, de ordem n^{-1} , dos estimadores de máxima verossimilhança para os parâmetros destes modelos são apresentadas no Capítulo 3.

2.3 CORREÇÃO DE BARTLETT PARA OS MODELOS NORMAIS LINEARES

2.3.1 OBTENÇÃO DA QUANTIDADE ε_{p+q}

Os fatores de correção de Bartlett para a estatística da razão de verossimilhança que são apresentados na Seção 1.3.1, dependem da quantidade ε_r , dada por (1.6), que, aparentemente, é uma função complicada dos cumulantes conjuntos κ 's de derivadas do logaritmo da função de verossimilhança total. O objetivo desta seção é obter ε_r em forma matricial e de fácil computação para a classe dos modelos normais lineares com estrutura de covariância não-escalar desconhecida. Como visto na seção anterior, o logaritmo da função de verossimilhança total depende dos vetores de parâmetros $\beta = (\beta_1, \dots, \beta_p)^\top$ e $\gamma = (\gamma_1, \dots, \gamma_q)^\top$, globalmente ortogonais. Assim, o vetor de parâmetros total $\theta = (\beta^\top, \gamma^\top)^\top$ tem dimensão $r = p + q$, ou seja, encontramos ε_{p+q} . A expressão de ε_{p+q} obtida aqui constitui uma generalização das fórmulas (3.6) - (3.8) apresentadas por Cordeiro (1993b), para o caso particular de um modelo normal heterocedástico multiplicativo.

Adotamos a notação introduzida na seção anterior para cumulantes e derivadas de cumulantes em relação aos componentes de $\theta^\top = (\beta^\top, \gamma^\top)$. Para evitar ambiguidade de índices entre os componentes do vetor β e os componentes do vetor γ , reservamos índices com letras minúsculas para representar componentes de β e índices com letras maiúsculas para os componentes de γ . Assim, por exemplo, $\kappa_{rs} = E(\partial^2 \ell / \partial \beta_r \partial \gamma_s)$, $\kappa_{rs}^{(t)} = \partial E(\kappa_{rs}) / \partial \beta_t$, $\kappa_{rs}^{(TU)} = \partial^2 E(\kappa_{rs}) / (\partial \gamma_T \partial \gamma_U)$. As submatrizes $K_{\beta, \beta}$ e $K_{\gamma, \gamma}$ da matriz de informação total de Fisher K para $(\beta^\top, \gamma^\top)^\top$ têm

elementos dados por $\kappa_{r,s} = -\kappa_{rs}$ e $\kappa_{R,S} = -\kappa_{RS}$, respectivamente, cujas expressões são apresentadas no Apêndice 2. As correspondentes submatrizes inversas, $K_{\beta,\beta}^{-1}$ e $K_{\gamma,\gamma}^{-1}$, têm elementos denotados por $\kappa^{r,s}$ e $\kappa^{R,S}$, respectivamente.

Denotamos por x_i a i -ésima coluna da matriz modelo X . Assim, podemos escrever $X\beta = (\sum_{j=1}^p x_{1j}\beta_j, \dots, \sum_{j=1}^p x_{ni}\beta_j)^\top$, implicando que $\partial(X\beta)/\partial\beta_i = x_i = (x_{1i}, \dots, x_{ni})^\top$ e $\partial^r(X\beta)/\partial\beta_{i_1} \dots \partial\beta_{i_r} = 0$, $\forall r \geq 2$, onde β_{i_j} pertence ao conjunto $\{\beta_1, \dots, \beta_p\}$.

Definimos $V_R = \partial V / \partial \gamma_R$, $V^R = \partial V^{-1} / \partial \gamma_R$, $V_{RS} = \partial^2 V / \partial \gamma_R \partial \gamma_S$ e $V^{RS} = \partial^2 V^{-1} / \partial \gamma_R \partial \gamma_S$, para $\gamma_R, \gamma_S \in \{\gamma_1, \dots, \gamma_q\}$. Vamos utilizar, também, as relações:

$$\begin{aligned} \text{(i)} \quad & V^R = -V^{-1}V_RV^{-1}; \\ \text{(ii)} \quad & \frac{\partial \log |V|}{\partial \gamma_R} = \text{tr}(V^{-1}V_R); \\ \text{(iii)} \quad & E[(Y - E(Y))^\top G(Y - E(Y))] = \text{tr}(V^{-1}G), \end{aligned}$$

onde V^{-1} é a matriz de variância-covariância do vetor aleatório Y e G é qualquer matriz positiva definida (vide Press, 1982 e Mardia e Marshall, 1984).

Com a finalidade de simplificar a obtenção da quantidade ε_{p+q} , definimos, também, $\bar{V}_R = V^{-1}V_R$, $\bar{V}^R = V^RV = -\bar{V}_R$, $\bar{V}_{RS} = V^{-1}V_{RS}$, $\bar{V}^{RS} = V^{RS}V$, $\bar{V}_{RST} = V^{-1}V_{RST}$, $\bar{V}^{RST} = V^{RST}V$, etc., $m_R = \text{tr}(\bar{V}_R)$, $m^R = \text{tr}(\bar{V}^R) = -m_R$, $m_{RS} = \text{tr}(\bar{V}_{RS})$, $m^{RS} = \text{tr}(\bar{V}^{RS})$, $m_{R,S} = \text{tr}(\bar{V}_R\bar{V}_S)$, $m^{R,S} = \text{tr}(\bar{V}^R\bar{V}^S)$, $m_{RS,T} = \text{tr}(\bar{V}_{RS}\bar{V}_T)$, $m^{RS,T} = \text{tr}(\bar{V}^{RS}\bar{V}^T)$, $m_{R,S,T} = \text{tr}(\bar{V}_R\bar{V}_S\bar{V}_T)$, $m^{R,S,T} = \text{tr}(\bar{V}^R\bar{V}^S\bar{V}^T)$, $m_{R,S,T,U} = \text{tr}(\bar{V}_R\bar{V}_S\bar{V}_T\bar{V}_U)$, $m^{R,S,T,U} = \text{tr}(\bar{V}^R\bar{V}^S\bar{V}^T\bar{V}^U)$, e assim por diante.

Os $\bar{V}$'s definidos acima satisfazem algumas relações (vide Apêndice 2) como, por exemplo, $\bar{V}_{RS} = \bar{V}^R\bar{V}^S - \bar{V}^{RS} + \bar{V}^S\bar{V}^R$, $\bar{V}^{RS} = \bar{V}_R\bar{V}_S - \bar{V}_{RS} + \bar{V}_S\bar{V}_R$, etc. Das relações existentes entre os $\bar{V}$'s seguem algumas equações entre os m 's (vide Apêndice 2), que, em algumas situações especiais, facilitam seus cálculos, tais como, $m_{RS} = 2m^{R,S} - m^{RS}$, $m^{RS} = 2m_{R,S} - m_{RS}$, $m_{RS,T} = m^{RS,T} - m^{R,S,T} - m^{S,R,T}$, $m_{R,S,T} = -m^{R,S,T}$ e assim por diante.

Da expansão de Lawley (1956) podemos escrever $E\{2[\ell(\hat{\beta}, \hat{\gamma}) - \ell(\beta, \gamma)]\} = p + q + \varepsilon_{p+q} + O(n^{-2})$, onde ε_{p+q} é um termo de ordem n^{-1} , dado por

$$\varepsilon_{p+q} = \sum'_{\beta, \gamma} (\ell_{rstu} - \ell_{rstuvw}) \quad (2.4)$$

com os índices r, s, t, u, v, w (embora todos minúsculos) referindo-se aos componentes de β e γ , ou seja, $\Sigma'_{\beta, \gamma}$ denotando o somatório sobre todas as combinações dos $(p + q)$ parâmetros de β e γ , os ℓ 's sendo obtidos de

$$\ell_{rstu} = \kappa^{rs} \kappa^{tu} \left(\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right) \quad (2.5)$$

e

$$\ell_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left[\kappa_{rtv} \left(\frac{1}{6} \kappa_{suw} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{1}{4} \kappa_{svw} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right]. \quad (2.6)$$

Encontramos agora o termo ε_{p+q} de forma a tornar seu cálculo mais simplificado. O procedimento básico para conseguirmos isto está em usar a diagonalidade em bloco da matriz de informação K , que torna muitos dos cumulantes mistos nulos. Observe-se que $K = \text{diag}\{K_{\beta, \beta}, K_{\gamma, \gamma}\}$, onde $K_{\beta, \beta} = X^T V X$ e $K_{\gamma, \gamma} = \{\frac{1}{2} m_{R, S}\}$. A obtenção dos cumulantes conjuntos de derivadas de $\ell(\theta) = \ell(\beta, \gamma)$ em relação só a β , só a γ e a β e γ , e as derivadas desses cumulantes necessárias ao cálculo de ε_{p+q} , são apresentadas no Apêndice 2. De agora em diante usamos subscritos β e γ para reforçar dependência de quantidades em relação a estes vetores. Definimos as matrizes $G_{\gamma}^{(RS)} = \{g_{tu}^{(RS)}\} = X^T V_{RS} X$ de dimensão $p \times p$ e $D_{\gamma}^{(RS)} = \{d_{IJ}^{(RS)}\}$ de ordem $q \times q$, onde $d_{IJ}^{(RS)}$ é dado por $d_{IJ}^{(RS)} = 2m_{RI, SJ} - m_{RS, IJ} + 4m_{RS, I, J} - 4m_{R, S, I, J} - 2m_{R, I, S, J}$, para R e S variando em $\gamma_1, \dots, \gamma_q$. Sejam $\tau_{1, \beta}^{(R)}$ q vetores coluna de dimensão q , cujo S -ésimo elemento é $\text{tr}(K_{\beta, \beta}^{-1} G_{\gamma}^{(RS)})$, e $\tau_{2, \gamma}^{(R)}$ q vetores coluna de dimensão q , cujo S -ésimo elemento é $\text{tr}(K_{\gamma, \gamma}^{-1} D_{\gamma}^{(RS)})$. Definimos as matrizes quadradas de dimensão q , $M_{\gamma, \beta}$ e N_{γ} , formadas pelos q vetores coluna $\tau_{1, \beta}^{(R)}$ e pelos q vetores coluna $\tau_{2, \gamma}^{(R)}$, respectivamente. Também, podemos escrever para a matriz inversa da informação relativa a γ , $K_{\gamma, \gamma}^{-1} = (\kappa_1^{\gamma, \gamma}, \dots, \kappa_q^{\gamma, \gamma})$, com $\kappa_R^{\gamma, \gamma}$ denotando o vetor coluna $(\kappa^{\gamma_1, \gamma_R} \dots \kappa^{\gamma_q, \gamma_R})^T$, para $R = 1, \dots, q$. Finalmente, definimos $m_{\gamma, \beta} = \text{vec}(M_{\gamma, \beta})$,

$n_\gamma = \text{vec}(N_\gamma)$ e $\kappa^\gamma = \text{vec}(K_{\gamma,\gamma}^{-1})$. Depois de alguma manipulação algébrica (vide Apêndice 2), (2.5) reduz-se a

$$\sum_{\beta,\gamma}' \ell_{rstu} = \frac{1}{2} \kappa^{\gamma^\top} m_{\gamma,\beta} + \frac{1}{8} \kappa^{\gamma^\top} n_\gamma. \quad (2.7)$$

Para a dedução do somatório $\sum_{\beta,\gamma}' \ell_{rstuvw}$ em (2.6) definimos as seguintes matrizes de ordem $q \times q$: $P_{\gamma,\beta} = \{p_{VW}\}$ com elemento geral (V, W) , $p_{VW} = \text{tr}(G_\gamma^{(V)} K_{\beta,\beta}^{-1} G_\gamma^{(W)} K_{\beta,\beta}^{-1})$, sendo $G_\gamma^{(V)}$ q matrizes de ordem $p \times p$ dadas por $G_\gamma^{(V)} = X^\top V_\gamma X$, $T_\gamma = \{t_{VW}\}$ com elemento geral (V, W) , $t_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T^{(V)} K_{\gamma,\gamma}^{-1} T^{(W)})$, sendo $T^{(V)}$ q matrizes de ordem $q \times q$ cujo elemento geral (I, J) é $t_{IJ}^{(V)} = m_{V,IJ}$, $U_\gamma = \{u_{VW}\}$ com elemento geral (V, W) , $u_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T^{(V)}) \text{tr}[K_{\gamma,\gamma}^{-1} (T^{(W)} - 4T_1^{(W)})]$, sendo $T_1^{(W)}$ q matrizes de ordem $q \times q$, cujo elemento geral (I, J) é $t_{IJ}^{(W)} = m_{W,IJ}$, $Q_\gamma = \{q_{VW}\}$ com elemento geral (V, W) $q_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)})$, $R_\gamma = \{r_{VW}\}$ com elemento geral (V, W) $r_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)\top})$ e $S_\gamma = \{s_{VW}\}$ com elemento geral $s_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)}) \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(W)})$. Sejam os vetores coluna de dimensão q $\tau_{1,\gamma,\beta}$, cujo R -ésimo elemento é $\text{tr}(K_{\beta,\beta}^{-1} G_\gamma^{(R)})$, $\tau_{3,\gamma}$, cujo R -ésimo elemento é $\text{tr}(K_{\gamma,\gamma}^{-1} E_\gamma^{(R)})$, sendo $E_\gamma^{(R)} = \{e_{IJ}^{(R)}\}$ uma matriz de ordem $q \times q$, com elemento geral $e_{IJ}^{(R)} = 2m_{RI,J} + 3m_{R,IJ} - 6m_{R,I,J}$, e $\tau_{4,\gamma}$, cujo R -ésimo elemento é $\text{tr}(K_{\gamma,\gamma}^{-1} F_\gamma^{(R)})$, sendo $F_\gamma^{(R)} = \{f_{IJ}^{(R)}\}$ uma matriz de ordem $q \times q$, com elemento geral $f_{IJ}^{(R)} = -2m_{RI,J} - m_{R,IJ} + 2m_{R,I,J}$. Assim, após um trabalho algébrico complicado entre vetores e matrizes (vide Apêndice 2), o somatório $\sum_{\beta,\gamma}' \ell_{rstuvw}$ resulta em

$$\begin{aligned} \sum_{\beta,\gamma}' \ell_{rstuvw} &= \frac{1}{2} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) - \frac{1}{8} \tau_{1,\gamma,\beta}^\top K_{\gamma,\gamma}^{-1} (2\tau_{1,\gamma,\beta} - \tau_{3,\gamma} - \tau_{4,\gamma}) \\ &\quad + \frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right]. \end{aligned} \quad (2.8)$$

Substituindo as expressões (2.7) e (2.8) em (2.4) obtemos

$$\begin{aligned} \varepsilon_{p+q} &= \frac{1}{8} \kappa^{\gamma^\top} n_\gamma - \frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right] \\ &\quad + \frac{1}{2} \kappa^{\gamma^\top} m_{\gamma,\beta} - \frac{1}{2} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) + \frac{1}{8} \tau_{1,\gamma,\beta}^\top K_{\gamma,\gamma}^{-1} (2\tau_{1,\gamma,\beta} - \tau_{3,\gamma} - \tau_{4,\gamma}). \end{aligned} \quad (2.9)$$

É importante notar que a expressão geral para ε_{p+q} em (2.4) pode ser decomposta em três parcelas, uma parcela em que o somatório envolve somente componentes do vetor β , outra

parcela em que o somatório envolve somente componentes do vetor γ e a terceira parcela é mista, ou seja, o somatório envolve componentes de β e de γ . Em vista disso, podemos decompor (2.4) em

$$\varepsilon_{p+q} = A_\beta + B_\gamma + C_{\beta,\gamma} + O(n^{-2}) , \quad (2.10)$$

sendo A_β , B_γ e $C_{\beta,\gamma}$ as três parcelas referidas acima, respectivamente. Confrontando (2.10) com (2.9) verificamos que, para o modelo (2.1), temos

$$\begin{aligned} A_\beta &= 0 , \\ B_\gamma &= \frac{1}{8} \kappa^\gamma{}^\top n_\gamma - \frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right] , \\ C_{\beta,\gamma} &= \frac{1}{2} \kappa^\gamma{}^\top m_{\gamma,\beta} - \frac{1}{2} \text{tr} (K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) + \frac{1}{8} \tau_{1,\beta}^\top K_{\gamma,\gamma}^{-1} (2\tau_{1,\beta} - \tau_{3,\gamma} - \tau_{4,\gamma}) . \end{aligned} \quad (2.11)$$

2.3.2 COMENTÁRIOS SOBRE A QUANTIDADE ε_{p+q}

A fórmula matricial para ε_{p+q} em (2.9), embora tenha uma aparência complicada, pode ser calculada facilmente se utilizarmos pacotes de computação simbólica que permitam executar operações entre matrizes e vetores, tais como o REDUCE (Rayna, 1987), ou o MATHEMATICA (Wolfram, 1991) ou o MAPLE (Abell e Braselton, 1994). A vantagem principal em expressarmos ε_{p+q} em notação matricial está em conseguirmos obter fatores de correção de Bartlett em forma fechada para diferentes testes de hipóteses de interesse prático, sempre que as matrizes de informação tiverem forma fechada.

Embora ε_{p+q} seja calculado facilmente, ele é de difícil interpretação. A dificuldade fundamental é que os termos individuais em ε_{p+q} não são invariantes sob reparametrização e, portanto, sua interpretação depende do sistema de coordenadas escolhido, apesar da invariância de ε_{p+q} .

Devemos observar, ainda, que o termo B_γ em (2.11) não depende do vetor de parâmetros β nem da matriz modelo X , sendo função apenas da matriz de covariância V^{-1} e de suas primeira e segunda derivadas em relação ao vetor de parâmetros γ . Já, a quantidade $C_{\beta,\gamma}$ em

(2.11) depende apenas de V^{-1} e de suas primeira e segunda derivadas em relação a γ e da matriz modelo X . Desta forma, ε_{p+q} não depende explicitamente do vetor de parâmetros β , embora o seu valor numérico dependa da estimativa de máxima verossimilhança de β .

2.4 ALGUNS CASOS ESPECIAIS

Consideremos as partições $\beta = (\beta_1^\top, \beta_2^\top)^\top$ e $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$, onde $\beta_1 = (\beta_1, \dots, \beta_{p_1})^\top$, $\beta_2 = (\beta_{p_1+1}, \dots, \beta_p)^\top$, $\gamma_1 = (\gamma_1, \dots, \gamma_{q_1})^\top$ e $\gamma_2 = (\gamma_{q_1+1}, \dots, \gamma_q)^\top$, com $0 \leq p_1 \leq p$ e $0 \leq q_1 \leq q$. Nosso interesse é testar a hipótese nula composta $H : \beta_1 = \beta_1^{(0)}$, $\gamma_1 = \gamma_1^{(0)}$ contra a hipótese alternativa A : pelo menos uma das igualdades é violada, onde $\beta_1^{(0)}$ e $\gamma_1^{(0)}$ são vetores especificados de dimensões p_1 e q_1 , respectivamente, e β_2 e γ_2 são vetores de parâmetros de perturbação com dimensões $p - p_1$ e $q - q_1$, respectivamente. O caso $q_1 = 0$ e $0 < p_1 \leq p$ corresponde ao teste da hipótese nula $H_1 : \beta_1 = \beta_1^{(0)}$, sendo β_2 e γ vetores de parâmetros de perturbação. Para $p_1 = 0$ e $0 < q_1 \leq q$ a hipótese nula de interesse é $H_2 : \gamma_1 = \gamma_1^{(0)}$, implicando que β e γ_2 são vetores de parâmetros de incômodo. Além dos testes das hipóteses H_1 e H_2 , outros testes podem ser obtidos com facilidade para valores diferentes de p , q , p_1 e q_1 . Observemos que qualquer uma das hipóteses nulas que mencionamos aqui devem ser testadas contra a hipótese alternativa A : violação de pelo menos uma igualdade.

Sejam $\hat{\beta}$ e $\hat{\gamma}$ os estimadores de máxima verossimilhança de β e γ sob o modelo irrestrito, obtidos resolvendo-se as equações em (2.3) e, $\tilde{\beta}_2$ e $\tilde{\gamma}_2$ os estimadores de máxima verossimilhança de β_2 e γ_2 , respectivamente, sob a hipótese nula H . A estatística da razão de verossimilhança para o teste de H é dada por

$$\omega = 2 \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2) - \ell(\beta_1^{(0)}, \tilde{\beta}_2, \gamma_1^{(0)}, \tilde{\gamma}_2) \right]. \quad (2.12)$$

Sabemos que, para grandes amostras, a distribuição de ω sob a hipótese nula H é aproximada-

mente $\chi^2_{p_1+q_1}$. Reescrevendo ω como

$$\omega = 2 \left\{ \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2) - \ell(\beta_1^{(0)}, \beta_2, \gamma_1^{(0)}, \gamma_2) \right] - \left[\ell(\beta_1^{(0)}, \tilde{\beta}_2, \gamma_1^{(0)}, \tilde{\gamma}_2) - \ell(\beta_1^{(0)}, \beta_2, \gamma_1^{(0)}, \gamma_2) \right] \right\}, \quad (2.13)$$

onde $\ell(\beta_1^{(0)}, \beta_2, \gamma_1^{(0)}, \gamma_2)$ denota o valor do logaritmo da função de verossimilhança calculado nos valores dos parâmetros especificados na hipótese nula e nos valores verdadeiros para os parâmetros restantes, segue da Seção 2.3.1, que até ordem n^{-1} ,

$$2E \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2) - \ell(\beta_1^{(0)}, \beta_2, \gamma_1^{(0)}, \gamma_2) \right] = p + q + \varepsilon_{p+q} + O(n^{-2}), \quad (2.14)$$

onde ε_{p+q} , de ordem n^{-1} , é obtido de (2.9). Dos resultados de Lawley (1956) temos, ainda, que

$$2E \left[\ell(\beta_1^{(0)}, \tilde{\beta}_2, \gamma_1^{(0)}, \tilde{\gamma}_2) - \ell(\beta_1^{(0)}, \beta_2, \gamma_1^{(0)}, \gamma_2) \right] = (p-p_1) + (q-q_1) + \varepsilon_{(p-p_1)+(q-q_1)} + O(n^{-2}), \quad (2.15)$$

onde $\varepsilon_{(p-p_1)+(q-q_1)}$, de ordem n^{-1} , é obtido de (2.4) observando, agora, que o somatório $\Sigma'_{\beta, \gamma}$ daquela expressão é tomado somente sobre os componentes de β_2 e γ_2 , isto é, sobre $p+q-(p_1+q_1)$ parâmetros. Substituindo (2.14) e (2.15) em (2.13) segue que, sob a hipótese nula H , o valor esperado da estatística da razão de verossimilhança ω tem a forma

$$E(\omega) = p_1 + q_1 + \varepsilon_{p+q} - \varepsilon_{(p-p_1)+(q-q_1)} + O(n^{-2}). \quad (2.16)$$

O fator de correção de Bartlett c associado à estatística ω é determinado por $c = (p_1 + q_1)^{-1} E(\omega)$, sendo que $E(\omega)$ é dado por (2.16), considerando termos até ordem n^{-1} . Assim,

$$c = 1 + \frac{\varepsilon_{p+q} - \varepsilon_{(p-p_1)+(q-q_1)}}{p_1 + q_1}. \quad (2.17)$$

A estatística da razão de verossimilhança modificada $\omega^* = c^{-1}\omega$ tem, sob H , distribuição $\chi^2_{p_1+q_1}$ até ordem n^{-1} . Quando o fator de correção c depender de parâmetros desconhecidos, estes devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob H . Neste caso, a estatística resultante $\tilde{\omega}^* = \tilde{c}^{-1}\omega$ ainda tem, sob H , distribuição $\chi^2_{p_1+q_1}$ até ordem n^{-1} .

Para o teste da hipótese nula $H_1 : \beta_1 = \beta_1^{(0)}$ contra A , sendo β_2 e γ vetores de parâmetros de perturbação e $\beta_1^{(0)}$ um vetor especificado de dimensão p_1 , a estatística da razão de verossimilhança é dada por:

$$\omega_1 = 2[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma})] , \quad (2.18)$$

onde $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\gamma}$ são as estimativas de máxima verossimilhança irrestrita de β_1 , β_2 e γ , respectivamente, $\tilde{\beta}_2$ e $\tilde{\gamma}$ são as estimativas de máxima verossimilhança de β_2 e γ , respectivamente, supondo H_1 verdadeira, ou seja, $\beta_1 = \beta_1^{(0)}$. Para amostras grandes, a distribuição de ω_1 , sob H_1 e até ordem n^{-1} é, aproximadamente, $\chi_{p_1}^2$. Podemos reescrever ω_1 como

$$\omega_1 = 2 \left\{ \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] - \left[\ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] \right\} , \quad (2.19)$$

onde $\ell(\beta_1^{(0)}, \beta_2, \gamma)$ é a log-verossimilhança calculada no valor verdadeiro dos parâmetros e sob H_1 . Dos resultados da Seção 2.3.1 segue que, sob H_1 ,

$$2E \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] = p + q + \varepsilon_{p+q} + O(n^{-2}) , \quad (2.20)$$

onde ε_{p+q} , de ordem n^{-1} , é obtido de (2.9). Dos resultados de Lawley (1956) vem que

$$2E \left[\ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] = p + q - p_1 - \varepsilon_{p+q-p_1} + O(n^{-2}) , \quad (2.21)$$

sendo o termo ε_{p+q-p_1} de ordem n^{-1} , calculado usando (2.4), observando que o somatório $\sum'_{\beta, \gamma}$ naquela expressão é estendido somente sobre os componentes de β_2 e γ , ou seja, sobre $p + q - p_1$ parâmetros. Calculando o valor esperado em (2.19) e usando (2.20) e (2.21) temos que o valor esperado da estatística da razão de verossimilhança ω_1 , sob H_1 e até ordem n^{-1} , tem a forma

$$E(\omega_1) = p_1 + \varepsilon_{p+q} - \varepsilon_{p+q-p_1} . \quad (2.22)$$

O fator de correção de Bartlett c_1 , associado à estatística ω_1 , é determinado por $c_1 = p_1^{-1} E(\omega_1)$, sendo que $E(\omega_1)$ é dado por (2.22) até ordem n^{-1} . Assim,

$$c_1 = 1 + \frac{\varepsilon_{p+q} - \varepsilon_{p+q-p_1}}{p_1} , \quad (2.23)$$

e a estatística da razão de verossimilhança aperfeiçoada $\omega_1^* = c_1^{-1}\omega_1$ tem, sob H_1 , distribuição $\chi_{p_1}^2$, até ordem n^{-1} . Se o fator c_1 depender de parâmetros desconhecidos, devemos substituí-los por suas respectivas estimativas de máxima verossimilhança sob H_1 , resultando $\tilde{\omega}_1^* = \tilde{c}_1^{-1}\omega_1$, com a mesma distribuição assintótica $\chi_{p_1}^2$, até ordem n^{-1} .

Consideramos, agora, que o interesse é testar a hipótese nula $H_2 : \gamma_1 = \gamma_1^{(0)}$ contra A , sendo, neste caso, β e γ_2 os vetores de parâmetros de incômodo e $\gamma_1^{(0)}$ um vetor de valores especificados de dimensão q_1 . A estatística da razão de verossimilhança, para o teste de H_2 contra A , é dado por

$$\omega_2 = 2 \left[\ell(\hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_2) - \ell(\tilde{\beta}, \gamma_1^{(0)}, \tilde{\gamma}_2) \right] . \quad (2.24)$$

Para grandes tamanhos de amostra, a distribuição de ω_2 , sob H_2 , é aproximadamente $\chi_{q_1}^2$. O fator de correção de Bartlett c_2 , associado a ω_2 , é determinado por $c_2 = q_1^{-1}E(\omega_2)$, com $E(\omega_2)$ calculado até ordem n^{-1} sob H_2 . Analogamente ao teste de H e ao teste de H_1 , descritos anteriormente, podemos reescrever ω_2 como

$$\omega_2 = 2 \left\{ \left[\ell(\hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_2) - \ell(\beta, \gamma_1^{(0)}, \gamma_2) \right] - \left[\ell(\tilde{\beta}, \gamma_1^{(0)}, \tilde{\gamma}_2) - \ell(\beta, \gamma_1^{(0)}, \gamma_2) \right] \right\} . \quad (2.25)$$

De Lawley (1956), temos que

$$2E \left[\ell(\tilde{\beta}, \gamma_1^{(0)}, \tilde{\gamma}_2) - \ell(\beta, \gamma_1^{(0)}, \gamma_2) \right] = p + q - q_1 - \varepsilon_{p+q-q_1} + O(n^{-2}) , \quad (2.26)$$

sendo o termo ε_{p+q-q_1} de ordem n^{-1} , calculado através de (2.4), onde o somatório $\sum'_{\beta, \gamma}$ naquela expressão estende-se, neste caso, somente sobre os componentes de β e γ_2 , isto é, sobre $p+q-q_1$ parâmetros. Calculando o valor esperado de (2.25) resulta, sob H_2 e até ordem n^{-1} ,

$$E(\omega_2) = q_1 + \varepsilon_{p+q} - \varepsilon_{p+q-q_1} . \quad (2.27)$$

Assim,

$$c_2 = 1 + \frac{\varepsilon_{p+q} - \varepsilon_{p+q-q_1}}{q_1} .$$

Encontramos, a seguir, a expressão (2.17) para alguns casos especiais que são de grande interesse prático.

2.4.1 DOIS PARÂMETROS UNIDIMENSIONAIS ORTOGONAIS

No caso em que os parâmetros β e γ são unidimensionais, ou seja, $p = q = 1$, a expressão de ε_{p+q} torna-se bastante simplificada. As variáveis aleatórias $Y_1, \dots, Y_n$ são independentes e identicamente distribuídas como $N(\beta, \sigma^2)$, ou seja, o vetor $Y = (Y_1, \dots, Y_n)^\top$ segue uma distribuição normal n -variada com vetor média $X\beta = 1_n\beta$ e matriz de covariância $V^{-1} = \gamma I_n$, com $\gamma = \sigma^2$, sendo 1_n o vetor coluna $n \times 1$ de uns e I_n a matriz identidade de ordem n . Podemos escrever a função de log-verossimilhança, dadas as observações $y_1, \dots, y_n$, por

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

Denotamos cumulantes de derivadas da log-verossimilhança em relação a β e a γ pela notação do tipo $\kappa_{\beta\beta} = E(\partial^2 \ell / \partial \beta^2)$, $\kappa_{\beta\gamma} = E(\partial^3 \ell / \partial \beta \partial \gamma^2)$, etc. Derivando ℓ duas vezes em relação a β e a γ , obtemos, após calcular valores esperados,

$$K = \begin{pmatrix} n/\gamma & 0 \\ 0 & n/2\gamma^2 \end{pmatrix} \quad \text{e} \quad K^{-1} = \begin{pmatrix} \gamma/n & 0 \\ 0 & 2\gamma^2/n \end{pmatrix},$$

a matriz de informação total de Fisher e sua inversa, respectivamente. Temos $K_{\beta,\beta} = \kappa_{\beta,\beta} = n/\gamma$, $K_{\gamma,\gamma} = \kappa_{\gamma,\gamma} = n/2\gamma^2$, $\kappa_{\beta,\beta}^{-1} = \kappa^{\beta,\beta} = \gamma/n$ e $\kappa_{\gamma,\gamma}^{-1} = \kappa^{\gamma,\gamma} = 2\gamma^2/n$. A matriz de precisão $V = I_n/\gamma$ tem primeira e segunda derivadas dadas por $V_\gamma = -I_n/\gamma^2$ e $V_{\gamma\gamma} = 2I_n/\gamma^3$, respectivamente. Esta situação permite dois testes de hipóteses correspondentes a $p_1 = p = 1$ e $q_1 = 0$ e $p_1 = 0$ e $q_1 = q = 1$, que conduzem às hipóteses nulas $H_1 : \beta = \beta^{(0)}$ e $H_2 : \gamma = \gamma^{(0)}$, respectivamente, onde $\beta^{(0)}$ e $\gamma^{(0)}$ são valores fixados.

(i) Teste de hipótese para o parâmetro β

Se o interesse é testar a hipótese nula $H_1 : \beta = \beta^{(0)}$ contra $A_1 : \beta \neq \beta^{(0)}$, sendo $\beta^{(0)}$ um valor fixado, o valor esperado da correspondente estatística da razão de verossimilhança ω_1 , sob H_1 , usando (2.16), é escrito por

$$E(\omega_1) = 1 + \varepsilon_{1+1} - \varepsilon_1 + O(n^{-2})$$

onde ε_1 é calculado por (2.4), com a soma $\Sigma'_{\beta,\gamma}$ tomada somente sobre γ . Assim, usando a partição de ε_{p+q} dada por (2.10), segue que $\varepsilon_{1+1} = A_\beta + B_\gamma + C_{\beta,\gamma}$ e $\varepsilon_1 = B_\gamma$, resultando, depois de substituição em (2.16),

$$\begin{aligned} E(\omega_1) &= 1 + A_\beta + B_\gamma + C_{\beta,\gamma} - B_\gamma + O(n^{-2}) \\ &= 1 + A_\beta + C_{\beta,\gamma} + O(n^{-2}) = 1 + C_{\beta,\gamma} + O(n^{-2}) \end{aligned}$$

já que, para o modelo normal linear $A_\beta = 0$. Assim, o fator de correção de Bartlett c_1 associado a ω_1 simplifica-se, até ordem n^{-1} , em

$$c_1 = 1 + C_{\beta,\gamma}. \quad (2.28)$$

A quantidade $C_{\beta,\gamma}$ em (2.28) é obtida usando a expressão (2.11), sendo, para este caso, $\kappa^\gamma = 2\gamma^2/n$, $m_{\gamma,\beta} = 2/\gamma^2$, $P_{\gamma,\beta} = 1/\gamma^2$, $\tau_{1,\gamma,\beta} = -1/\gamma$, $\tau_{3,\gamma} = -8/\gamma$ e $\tau_{4,\gamma} = 8/\gamma$, que resulta, depois de substituição em (2.11), $C_{\beta,\gamma} = 3/2n$.

Finalmente, considerando termos até ordem n^{-1} , temos

$$E(\omega_1) = c_1 = 1 + \frac{3}{2n}, \quad (2.29)$$

e a estatística aperfeiçoada $\omega_1^* = (1 + \frac{3}{2n})^{-1}\omega_1$ tem, sob H_1 e até ordem n^{-1} , distribuição χ_1^2 .

O resultado acima coincide com o valor de $E(\omega_1)$ obtido diretamente, sem usar a fórmula de Lawley (vide Apêndice 1), constituindo assim uma verificação parcial da expressão para ε_{p+q} em (2.9).

(ii) Teste de hipótese para o parâmetro γ

Para testar a hipótese $H_2 : \gamma = \gamma^{(0)}$ contra $A_2 : \gamma \neq \gamma^{(0)}$, sendo $\gamma^{(0)}$ um valor fixado, procedemos de maneira análoga ao teste de hipótese do parâmetro β . O valor esperado da correspondente estatística da razão de verossimilhança ω_2 , sob a hipótese nula H_2 , é dado por

$$E(\omega_2) = 1 + \varepsilon_{1+1} - \varepsilon_1 + O(n^{-2}) = 1 + B_\gamma + C_{\beta,\gamma} + O(n^{-2}),$$

já que, agora, $\varepsilon_1 = A_\beta = 0$, implicando que a correção de Bartlett c_2 , associada a ω_2 , seja escrita, considerando termos até ordem n^{-1} , como

$$c_2 = 1 + B_\gamma + C_{\beta,\gamma}.$$

As quantidades B_γ e $C_{\beta,\gamma}$ são obtidas de (2.11), sendo $C_{\beta,\gamma} = 3/2n$, calculado no item (i) anterior, para o teste de hipótese para β . Para o cálculo de B_γ temos que, $n_\gamma = 12/\gamma^2$, $T_\gamma = 16/\gamma^2$, $U_\gamma = -16/\gamma^2$, $Q_\gamma = R_\gamma = S_\gamma = 4/\gamma^2$, resultando $B_\gamma = 1/3n$, que coincide com o valor de B_γ obtido pelo cálculo direto de $B_\gamma = \ell_{\gamma\gamma\gamma\gamma} - \ell_{\gamma\gamma\gamma\gamma\gamma\gamma}$, com $\ell_{\gamma\gamma\gamma\gamma}$ e $\ell_{\gamma\gamma\gamma\gamma\gamma\gamma}$ dados por (2.5) e (2.6), respectivamente, fazendo $r = s = t = u = v = w = \gamma$.

Assim,

$$c_2 = E(\omega_2) = 1 + \frac{1}{3n} + \frac{3}{2n} + O(n^{-2}) = 1 + \frac{11}{6n} + O(n^{-2}), \quad (2.30)$$

como já é conhecido, produzindo a estatística aperfeiçoada $\omega_2^* = (1 + (11/6n))^{-1}\omega_2$ com distribuição χ_1^2 até ordem n^{-1} sob a hipótese nula H_2 .

2.4.2 PARÂMETRO γ UNIDIMENSIONAL

Nesta seção consideramos o caso em que o vetor de variáveis aleatórias $Y = (Y_1, \dots, Y_n)^\top$ pode ser modelado por uma normal multivariada com $E(Y) = X\beta$ e $\text{Var}(Y) = \sigma^2 I_n$, ou seja, as variáveis aleatórias $Y_1, \dots, Y_n$ são independentes com a mesma variância $\sigma^2 = \gamma$, isto é, temos $q = 1$ e $p > 1$. Nesta situação, $K_{\beta,\beta} = X^\top X/\gamma$, $K_{\beta,\beta}^{-1} = \gamma(X^\top X)^{-1}$ e $K_{\gamma,\gamma}$, $K_{\gamma,\gamma}^{-1}$, V_γ e $V_{\gamma\gamma}$ são os mesmos apresentados na Seção 2.4.1. Para o cálculo da quantidade ε_{p+1} em (2.9), ou, equivalentemente, em (2.10), a parcela B_γ , expressa em (2.11) não depende do vetor de parâmetros β e nem da matriz modelo X , assumindo, portanto, o mesmo valor obtido na Seção 2.4.1, ou seja, $B_\gamma = 1/3n$. Para o cálculo da parcela $C_{\beta,\gamma}$ de (2.11) temos, depois de alguma manipulação algébrica, que $m_{\gamma,\beta} = 2p/\gamma^2$, $P_{\gamma,\beta} = p/\gamma^2$, $\tau_{\gamma,\beta} = -p/\gamma$, $\tau_{3\gamma} = -8/\gamma$ e $\tau_{4\gamma} = 8/\gamma$, resultando, depois de substituição, $C_{\beta,\gamma} = p/n + p^2/2n$.

Assim, segue

$$\varepsilon_{p+1} = \frac{1}{3n} + \frac{p}{n} + \frac{p^2}{2n} = \frac{3p(p+2) + 2}{6n} . \quad (2.31)$$

Se temos interesse em testar a hipótese nula $H_3 : \beta = \beta^{(0)}$ contra a hipótese alternativa $A_3 : \beta \neq \beta^{(0)}$, com $\beta^{(0)}$ um vetor de valores especificados e γ o parâmetro de perturbação, utilizando o desenvolvimento para testar H contra A na Seção 2.4, temos que a esperança da estatística da razão de verossimilhança ω_3 , para o teste de H_3 contra A_3 , é

$$E(\omega_3) = p + \varepsilon_{p+1} - \varepsilon_1 + O(n^{-2}) , \quad (2.32)$$

onde ε_{p+1} é dado por (2.31) e $\varepsilon_1 = B_\gamma = 1/3n$, uma vez que ε_1 é obtido de (2.9) considerando as parcelas que envolvem somente o parâmetro unidimensional γ . Assim,

$$E(\omega_3) = p + \frac{p}{n} + \frac{p^2}{2n} + O(n^{-2}) = p + \frac{p(p+2)}{2n} + O(n^{-2}) . \quad (2.33)$$

O fator de correção de Bartlett c_3 , associado a ω_3 , é determinado por $c_3 = p^{-1} E(\omega_3) = 1 + (p+2)/2n$. A estatística aperfeiçoada $\omega_3^* = \omega_3(1 + (p+2)/2n)^{-1}$ tem aproximadamente distribuição χ_p^2 , sob H_3 até ordem n^{-1} .

O valor c_3 obtido acima é o mesmo encontrado por Cordeiro (1993b) no final da Seção 4, quando lá se considera $p_1 = 0$, constituindo, portanto, outra verificação parcial da expressão (2.9).

Outro teste de grande interesse prático consiste em verificar a homogeneidade das médias. Neste caso estamos interessados em testar a hipótese nula $H_4 : \beta_2 = \dots = \beta_p = 0$, com $x_{l1} = 1$ para $l = 1, \dots, n$, contra A_4 : violação de pelo menos uma igualdade, sendo β_1 e γ os dois parâmetros de incômodo. Seguindo o mesmo raciocínio desenvolvido nos casos anteriores, temos que, sob H_4 e até ordem n^{-1} ,

$$E(\omega_4) = p - 1 + \varepsilon_{p+1} - \varepsilon_{1+1} , \quad (2.34)$$

sendo ω_4 a estatística da razão de verossimilhança para testar H_4 contra A_4 . O fator de correção de Bartlett c_4 , associado a ω_4 , é obtido pela relação $c_4 = (p-1)^{-1} E(\omega_4)$ e a estatística

aperfeiçoada $\omega_4^* = c_4^{-1}\omega_4$ tem, sob a hipótese de homogeneidade das médias e até ordem n^{-1} , distribuição χ_{p-1}^2 , ou seja, $c_4 = 1 + (\varepsilon_{p+1} - \varepsilon_{1+1})/(p-1)$.

As quantidades ε_{p+1} e ε_{1+1} em (2.34) já foram obtidas para os casos anteriores considerados nesta seção. Assim, de (2.31) e de (2.30), resulta, até ordem n^{-1} ,

$$E(\omega_4) = p - 1 + \frac{p^2 + 2p - 3}{2n} \quad (2.35)$$

e

$$c_4 = 1 + \frac{p^2 + 2p - 3}{2n(p-1)}.$$

O resultado em (2.35) coincide com a expressão (4.3) obtida por Cordeiro (1993b) para o mesmo teste de hipótese que estamos considerando, o que demonstra mais uma vez a exatidão da fórmula (2.9).

Finalizando este capítulo, ressaltamos que nosso objetivo aqui foi encontrar uma expressão matricial para a quantidade ε_{p+q} , nos modelos normais lineares com estrutura de covariância dependendo de um conjunto de parâmetros desconhecidos e, a partir dela, obter expressões bastante simples para fatores de correção de Bartlett, associados a vários testes de hipóteses para os parâmetros em alguns casos especiais dos modelos normais lineares. Uma verificação total da expressão de ε_{p+q} , dada em (2.9), com a expressão (3.6) - (3.8) obtida por Cordeiro (1993b), para o caso particular de modelo normal heterocedástico multiplicativo, deverá ser feita em trabalho futuro. Pretendemos, também, em pesquisa futura, efetuar um estudo de simulação para verificar o desempenho das estatísticas de máxima verossimilhança corrigidas pelos fatores de correção de Bartlett nos casos especiais considerados, comparado ao da estatística não corrigida. A grande importância das fórmulas (2.10) - (2.11), na prática, é que se usadas em conjunção com “softwares” de manipulação algébrica (REDUCE, MATHEMATICA, MAPLE, etc.) podemos obter formas fechadas para as correções de Bartlett em inúmeros casos especiais do modelo (2.1), quando a matriz de informação tem forma fechada.

CAPÍTULO 3

VIÉS EM MODELOS DE REGRESSÃO NORMAL LINEAR COM MATRIZ DE COVARIÂNCIA DESCONHECIDA

3.1 INTRODUÇÃO

No Capítulo 2 desenvolvemos o aperfeiçoamento do teste da razão de verossimilhança nos modelos normais lineares, cujos erros têm estrutura de covariância não-escalar e dependente de um conjunto de parâmetros desconhecidos. Fornecemos uma fórmula geral em notação matricial para a correção de Bartlett em testes de hipóteses nestes modelos. Um aspecto não menos relevante a ser explorado nestes modelos é o viés dos estimadores de máxima verossimilhança dos parâmetros em amostras pequenas. Nos últimos 20 anos tem aparecido na literatura uma grande quantidade de resultados referentes às propriedades em amostras finitas dos estimadores

de máxima verossimilhança não-lineares. Uma dificuldade óbvia com estimadores não-lineares é que eles não podem ser expressos como funções explícitas dos dados.

Sabemos que os estimadores de máxima verossimilhança são viesados quando o tamanho n da amostra ou a informação total de Fisher é pequena. Usualmente, o viés é ignorado na prática, pela alegação de que ele é desprezível quando comparado aos erros padrão das estimativas. De fato, o viés em geral é de ordem n^{-1} , enquanto que o desvio padrão assintótico é de ordem $n^{-1/2}$. Entretanto, para alguns modelos, termos que compõem o viés podem ser consideravelmente maiores que os correspondentes termos do desvio padrão e, portanto, é útil obtermos fórmulas para calcular vieses de segunda ordem de estimadores frequentemente usados em pesquisa aplicada. Para problemas com função de verossimilhança envolvendo um único parâmetro a ser estimado, o viés de ordem n^{-1} da estimativa de máxima verossimilhança em uma única amostra aleatória é apresentado primeiro por Bartlett (1953). Haldane e Smith (1956) deduzem expressões similares, de mesma ordem, para os quatro primeiros cumulantes da estimativa de máxima verossimilhança. Vários trabalhos tratando de viés e precisão de estimativas são conduzidos por Shenton e Bowman (1963, 1969, 1977). Shenton e Bowman (1963) estudam o caso uniparamétrico, apresentando os quatro primeiros momentos centrais da estimativa de máxima verossimilhança até ordens n^{-2} , n^{-3} , n^{-3} e n^{-4} , respectivamente. Extensão para o viés multiparamétrico de ordem n^{-1} foi apresentado por Shenton e Wallington (1962) e Cox e Snell (1968). Recentemente, Cordeiro e McCullagh (1991) e Cordeiro e Klein (1994) deduzem fórmulas para o viés de segunda ordem das estimativas de máxima verossimilhança dos parâmetros em modelos lineares generalizados (McCullagh e Nelder, 1989) e em modelos ARMA, respectivamente.

O principal objetivo deste capítulo é obter uma fórmula geral para o viés assintótico dos estimadores de máxima verossimilhança de parâmetros dos modelos normais lineares apresentados no Capítulo 2. A utilidade prática desta fórmula consiste em conseguir expressões em forma fechada para o viés de segunda ordem em termos dos parâmetros em inúmeros casos

particulares. Esta fórmula combinada com o uso de um sistema computacional algébrico, como o MATHEMATICA (Wolfram, 1991) ou o MAPLE (Char *et al.*, 1988), ou uma linguagem numérica que permita álgebra matricial como, por exemplo, APL ou GAUSS, possibilita um modo mais fácil para calcular o viés das estimativas em modelos normais lineares com matriz de covariância não-escalar e dependente de um conjunto de parâmetros desconhecidos.

A metodologia básica usada para encontrar este viés segue o trabalho desenvolvido por Cox e Snell (1968). A discussão neste capítulo procede da seguinte forma. Na Seção 3.2, apresentamos o desenvolvimento de uma fórmula matricial simples para o viés de ordem n^{-1} das estimativas de máxima verosimilhança dos parâmetros da estrutura de covariância. A fórmula depende somente da matriz modelo para o vetor de médias, da estrutura de covariância e das primeira e segunda derivadas parciais em relação aos parâmetros desta matriz de covariância. Esta fórmula é simples o suficiente para obtermos várias correções para o viés em forma fechada em uma variedade de modelos importantes. Na Seção 3.3 apresentamos aplicações da fórmula geral, obtida na Seção 3.2, em alguns modelos especiais. Finalmente, na Seção 3.4, apresentamos alguns estudos de simulação comprovando a importância da fórmula de correção pelo viés na prática.

3.2 VIÉS DA ESTIMATIVA DO PARÂMETRO γ

Consideremos o modelo descrito em (2.1) na Seção 2.2, que, para facilidade, é descrito novamente abaixo,

$$Y = X\beta + U \quad (3.1)$$

onde Y , X , β e U estão definidos logo após (2.1). O vetor aleatório $Y = (Y_1, \dots, Y_n)^\top$ segue uma distribuição normal n -variada, com vetor média $X\beta$ ($n \times 1$) e matriz de covariância $n \times n$ $V^{-1} = V^{-1}(\gamma)$ não-singular, sendo γ um vetor paramétrico q -dimensional desconhecido.

Seja $y = (y_1, \dots, y_n)^\top$ o vetor das observações de $Y = (Y_1, \dots, Y_n)^\top$. As equações de máxima verossimilhança para $\hat{\beta}$ e $\hat{\gamma}$, resultantes da obtenção de $\ell(\theta)$ em (2.2), em relação a β e a γ , são

$$\hat{\beta} = (X^\top \hat{V} X)^{-1} X^\top \hat{V} y \quad (3.2)$$

e

$$tr(\hat{V} \hat{V}_R) = (y - X \hat{\beta})^\top \hat{V}_R (y - X \hat{\beta}), \quad (3.3)$$

para $R = 1, \dots, q$, onde $\hat{V} = V(\hat{\gamma})$ e $\hat{V}_R = V_R(\hat{\gamma})$, sendo $V_R(\gamma) = \partial V / \partial \gamma_R$. Os estimadores $\hat{\beta}$ e $\hat{\gamma}$ podem ser calculados numericamente por qualquer algoritmo iterativo aplicado às equações (3.2) – (3.3).

Nesta seção nossa atenção está dirigida à correção pelo viés do estimador do vetor de parâmetros γ do modelo (3.1). Nosso objetivo é obter uma expressão matricial geral para o viés de ordem n^{-1} do estimador de máxima verossimilhança de γ . As quantidades envolvidas nesta expressão não definidas aqui estão introduzidas no Capítulo 2.

Para o modelo (3.1) denotamos por $B_1(\hat{\theta}_r)$ o viés de ordem n^{-1} de $\hat{\theta}_r$, para $r = 1, \dots, p + q$, onde $\hat{\theta}_r$ representa qualquer componente do vetor $\hat{\beta}$ e $\hat{\gamma}$, ou seja, qualquer componente do vetor $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$. Da expressão geral obtida por Cox e Snell (1968), para os vieses multiparamétricos de ordem n^{-1} dos estimadores de máxima verossimilhança, podemos escrever, usando a notação da Seção 2.2,

$$B_1(\hat{\theta}_r) = \sum'_{\beta, \gamma} \kappa^{ri} \kappa^{jl} (\kappa_{ij}^{(l)} - \frac{1}{2} \kappa_{ijl}), \quad (3.4)$$

para $r = 1, \dots, p + q$, onde o somatório $\sum'$ é tomado sobre todos $p + q$ parâmetros $\beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_q$. No que segue, a interpretação de superescritos é clara: indicam elementos da correspondente matriz inversa, e nunca se referem a uma potência em qualquer expressão de somatório. Na expressão anterior, $-\kappa^{ij}$ representa um elemento típico da inversa da matriz de informação de Fisher $K = \{-\kappa_{ij}\}$.

No lado direito da equação (3.4), estimadores consistentes dos parâmetros β e γ podem ser inseridos para definir os estimadores de máxima verossimilhança corrigidos $\tilde{\theta}_r = \hat{\theta}_r - \hat{B}_1(\hat{\theta}_r)$, para $r = 1, \dots, p + q$, onde $\hat{B}_1(\cdot)$ denota o valor de $B_1(\cdot)$ calculado no ponto $\hat{\theta}$. Os estimadores corrigidos $\tilde{\theta}_r$'s têm vieses menores que os correspondentes estimadores não corrigidos $\hat{\theta}_r$'s. O estimador corrigido $\tilde{\theta}$ não tem sido usado exaustivamente em modelos multiparamétricos, possivelmente, por causa da dificuldade em se avaliarem os valores esperados envolvidos na expressão (3.4). O cálculo de $B_1(\hat{\theta})$ para os estimadores de máxima verossimilhança nos modelos de regressão logística tem sido trabalhado por diversos autores, incluindo Anderson e Richardson (1979), McLachlan (1980), Schaffer (1983) e Copas (1988). Vieses de segunda ordem dos estimadores de máxima verossimilhança em modelos lineares generalizados e em modelos ARMA foram obtidos, recentemente, por Cordeiro e McCullagh (1991) e Cordeiro e Klein (1994), respectivamente. Um estudo importante feito por Cook *et al.* (1986) apresenta estes vieses para modelos de regressão não-linear.

Apresentamos, agora, a dedução do viés de ordem n^{-1} das estimativas $\hat{\beta}$ e $\hat{\gamma}$ no modelo (3.1). Como no Capítulo 2, reservamos índices minúsculos para representar os componentes do vetor β e índices maiúsculos para representar os componentes do vetor γ . Como os vetores paramétricos β e γ são globalmente ortogonais temos que $\kappa_{Rs} = \kappa^{rS} = 0$, implicando que para o cálculo de $B_1(\hat{\beta}_i)$ precisamos levar em conta somente as duas somas seguintes:

$$\sum_{\beta}' \kappa^{ij} \kappa^{lm} (\kappa_{jl}^{(m)} - \frac{1}{2} \kappa_{jlm})$$

e

$$\sum_{\beta, \gamma}' \kappa^{ij} \kappa^{LM} (\kappa_{jL}^{(M)} - \frac{1}{2} \kappa_{jLM}),$$

onde $\sum_{\beta}'$ denota o somatório apenas sobre os componentes de β e $\sum_{\beta, \gamma}'$ representa o somatório sobre todas as combinações dos $(p + q)$ parâmetros em β e γ . Como já visto no Capítulo 2, do modelo e da ortogonalidade entre β e γ resulta que $\kappa_{rst} = \kappa_{rST} = \kappa_{rs}^{(t)} = \kappa_{rS}^{(T)} = 0$. Assim, as duas últimas somas implicam que o viés de ordem n^{-1} de $\hat{\beta}$ é zero, isto é, $B_1(\hat{\beta}) = 0$. Este

resultado é esperado para o modelo normal linear com erros independentes, mas não é óbvio que isto acontece para qualquer matriz de covariância V^{-1} já que a estimativa $\hat{\beta}$ é obtida, a menos deste caso, de uma equação não-linear e pela dependência de $\hat{\beta}$ com y e $\hat{V}$. Embora o viés de ordem n^{-1} de $\hat{\beta}$ seja igual a zero, seus vieses de ordem mais alta podem ser diferentes de zero. Este é um tópico a ser pesquisado no futuro.

Para avaliar o viés de ordem n^{-1} da estimativa $\hat{\gamma}$, podemos reescrever a equação (3.4) como

$$B_1(\hat{\gamma}_I) = \sum_{\beta, \gamma}' \kappa^{Is} \kappa^{tu} (\kappa_{st}^{(u)} - \frac{1}{2} \kappa_{stu}),$$

para qualquer componente I de γ , onde os índices s, t, u variam em todos os parâmetros de $\theta^\top = (\beta^\top, \gamma^\top)$. Em vista da diagonalidade em bloco da matriz de informação para θ , podemos obter, para o I -ésimo componente do vetor de parâmetros γ , o viés $B_1(\hat{\gamma}_I)$ avaliando apenas as duas somas seguintes:

$$S_1 = \sum_{\beta, \gamma}' \kappa^{IR} \kappa^{st} (\kappa_{Rs}^{(t)} - \frac{1}{2} \kappa_{Rst})$$

e

$$S_2 = \sum_{\gamma}' \kappa^{IR} \kappa^{ST} (\kappa_{RS}^{(T)} - \frac{1}{2} \kappa_{RST}),$$

onde $\sum_{\gamma}'$ denota somatório variando em $\gamma_1, \dots, \gamma_q$.

Podemos obter uma expressão simples em notação matricial para $B_1(\hat{\gamma})$, substituindo os cumulantes κ 's e derivadas de cumulantes, introduzidos na Seção 2.2, nas somas S_1 e S_2 acima. Destacando que as submatrizes de informação para β e γ , são dadas, respectivamente por $K_{\beta, \beta} = \{\kappa_{r,s}\} = \{x_r^\top V x_s\}$ e $K_{\gamma, \gamma} = \{\kappa_{R,S}\} = \{\frac{1}{2} m_{R,S}\}$ (vide Apêndice 2), as expressões resultantes para S_1 e S_2 seguem de cálculos elementares entre matrizes, e são dadas por

$$S_1 = \frac{1}{2} \rho_I^\top K_{\gamma, \gamma}^{-1} \tau_{1\gamma, \beta} \quad \text{e} \quad S_2 = \frac{1}{4} \rho_I^\top K_{\gamma, \gamma}^{-1} \tau_{2\gamma},$$

onde ρ_I é um vetor $q \times 1$ com 1 no I -ésimo componente e zero nos restantes, $\tau_{1\gamma, \beta}$ é um vetor $q \times 1$ cujo R -ésimo componente típico é $\text{tr}(K_{\beta, \beta}^{-1} G_\gamma^{(R)})$, sendo $G_\gamma^{(R)} = X^\top V_R X$ uma matriz $p \times p$, para

$R = 1, \dots, q$, e $\tau_{2\gamma}$ é um vetor $q \times 1$ cujo S -ésimo componente é $\text{tr}(K_{\gamma,\gamma}^{-1} A_{\gamma}^{(S)})$, sendo $A_{\gamma}^{(S)}$ uma matriz $q \times q$, cujo elemento típico (U, V) é definido por $2m_{S,U,V} - m_{S,UV}$ para $S, U, V = 1, \dots, q$.

Podemos, então, escrever em notação matricial

$$B_1(\hat{\gamma}) = K_{\gamma,\gamma}^{-1} \tau_{\gamma,\beta}, \quad (3.5)$$

onde $\tau_{\gamma,\beta} = \frac{1}{4}(2\tau_{1\gamma,\beta} + \tau_{2\gamma})$.

É importante que sejam salientados alguns pontos pertinentes à equação (3.5). Primeiro, $B_1(\hat{\gamma})$ é uma função da matriz modelo X , da matriz de precisão V (ou da matriz de covariância V^{-1}) e da primeira e segunda derivadas parciais de V ou V^{-1} . Ela depende explicitamente da matriz de informação para γ . Em modelos especiais, cuja matriz de informação para θ tem forma fechada, é possível obter uma expressão em forma fechada para $B_1(\hat{\gamma})$. Embora o cálculo de $B_1(\hat{\gamma})$ seja fácil, porque sua expressão envolve somente operações simples com vetores e matrizes, ele tem difícil interpretação. Segundo, a equação (3.5) em conjunção com um sistema de computação algébrica, tal como o REDUCE (Hearn, 1984) ou o MAPLE (Char *et al.*, 1988), permitem calcular $B_1(\hat{\gamma})$ de forma bastante simplificada. Numericamente, podemos obter $B_1(\hat{\gamma})$ via o pacote estatístico APL ou GAUSS. Terceiro, a equação (3.5) permite avaliar a influência de termos de segunda ordem na locação da estimativa $\hat{\gamma}$. Para estruturas de covariância especiais, este termo pode ser relativamente grande e podemos estar interessados em estimá-lo (diretamente ou por algum algoritmo iterativo) ou, pelo menos, levá-lo em conta quando fazemos inferências. O lado direito da equação (3.5) pode ser estimado no ponto $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$ para definir o vetor corrigido pelo viés $\tilde{\gamma} = \hat{\gamma} - \hat{B}_1(\hat{\gamma}) = \hat{\gamma} - \hat{K}_{\gamma,\gamma}^{-1} \hat{\tau}_{\gamma,\beta}$. Espera-se que a estimativa corrigida $\tilde{\gamma}$ tenha propriedades amostrais melhores do que a não corrigida $\hat{\gamma}$ (vide estudos de simulação na Seção 3.4). O termo do viés $\hat{B}_1(\hat{\gamma})$ tem o efeito de direcionar $\tilde{\gamma}$ para o parâmetro verdadeiro γ .

3.3 ILUSTRAÇÕES

É bastante útil considerarmos alguns exemplos que ilustrem as técnicas apresentadas na seção anterior e permitam um melhor esclarecimento da notação adotada. Alguns outros casos importantes também são obtidos com facilidade, graças à vantagem da fórmula (3.5), que requer somente operações elementares com matrizes e vetores.

Em particular, a equação (3.5) cobre o modelo mais simples quando temos n variáveis aleatórias independentes e identicamente distribuídas, $Y_1, \dots, Y_n$, como uma distribuição normal com média desconhecida μ e variância desconhecida σ^2 , $N(\mu, \sigma^2)$, digamos. Neste caso, $X = 1_n$, $\beta = \mu$, $\gamma = \sigma^2$, $V^{-1} = \gamma I_n$, $V = \gamma^{-1} I_n$, $V_\gamma = \partial V / \partial \gamma = -\gamma^{-2} I_n$ e $V_{\gamma\gamma} = \partial^2 V / \partial \gamma^2 = 2\gamma^{-3} I_n$, onde 1_n é um vetor $n \times 1$ de uns e I_n é uma matriz identidade de ordem n . As estimativas de máxima verossimilhança de μ e σ^2 para uma amostra aleatória $y_1, \dots, y_n$ são, respectivamente, $\hat{\mu} = \bar{y} = \sum y_i / n$ e $\hat{\sigma}^2 = \sum (y_i - \bar{y})^2 / n$. É bastante conhecido que $\hat{\mu} \sim N(\mu, \sigma^2/n)$ e $\hat{\sigma}^2 \sim \sigma^2 \chi_{n-1}^2 / n$ e, portanto, $E(\hat{\mu}) = \mu$ e $E(\hat{\sigma}^2) = \sigma^2 - \sigma^2/n$. A matriz de informação para $\theta = (\beta^\top, \gamma^\top)^\top$ tem submatrizes $K_{\beta,\beta} = n/\gamma$ e $K_{\gamma,\gamma} = n/2\gamma^2$. Facilmente, podemos obter, $G_\gamma^{(\gamma)} = -n/\gamma^2$, $A_\gamma^{(\gamma)} = 0$ e, assim, $\tau_{1\gamma,\beta} = -1/\gamma$ e $\tau_{2\gamma} = 0$, resultando $\tau_{\gamma,\beta} = -1/2\gamma$. Substituindo as quantidades apropriadas na expressão (3.5), encontramos $B_1(\hat{\gamma}) = -\gamma/n = -\sigma^2/n$, concordando com o resultado já conhecido. Este resultado representa uma comprovação parcial da fórmula (3.5).

Como um segundo exemplo, consideramos o modelo heterocedástico multiplicativo $Y \sim N(X\beta, V^{-1})$, introduzido na Seção 2.2, para o qual $V^{-1} = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$ e $\sigma_l^2 = \exp(w_l^\top \gamma)$ para $l = 1, \dots, n$. Cordeiro(1993b) mostrou que

$$B_1(\hat{\gamma}) = -(W^\top W)^{-1} W^\top \tau^*, \quad (3.6)$$

onde $\tau^* = (A_d + VB_d)1_n$, $W = (w_1, \dots, w_n)^\top$ é uma matriz $n \times q$, $A = WK_{\gamma,\gamma}^{-1}W^\top/2 = W(W^\top W)^{-1}W^\top$ e $B = XK_{\beta,\beta}^{-1}X^\top = X(X^\top VX)^{-1}X^\top$ são matrizes $n \times n$, A_d e B_d são matrizes diagonais com os correspondentes elementos diagonais de A e B , respectivamente. As matrizes

$2A$ e B têm interpretações simples, sendo as estruturas de covariância assintótica de $W\hat{\gamma}$ e $X\hat{\beta}$, respectivamente.

Mostramos agora que a equação (3.6) segue diretamente da equação (3.5). Para isto, é necessário provar apenas que $\tau_{\gamma,\beta} = -\frac{1}{2}W^\top \tau^*$, uma vez que, neste caso, temos $K_{\gamma,\gamma} = W^\top W/2$. Usamos a notação $D_R = \text{diag}\{w_{1R}, \dots, w_{nR}\}$, para $R = 1, \dots, q$. Para este modelo temos $\bar{V}_R = -D_R$, $\bar{V}_{RS} = \bar{V}_R \bar{V}_S = D_R D_S$ e $\bar{V}_R \bar{V}_S \bar{V}_T = \bar{V}_{RS} \bar{V}_T = -D_R D_S D_T$. Também, $A_\gamma^{(R)} = -W^\top D_R W$ e $G_\gamma^{(R)} = X^\top V_R X$ para $R = 1, \dots, q$. Assim, $\tau_{\gamma,\beta} = \frac{1}{4}(2\tau_{1\gamma,\beta} + \tau_{2\gamma})$ tem um componente típico S dado por $\frac{1}{4}\{2 \text{tr}(K_{\beta,\beta}^{-1} X^\top V_S X) - \text{tr}(K_{\gamma,\gamma}^{-1} W^\top D_S W)\}$, que pode ser escrito como $-\frac{1}{2}\text{tr}(D_S \{VB + A\})$, para $S = 1, \dots, q$. Por outro lado, $-\frac{1}{2}W^\top \tau^* = -\frac{1}{2}W^\top (VB_d + A_d)$, cujo S -ésimo elemento é idêntico ao correspondente componente de $\tau_{\gamma,\beta}$. Portanto, concluímos que (3.5) é uma generalização de (3.6).

Como um terceiro exemplo, consideramos um modelo auto-regressivo de ordem 1, $AR(1)$, $Y_t - \mu = \rho(Y_{t-1} - \mu) + U_t$, $|\rho| < 1$, onde $U_t \sim N(0, \sigma^2)$ para $t = 1, \dots, n$. Aqui X reduz-se a um vetor coluna $n \times 1$ de uns, $\beta = \mu$ e $\gamma = (\rho, \sigma^2)^\top$. A matriz de covariância V^{-1} tem a forma simples

$$V^{-1} = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}$$

cujas inversa é

$$V = \sigma^{-2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Podemos encontrar $K_{\beta,\beta} = 2 - 2\rho + (n-2)(1 - 2\rho + \rho^2)$ e

$$K_{\gamma,\gamma} = \frac{\rho}{\sigma^2} \begin{pmatrix} \frac{\rho}{2} m_{\rho,\rho} & -\frac{1}{2\sigma^2} m_{\rho} \\ -\frac{1}{2\sigma^2} m_{\rho} & \frac{n}{2\sigma^4} \end{pmatrix},$$

onde $m_{\rho} = -2\rho/(1 - \rho^2)$, $m_{\rho,\rho} = 2\{n(1 - \rho^2) + 3\rho^2 - 1\}/(1 - \rho^2)^2$ e $m_{\rho\rho} = 2(n - 2)/(1 - \rho^2)$.

Além disso,

$$A_{\gamma}^{(\rho)} = \frac{\rho}{\sigma^2} \begin{pmatrix} \frac{\rho}{2} m_{\rho,\rho,\rho} - m_{\rho\rho,\rho} & 2m_{\rho,\rho,\sigma^2} - m_{\rho\sigma^2,\rho} \\ 2m_{\rho,\rho,\sigma^2} - m_{\rho\sigma^2,\rho} & 2m_{\rho,\sigma^2,\sigma^2} - m_{\sigma^2\sigma^2,\rho} \end{pmatrix}$$

e

$$A_{\gamma}^{(\sigma^2)} = \frac{\rho}{\sigma^2} \begin{pmatrix} \frac{\rho}{2} m_{\rho,\rho,\sigma^2} - m_{\rho\rho,\sigma^2} & 2m_{\rho,\sigma^2,\sigma^2} - m_{\rho\sigma^2,\sigma^2} \\ 2m_{\rho,\sigma^2,\sigma^2} - m_{\rho\sigma^2,\sigma^2} & 2m_{\sigma^2,\sigma^2,\sigma^2} - m_{\sigma^2\sigma^2,\sigma^2} \end{pmatrix},$$

onde $m_{\rho,\rho\rho} = -4(n - 2)\rho/(1 - \rho^2)^2$, $m_{\rho,\rho,\rho} = -\rho\{6n(1 - \rho^2) + 14\rho^2 - 6\}/(1 - \rho^2)^3$, $m_{\rho,\rho,\sigma^2} = m_{\rho,\rho\sigma^2} = -\{2n(1 - \rho^2) + 6\rho^2 - 2\}/\sigma^2(1 - \rho^2)^2$, $m_{\rho,\sigma^2,\sigma^2} = m_{\sigma^2,\rho\sigma^2} = -2\rho/\sigma^4(1 - \rho^2)$, $m_{\rho\rho,\sigma^2} = -2(n - 2)/\sigma^2(1 - \rho^2)$, $m_{\rho,\sigma^2\sigma^2} = -4\rho/\sigma^4(1 - \rho^2)$, $m_{\sigma^2,\sigma^2\sigma^2} = 2m_{\sigma^2\sigma^2,\sigma^2} = -2n/\sigma^6$. Temos, também, $G_{\gamma}^{(\rho)} = 2\sigma^{-2}\{-1 + (n - 2)(\rho - 1)\}$ e $G_{\gamma}^{(\sigma^2)} = -\sigma^{-4}\{2(1 - \rho) + (n - 2)(1 - 2\rho + \rho^2)\}$.

Assim,

$$\tau_{1\gamma,\beta} = \{2 - 2\rho + (n - 2)(1 - 2\rho + \rho^2)\}^{-1} \begin{pmatrix} 2\{(n - 2)(\rho - 1) - 1\}/\sigma^2 \\ -\{2(1 - \rho) + (n - 2)(1 - 2\rho + \rho^2)\}/\sigma^4 \end{pmatrix}$$

e

$$\tau_{2\gamma} = \frac{4}{(nm_{\rho,\rho} - m_{\rho}^2)} \begin{pmatrix} \{n(2m_{\rho,\rho,\rho} - m_{\rho\rho,\rho}) - 2m_{\rho}m_{\rho,\rho}\}/2 \\ \{n(m_{\rho\rho} - 2m_{\rho,\rho}) - 2m_{\rho}^2\}/2\sigma^2 \end{pmatrix}.$$

Podemos obter uma expressão simples para $B_1(\hat{\gamma})$ introduzindo os m 's nas expressões para $K_{\gamma,\gamma}$ e $\tau_{2\gamma}$. Após efetuarmos um trabalho algébrico bastante intenso, envolvendo produtos de

matrizes complicadas, encontramos as seguintes expressões para os vieses de ordem n^{-1} de $\hat{\rho}$ e $\hat{\sigma}^2$ no modelos AR(1):

$$b_{\rho} = B_1(\hat{\rho}) = \frac{-2\rho(1 - \rho^2)\{n^3(1 - \rho^2) - 2n^2(1 - 2\rho^2) + n(1 - 5\rho^2) + 2\rho^2\}}{\{n^2(1 - \rho^2) - n(1 - 3\rho^2) - 2\rho^2\}^2}$$

e

$$b_{\sigma^2} = B_2(\hat{\sigma}^2) = \frac{-n(1 - \rho^2)\sigma^2}{n^2(1 - \rho^2) - n(1 - 3\rho^2) - 2\rho^2}.$$

Estas fórmulas coincidem com aquelas dadas em Cordeiro e Klein (1994). Podemos aproximar estes vieses considerando apenas termos de ordem n^{-1} e, assim, as duas equações anteriores reduzem-se a

$$b_{\rho} = E(\hat{\rho} - \rho) = -2\rho/n \quad \text{e} \quad b_{\sigma^2} = E(\hat{\sigma}^2 - \sigma^2) = -\sigma^2/n.$$

Estes resultados coincidem com as fórmulas de Tanaka (1984) para os vieses de ordem n^{-1} de $\hat{\rho}$ e $\hat{\sigma}^2$.

Finalizando, consideramos um modelo de médias móveis de ordem 1, MA(1), $Y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, $|\theta| < 1$, onde $\varepsilon_t \sim N(0, \sigma^2)$ para $t = 1, \dots, n$. Temos $\gamma = (\theta, \sigma^2)^T$, $X = 1_n$ e a matriz de covariância V^{-1} tem a forma simples

$$V^{-1} = \sigma^2 \begin{pmatrix} 1 + \theta^2 & -\theta & 0 & \dots & 0 & 0 \\ -\theta & 1 + \theta^2 & -\theta & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & -\theta & 1 + \theta^2 \end{pmatrix},$$

uma vez que $\text{Var}(Y_t) = \sigma^2(1 + \theta^2)$ e $\text{Cov}(Y_t, Y_{t-k}) = \rho_k(1 + \theta^2)$, com $\rho_k = -\theta/(1 + \theta^2)$ para $k = 1$ e $\rho_k = 0$ para $k \geq 2$. A inversa da matriz V^{-1} , acima, possui um elemento típico calculado por (vide Galbraith e Galbraith, 1974)

$$v_{r,s} = \frac{1}{\sigma^2(1 - \theta^2)(1 - \theta^{2(n+1)})} \theta^{|s-r|} (1 - \theta^{2r})(1 - \theta^{2(n+1-s)}).$$

Procedendo de maneira análoga àquela para o modelo AR(1), encontramos, após manipulação algébrica extremamente trabalhosa que omitimos aqui,

$$b_{\theta} = b_{\theta}(\theta, n) = \frac{\theta(1 - \theta^2)[n^3(1 - \theta^2) + n^2(1 - 5\theta^2) - 2n(1 + 4\theta^2) - 4\theta^2]}{[n^2(1 - \theta^2) - n(1 + 3\theta^2) - 2\theta^2]^2}$$

e

$$b_{\sigma^2} = b_{\sigma^2}(\theta, \sigma^2, n) = \frac{-n\sigma^2(1 - \theta^2)}{n^2(1 - \theta^2) - n(1 + 3\theta^2) - 2\theta^2}.$$

Assim, até ordem n^{-1} , as expressões acima reduzem-se a

$$b_{\theta} = E(\hat{\theta} - \theta) = \frac{\theta}{n} \quad \text{e} \quad b_{\sigma^2} = E(\hat{\sigma}^2 - \sigma^2) = -\frac{\sigma^2}{n},$$

que coincidem com os resultados de Tanaka (1984).

3.4 ALGUNS ESTUDOS DE SIMULAÇÃO

Apresentamos nesta seção um estudo de simulação cujo objetivo é verificar o desempenho da estimativa de máxima verossimilhança corrigida pelo viés ($\tilde{\theta}$) comparado ao da não corrigida ($\hat{\theta}$). O estudo é realizado para alguns modelos e considerando diferentes tamanhos de amostras.

3.4.1 MODELO HETEROCEDÁSTICO MULTIPLICATIVO

O modelo heterocedástico multiplicativo estudado é descrito pelas equações,

$$\begin{aligned} \mu_l &= \beta_0 + \beta_1 x_{l1} + \beta_2 x_{l2} + \beta_3 x_{l3}, \\ \sigma_l^2 &= \exp\{\gamma_0 + \gamma_1 \omega_{l1} + \gamma_2 \omega_{l2}\}, \end{aligned} \quad (3.7)$$

$l = 1, \dots, n$, sendo μ_l e σ_l^2 a média e a variância da l -ésima observação, respectivamente. Para as simulações, os parâmetros verdadeiros do modelo são fixados em $\beta_0 = 5$, $\beta_1 = \beta_2 = 2$, $\beta_3 = 3$, $\gamma_0 = 2$, $\gamma_1 = 1$, $\gamma_2 = 3$ e as variáveis x_{li} 's e ω_{li} 's são geradas das distribuições uniformes nos intervalos $(0, 50)$ e $(0, 5)$, respectivamente, sendo mantidas fixas para cada n . Os tamanhos de amostras n considerados são 10, 20, 30 e 40. Para cada n , 10.000 amostras são simuladas de uma distribuição normal seguindo (3.7). Em cada simulação, computamos: i) os valores $\hat{\beta}_i$'s e os $\hat{\gamma}_i$'s de

$$\hat{\beta} = (X^T \hat{V} X)^{-1} X^T \hat{V} y \quad \text{e} \quad W^T 1_n = W^T \hat{V} \hat{v},$$

onde $v = (v_1, \dots, v_n)^T$, com $v_l = (y_l - x_l^T \beta)^2$ e (ii) $\hat{B}_1(\hat{\gamma})$, onde $\hat{B}_1(\hat{\gamma}) = -(W^T W)^{-1} W^T \hat{\tau}^*$ com $\hat{\tau}^* = (A_d + \hat{V} B_d) 1_n$, e $\tilde{\gamma} = \hat{\gamma} - \hat{B}_1(\hat{\gamma})$. O programa de cálculo de $\hat{\beta}$ e $\hat{\gamma}$ e de simulação foi escrito em FORTRAN, com a utilização das sub-rotinas do programa NAG, implementadas no computador VAX. Para cada n calculamos as médias amostrais de $\hat{\gamma}$ e $\tilde{\gamma}$, baseadas nos 10.000 valores estimados. Os resultados obtidos estão apresentados na Tabela 3.1.

Tabela 3.1: Média dos valores simulados das estimativas de máxima verossimilhança $\hat{\gamma}$ e máxima verossimilhança corrigida pelo viés $\tilde{\gamma}$, no modelo heterocedástico multiplicativo.

n	$\gamma_0 = 2$		$\gamma_1 = 1$		$\gamma_2 = 3$	
	$\hat{\gamma}_0$	$\tilde{\gamma}_0$	$\hat{\gamma}_1$	$\tilde{\gamma}_1$	$\hat{\gamma}_2$	$\tilde{\gamma}_2$
10	2,39	2,15	1,35	1,23	3,61	3,28
20	2,21	2,09	1,17	1,10	3,38	3,13
30	2,15	2,04	1,11	1,06	3,20	3,06
40	2,09	1,98	1,07	1,02	3,12	2,98

Da observação da Tabela 3.1, podemos concluir que o desempenho das estimativas corrigidas pelo viés $\tilde{\gamma}$ é superior ao das estimativas usuais $\hat{\gamma}$, no sentido de estarem mais próximas dos valores verdadeiros. A correção pelo viés das estimativas de máxima verossimilhança é, portanto, efetiva em deslocar as estimativas não corrigidas na direção dos valores verdadeiros do modelo, até mesmo para tamanhos de amostra pequenos, ou moderadamente pequenos, como é o caso de $n = 10$ ou 20 . Para tamanhos de amostra maiores, $n = 30$ ou 40 , esta melhora não é substancial, como esperado.

3.4.2 MODELO AUTO-REGRESSIVO DE ORDEM 1

Consideramos, agora, o modelo AR(1) descrito por

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + U_t, \quad (3.8)$$

para $t = 1, \dots, n$, com $|\rho| < 1$ e U_t variáveis aleatórias i.i.d. $N(0, \sigma^2)$. Suponhamos que a média μ seja conhecida. Neste caso, $\gamma = (\rho, \sigma^2)^\top$. Para corrigir as estimativas de máxima verossimilhança usuais $\hat{\rho}$ e $\hat{\sigma}^2$, usamos as expressões exatas dos vieses até ordem n^{-1} , $b_\rho = b_\rho(\rho, n)$ e $b_{\sigma^2} = b_{\sigma^2}(\rho, \sigma^2, n)$, resultantes da aplicação da fórmula (3.5) e dadas na Seção 3.3. As estimativas de máxima verossimilhança corrigidas são calculadas por $\tilde{\rho} = \hat{\rho} - \hat{b}_\rho$ e $\tilde{\sigma}^2 = \hat{\sigma}^2 - \hat{b}_{\sigma^2}$, onde $\hat{b}_\rho = b_\rho(\hat{\rho}, n)$ e $\hat{b}_{\sigma^2} = b_{\sigma^2}(\hat{\rho}, \hat{\sigma}^2, n)$. A fim de verificar a eficiência da correção das estimativas, procedemos a três estudos de simulação, onde num deles comparamos as estimativas $\hat{\rho}$ e $\tilde{\rho}$ e, nos outros dois, as estimativas $\hat{\sigma}^2$ e $\tilde{\sigma}^2$. Utilizando o pacote computacional GLIM geramos 10.000 amostras, seguindo (3.8), para cada uma das combinações de valores de n ($n = 5, 10, 20, 30, 40$) e de ρ ($\rho = 0, 1; 0, 2; 0, 4; 0, 6; 0, 8$ e $0, 9$), fixando a média geral μ igual a 4,0 e a variância do erro σ^2 igual a 25, comuns às observações de todas as simulações. Calculamos, para cada uma das amostras simuladas, os valores das estimativas $\hat{\rho}$ e $\tilde{\rho}$, e computamos as médias amostrais baseadas nos 10.000 valores simulados. A estimativa $\hat{\rho}$ é calculada via uma macro do GLIM que ajusta o modelo AR(1) aos dados simulados. A Tabela 3.2 mostra os resultados obtidos.

Tabela 3.2: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\rho}$ e não corrigidas $\hat{\rho}$, do parâmetro ρ do modelo AR(1), com $\mu = 4,0$ e $\sigma^2 = 25$.

n	ρ		0,10		0,20		0,40		0,60		0,80		0,90	
	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$
5	0,06	0,08	0,14	0,22	0,27	0,35	0,38	0,51	0,47	0,69	0,71	0,80		
10	0,07	0,09	0,16	0,21	0,29	0,36	0,47	0,54	0,58	0,72	0,75	0,85		
20	0,08	0,09	0,17	0,19	0,31	0,36	0,50	0,55	0,63	0,70	0,78	0,85		
30	0,08	0,11	0,18	0,21	0,35	0,38	0,54	0,57	0,74	0,77	0,83	0,88		
40	0,09	0,10	0,18	0,19	0,36	0,38	0,55	0,58	0,75	0,78	0,84	0,87		

Através de análise desta tabela verificamos que, em todos os casos considerados, as estimativas de máxima verossimilhança corrigidas $\tilde{\rho}$ estão mais próximas dos valores verdadeiros de ρ do que as estimativas usuais $\hat{\rho}$. O tamanho da proximidade varia de acordo com o tamanho da

amostra e do valor verdadeiro de ρ . Por exemplo, para tamanhos de amostra grandes, 30 ou 40, as duas estimativas $\hat{\rho}$ e $\tilde{\rho}$ estão razoavelmente próximas de ρ , com um desempenho ligeiramente melhor para a estimativa corrigida. Entretanto, quando os tamanhos de amostra são pequenos, 5 ou 10, e para alguns valores de ρ , 0,4, 0,6 ou 0,8, a estimativa de máxima verossimilhança corrigida pelo viés $\tilde{\rho}$ tem um desempenho muito melhor do que a não corrigida $\hat{\rho}$, no sentido de estar mais próxima do valor verdadeiro ρ . Em praticamente todos os casos, as estimativas não corrigidas subestimam os parâmetros (exceção para $p = 0,10$ e $n = 30$, $p = 0,20$ e $n = 5$, 10 e 30).

Os outros dois estudos de simulação foram conduzidos com a finalidade de comparar os desempenhos das estimativas de máxima verossimilhança corrigida e não corrigida de σ^2 , $\tilde{\sigma}^2$ e $\hat{\sigma}^2$, respectivamente. Através do pacote computacional GLIM geramos 10.000 amostras para cada uma das combinações de valores de ρ ($\rho = 0,2; 0,4; 0,6$ e $0,8$) e de tamanho das amostras n ($n = 5, 10, 20, 30$ e 40), sendo mantidos fixos em todas as simulações, a média geral μ igual a 4,0 e o valor verdadeiro da variância do erro σ^2 , tomado como igual a 10, num estudo, e igual a 50, no outro. Para cada amostra, calculamos os valores das estimativas $\hat{\sigma}^2$ e $\tilde{\sigma}^2$, sendo as médias amostrais destas estimativas computadas com base nos 10.000 valores encontrados. Os resultados obtidos estão na Tabela 3.3 ($\sigma^2 = 10$) e na Tabela 3.4 ($\sigma^2 = 50$). Um exame destas tabelas permite-nos concluir que em todas as situações estudadas o desempenho da estimativa de máxima verossimilhança corrigida $\tilde{\sigma}^2$ é superior ao da não corrigida $\hat{\sigma}^2$, no sentido de estar mais próxima do valor verdadeiro do parâmetro σ^2 . Convém notar, ainda, que esta melhora na estimativa, quando usamos a correção pelo viés, é mais acentuada para valores de ρ mais próximos de zero e tamanho de amostra pequeno, por exemplo, para $n = 5$ e $\rho = 0,2, 0,4$ e $0,6$, quer σ^2 seja igual a 10 ou a 50. Evidentemente, a correção das estimativas tem menor efeito quando o tamanho da amostra é grande, $n = 30$ e 40 .

Tabela 3.3: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\sigma}^2$ e não corrigidas $\hat{\sigma}^2$, do parâmetro σ^2 do modelo AR(1), com $\mu = 4,0$ e $\sigma^2 = 10$.

n	ρ	0,2		0,4		0,6		0,8	
		$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$
5		6,34	8,53	6,75	8,71	6,73	8,69	7,42	8,83
10		7,62	8,60	8,83	9,52	8,80	9,41	8,85	9,64
20		8,28	8,75	8,87	9,49	8,92	9,46	9,25	9,67
30		8,52	8,89	9,14	9,41	9,28	9,55	9,53	9,81
40		8,90	9,15	9,18	9,39	9,31	9,52	9,56	9,78

Tabela 3.4: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\sigma}^2$ e não corrigidas $\hat{\sigma}^2$, do parâmetro σ^2 do modelo AR(1), com $\mu = 4,0$ e $\sigma^2 = 50$.

n	ρ	0,2		0,4		0,6		0,8	
		$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$
5		29,37	41,45	33,48	44,26	32,78	42,78	35,93	44,81
10		38,04	43,51	41,44	46,71	41,63	46,50	42,83	47,51
20		42,67	45,28	44,29	46,83	44,62	46,93	45,60	48,32
30		43,81	45,51	45,73	47,34	46,31	47,85	47,31	48,75
40		45,76	47,04	46,24	47,81	47,10	48,39	47,84	49,04

3.4.3 MODELO DE MÉDIAS MÓVEIS DE ORDEM 1

Estudamos, também, o modelo MA(1) descrito pela equação

$$Y_t - \mu = \varepsilon_t - \theta \varepsilon_{t-1}, \quad (3.9)$$

para $t = 1, \dots, n$, com $|\theta| < 1$ e ε_t sendo variáveis aleatórias i.i.d. $N(0, \sigma^2)$. Supomos que a média μ seja conhecida. Como no modelo AR(1), a correção das estimativas de máxima verossimilhança é feita através das fórmulas exatas dos vieses até ordem n^{-1} , obtidas de (3.5), e expressas por $b_\theta = b_\theta(\theta, n)$ e $b_{\sigma^2} = b_{\sigma^2}(\theta, \sigma^2, n)$ na Seção 3.3. Assim, as estimativas de máxima

verossimilhança corrigidas de θ e σ^2 são dadas por $\tilde{\theta} = \hat{\theta} - \hat{b}_\theta$ e $\tilde{\sigma}^2 = \hat{\sigma}^2 - \hat{b}_{\sigma^2}$, respectivamente, onde $\hat{b}_\theta = b_\theta(\hat{\theta}, n)$ e $\hat{b}_{\sigma^2} = b_{\sigma^2}(\hat{\theta}, \hat{\sigma}^2, n)$. Realizamos três estudos de simulação com o objetivo de comparar o desempenho das estimativas $\hat{\theta}$ e $\tilde{\theta}$ e das estimativas $\hat{\sigma}^2$ e $\tilde{\sigma}^2$. Como nos estudos anteriores, todas as amostras são simuladas usando o pacote computacional GLIM, e seguindo o modelo (3.9). No primeiro estudo, 10.000 amostras foram geradas, para cada combinação de valores de θ ($\theta = 0,1; 0,2; 0,4; 0,6; 0,8$ e $0,9$) e de tamanho de amostra n ($n = 5, 10, 20, 30$ e 40), sendo fixada, em cada caso, a média geral μ igual a $4,0$ e a variância σ^2 igual a 25 . Para cada uma das 10.000 amostras simuladas, encontramos as estimativas $\hat{\theta}$ e $\tilde{\theta}$ e computamos as médias amostrais destas estimativas. A estimativa $\hat{\theta}$ é obtida através de macro especial que ajusta o modelo MA(1) aos dados simulados. A Tabela 3.5 fornece os valores obtidos.

Tabela 3.5: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\theta}$ e não corrigidas $\hat{\theta}$, do parâmetro θ do modelo MA(1), com $\mu = 4,0$ e $\sigma^2 = 25$.

n	θ		$0,10$		$0,20$		$0,40$		$0,60$		$0,80$		$0,90$	
	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$	$\tilde{\theta}$
5	0,17	0,13	0,36	0,28	0,71	0,53	0,89	0,51	0,71	0,85	0,75	0,84		
10	0,16	0,14	0,31	0,25	0,59	0,51	0,83	0,72	0,97	0,68	0,83	0,95		
20	0,14	0,14	0,30	0,26	0,48	0,46	0,76	0,70	0,90	0,84	0,98	0,85		
30	0,14	0,13	0,25	0,24	0,46	0,45	0,69	0,67	0,87	0,83	0,95	0,89		
40	0,13	0,12	0,24	0,23	0,45	0,42	0,68	0,66	0,84	0,83	0,94	0,92		

Verificamos, como esperado, que, para tamanhos de amostra grandes ($n = 30$ ou 40), a correção pelo viés produz pouca melhora nas estimativas de máxima verossimilhança. No entanto, quando os tamanhos das amostras são pequenos ($n = 5$ ou 10), a superioridade das estimativas corrigidas é bastante efetiva, no sentido de torná-las mais próximas dos valores verdadeiros do parâmetro θ , particularmente para θ próximo de zero. As estimativas não corrigidas excedem os parâmetros verdadeiros em quase todos os casos, exceto para $n = 5$ e $\theta = 0,8$ e $n = 5$ e 10 e $\theta = 0,9$. Nos outros dois estudos comparamos o desempenho

das estimativas $\hat{\sigma}^2$ e $\tilde{\sigma}^2$. Geramos 10.000 amostras para cada combinação de valores de θ ($\theta = 0, 20, 0, 40, 0, 60$ e $0, 80$) e de tamanho de amostra n ($n = 5, 10, 20, 30$ e 40), mantendo fixa a média geral μ igual a 4,0 e a variância σ^2 , igual a 10 num caso, e igual a 50 no outro. Para cada uma das 10.000 amostras simuladas encontramos as estimativas de máxima verossimilhança corrigidas $\tilde{\sigma}^2$ e não corrigidas $\hat{\sigma}^2$ e calculamos as médias amostrais destas estimativas. Os resultados encontrados estão dispostos nas Tabelas 3.6 e 3.7, referentes a $\sigma^2 = 10$ e $\sigma^2 = 50$, respectivamente.

Tabela 3.6: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\sigma}^2$ e não corrigidas $\hat{\sigma}^2$, do parâmetro σ^2 do modelo MA(1), com $\mu = 4, 0$ e $\sigma^2 = 10$.

n	θ	0,20		0,40		0,60		0,80	
		$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$
5		6,09	8,69	5,94	8,73	5,56	11,45	11,31	9,27
10		7,74	8,75	7,81	8,90	6,79	8,33	6,34	11,36
20		8,46	8,93	8,83	9,31	7,93	8,58	8,70	9,51
30		8,85	9,14	9,19	9,47	8,31	8,69	8,97	9,38
40		8,97	9,21	9,31	9,49	8,56	8,84	9,35	9,63

Tabela 3.7: Médias das estimativas simuladas de máxima verossimilhança corrigidas $\tilde{\sigma}^2$ e não corrigidas $\hat{\sigma}^2$, do parâmetro σ^2 do modelo MA(1), com $\mu = 4, 0$ e $\sigma^2 = 50$.

n	θ	0,20		0,40		0,60		0,80	
		$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$
5		32,79	44,07	30,52	46,51	29,31	55,14	56,32	44,45
10		39,54	45,21	41,21	47,31	41,05	48,71	27,96	55,17
20		44,05	46,71	44,95	47,83	43,61	46,88	41,82	45,81
30		46,58	48,54	46,17	48,04	45,44	47,62	44,41	46,44
40		47,64	48,92	47,58	48,73	46,51	47,90	46,88	47,92

Os valores das Tabelas 3.6 e 3.7 mostram que as estimativas de máxima verossimilhança de σ^2 corrigidas pelo viés têm sempre um desempenho melhor do que as não corrigidas, no sentido

de estarem mais próximas do valor verdadeiro de σ^2 . No entanto, a correção é bastante efetiva para tamanhos de amostra pequenos e em algumas situações particulares: i) quando $\sigma^2 = 10$, para $n = 5$ e $\theta = 0, 2, 0,4$ e $0,6$ e para $n = 10$ e $\theta = 0,6$ e $0,8$ e ii) quando $\sigma^2 = 50$, para $n = 5$ e $\theta = 0, 2, 0,4$ e $0,6$ e para $n = 10$ e $\theta = 0, 2$ e $0,8$.

Em todos os casos descritos nas Tabelas 3.1 a 3.7, a correção pelo viés sinalizou corretamente, ou seja, conduziu a estimativa corrigida na direção do parâmetro verdadeiro. Através dos estudos de simulação apresentados nesta seção podemos concluir que há evidências de que o desempenho das estimativas de máxima verossimilhança corrigidas pelos vieses é superior ao das estimativas não corrigidas, mostrando que a correção proposta aqui é efetiva para melhorar as estimativas de máxima verossimilhança, até ordem n^{-1} , no sentido de torná-las mais próximas dos parâmetros verdadeiros, particularmente, se temos um número reduzido de observações.

CAPÍTULO 4

CORREÇÕES DE BARTLETT PARA MODELOS NA FAMÍLIA EXPONENCIAL UNIPARAMÉTRICA

4.1 INTRODUÇÃO

O fator de correção de Bartlett para a estatística da razão de verossimilhança, como mencionado anteriormente, envolve, em geral, funções complicadas de cumulantes conjuntos de derivadas da função de log-verossimilhança até quarta ordem. É, portanto, fundamental obter fórmulas simplificadas para estes fatores em classes importantes de modelos, para que a metodologia de aperfeiçoamento dos testes, visando à melhoria de sua aproximação assintótica, não se restrinja apenas ao campo teórico, mas seja realmente utilizada em problemas práticos.

Com isto em mente, desenvolvemos, no Capítulo 2, uma expressão matricial simples para a quantidade ε_{p+q} na classe dos modelos normais lineares em que a estrutura de covariância é

não-escalar, mas depende de um conjunto de parâmetros. Neste capítulo nossa atenção está voltada para as distribuições uniparamétricas na família exponencial. Esta família de distribuições goza de vasta aplicação prática e de muitas propriedades matemáticas úteis, como apresentamos no decorrer do capítulo, ou podem ser encontradas em Bickel e Doksum (1977), por exemplo. O livro de Barndorff-Nielsen (1978) expõe um desenvolvimento matemático completo das idéias sobre esta família incluindo inúmeros conceitos importantes como ancilaridade, convexidade, otimização, estimação, testes e suficiência. Mais precisamente, considere a função de probabilidade ou função densidade escrita por

$$\pi(y; \theta) = \frac{1}{\xi(\theta)} \exp\{-\alpha(\theta)d(y) + v(y)\}, \quad (4.1)$$

onde θ é um parâmetro escalar, $\xi(\cdot)$, $\alpha(\cdot)$, $d(\cdot)$ e $v(\cdot)$ são funções conhecidas e assume-se que o suporte de $\pi(y; \theta)$ não depende de θ . Também, supomos que as funções α e ξ têm as quatro primeiras derivadas contínuas com respeito a θ , $\xi(\cdot)$ é positiva e $d\alpha(\theta)/d(\theta)$ e $d\beta(\theta)/d(\theta)$ são diferentes de zero no espaço paramétrico, onde $\beta(\theta)$ é definido como $\beta(\theta) = (d\xi(\theta)/d(\theta))(\xi(\theta)d\alpha(\theta)/d\theta)^{-1}$. A família de distribuições (4.1) compreende muitas distribuições importantes que utilizamos como casos especiais (vide Seção 4.3). O problema considerado é o de testar a hipótese nula simples $H : \theta = \theta^{(0)}$ contra uma alternativa bilateral, onde $\theta^{(0)}$ é um valor especificado. Para esta situação, escrevemos a correção de Bartlett como

$$c = 1 + \frac{\rho(\theta)}{12n}, \quad (4.2)$$

e fornecemos uma expressão em forma fechada para $\rho(\theta)$ na Seção 4.2. Nosso objetivo aqui é obter uma nova expressão simples para $\rho(\theta)$, que é algebricamente mais interessante do que as fórmulas gerais desenvolvidas por Lawley (1956) e McCullagh e Cox (1986). Nossa fórmula é facilmente aplicável e envolve somente operações triviais com certas funções e suas derivadas. Consideramos, também, como uma aplicação do nosso resultado, o teste de homogeneidade de parâmetros em p populações independentes.

Na Seção 4.3 mostramos que nosso resultado pode ser usado facilmente para obtermos

expressões simples em muitos casos especiais. As distribuições que tratamos são muito úteis para modelar dados em várias ciências aplicadas. Considere, por exemplo, a distribuição em série logarítmica introduzida por Fisher *et al.* (1943) para estudar amostragem de borboletas. Esta distribuição tem muitas aplicações em ecologia (vide Williams, 1944, 1964) e negócios (vide Chatfield, 1986, e Chatfield *et al.*, 1966). As outras distribuições consideradas têm aplicações igualmente importantes e um breve relato do uso e propriedades de algumas delas é feito na Seção 4.3. Em alguns casos, o cálculo de $\rho(\theta)$ requer a avaliação de funções poligamas, zeta ou de Bessel. Fornecemos, em tais situações, algumas aproximações simples que simplificam bastante suas expressões e são úteis na prática.

Para algumas das distribuições estudadas fornecemos gráficos de $\rho(\theta)$ versus θ . Assim, podemos verificar como a correção varia com θ . Isto é feito na Seção 4.4. A Seção 4.5 considera correções de Bartlett para algumas classes de funções de variância em modelos na família exponencial natural. Finalizando, estudos de simulação comprovando a importância das correções desenvolvidas são apresentados na Seção 4.6.

4.2 OBTENÇÃO DA CORREÇÃO DE BARTLETT

Seja $Y_1, \dots, Y_n$ um conjunto de n variáveis aleatórias independentes e identicamente distribuídas com função densidade (ou de probabilidade) dada por

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

onde θ é um parâmetro escalar. Observe-se que qualquer distribuição de probabilidade uniparamétrica pode ser incluída nesta forma, pois a função $t(y; \theta)$ é genérica. O logaritmo da função de verossimilhança, dados os valores observados $y_1, \dots, y_n$ de $Y_1, \dots, Y_n$, respectivamente, é escrito por

$$\ell(\theta) = \sum_{i=1}^n t(y_i; \theta) . \quad (4.3)$$

A estatística da razão de verossimilhança para testar a hipótese nula $H : \theta = \theta^{(0)}$ contra a hipótese alternativa $A : \theta \neq \theta^{(0)}$, sendo $\theta^{(0)}$ um valor especificado, é

$$\omega = 2 \sum_{i=1}^n \{t(y_i; \hat{\theta}) - t(y_i; \theta^{(0)})\} ,$$

onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ . Para obtermos uma expressão para o fator de correção de Bartlett c associado à estatística ω , necessitamos, primeiramente, introduzir alguma notação. No que segue, $v_r = v_r(\theta) = E\{t^{(r)}(y; \theta)\}$, $r = 1, 2, 3$ e 4 , com $t^{(r)}(y; \theta) = d^r t(y; \theta)/d\theta^r$. A notação que segue é usada para cumulantes de derivadas do logaritmo da função de verossimilhança $\ell = \ell(\theta)$ (Lawley, 1956; Hayakawa, 1977): $\kappa_{\theta\theta} = E(d^2 \ell/d\theta^2)$, $\kappa_{\theta\theta\theta} = E(d^3 \ell/d\theta^3)$, $\kappa_{\theta\theta\theta\theta} = E(d^4 \ell/d\theta^4)$, $\kappa_{\theta\theta}^{(\theta)} = d\kappa_{\theta\theta}/d\theta$, $\kappa_{\theta\theta}^{(\theta\theta)} = d^2 \kappa_{\theta\theta}/d\theta^2$, $\kappa^{\theta\theta} = 1/\kappa_{\theta\theta}$.

Da expansão de Lawley (1956) segue que, em problemas regulares, conforme já visto em capítulo anteriores, $E(\omega) = 1 + \varepsilon_1 + O(n^{-2})$ e $c = 1 + \varepsilon_1$, com $\varepsilon_1 = \ell_{\theta\theta\theta\theta} - \ell_{\theta\theta\theta\theta\theta}$, onde

$$\begin{aligned} \ell_{\theta\theta\theta\theta} &= (\kappa^{\theta\theta})^2 \left(\frac{1}{4} \kappa_{\theta\theta\theta\theta} - \kappa_{\theta\theta\theta}^{(\theta)} + \kappa_{\theta\theta}^{(\theta\theta)} \right) , \\ \ell_{\theta\theta\theta\theta\theta} &= (\kappa^{\theta\theta})^3 \left\{ \kappa_{\theta\theta\theta} \left(\frac{\kappa_{\theta\theta\theta}}{6} - \kappa_{\theta\theta}^{(\theta)} \right) + \kappa_{\theta\theta\theta} \left(\frac{\kappa_{\theta\theta\theta}}{4} - \kappa_{\theta\theta}^{(\theta)} \right) + 2 \left(\kappa_{\theta\theta}^{(\theta)} \right)^2 \right\} . \end{aligned}$$

Para a função em (4.3) temos

$$\frac{d^2 \ell}{d\theta^2} = \sum_{i=1}^n t^{(2)}(y_i; \theta), \quad \frac{d^3 \ell}{d\theta^3} = \sum_{i=1}^n t^{(3)}(y_i; \theta) \quad \text{e} \quad \frac{d^4 \ell}{d\theta^4} = \sum_{i=1}^n t^{(4)}(y_i; \theta) ,$$

resultando,

$$\begin{aligned} \kappa_{\theta\theta} &= nv_2 , \quad \kappa_{\theta\theta\theta} = nv_3 , \quad \kappa_{\theta\theta\theta\theta} = nv_4 , \\ \kappa_{\theta\theta}^{(\theta)} &= nv'_2 , \quad \kappa_{\theta\theta}^{(\theta\theta)} = nv''_2 , \quad \kappa_{\theta\theta\theta}^{(\theta)} = nv'_3 , \end{aligned}$$

onde v'_2 , v''_2 e v'_3 denotam, respectivamente, as primeira e segunda derivadas de v_2 e a primeira derivada de v_3 . Substituindo as quantidades acima na expressão de ε_1 e aplicando a fórmula de Lawley a uma distribuição uniparamétrica geral, obtemos, depois de alguma manipulação algébrica, que c é dado por (4.2) com

$$\rho(\theta) = \frac{3v_2(v_4 - 4v'_3 + 4v''_2) - v_3(5v_3 - 24v'_2) - 24v_2'^2}{v_2^3} . \quad (4.4)$$

O próximo passo é assumir uma forma exponencial para $\pi(\cdot; \cdot)$. Seja $\pi(y; \theta)$ como definido em (4.1); assim, $t(y; \theta) = -\log \xi(\theta) - \alpha(\theta)d(y) + v(y)$. Definimos $\beta = \beta(\theta) = \xi' / (\xi \alpha')$, onde $\xi' = \xi'(\theta) = d\xi(\theta)/d\theta$ e $\alpha' = \alpha'(\theta) = d\alpha(\theta)/d\theta$. Obtemos a estimativa de máxima verossimilhança $\hat{\theta}$ resolvendo a equação $-n^{-1} \sum_{i=1}^n d(y_i) = \beta(\theta)$, e o critério da razão de verossimilhança para testar $H_0 : \theta = \theta^{(0)}$ pode ser escrito como

$$\omega = 2n\beta(\hat{\theta})\{\alpha(\hat{\theta}) - \alpha(\theta^{(0)})\} + 2n \log\{\xi(\theta^{(0)})/\xi(\hat{\theta})\}. \quad (4.5)$$

Temos, então, que $t'(y; \theta) = -\alpha'\{\beta + d(y)\}$, e, desde que $E\{t'(y; \theta)\} = 0$ (por condições de regularidade), segue que $E\{d(y)\} = -\beta$, como esperado. Não é difícil mostrarmos que

$$v_2 = -\alpha'\beta', \quad v_3 = -2\alpha''\beta' - \alpha'\beta'' \text{ e } v_4 = -3(\alpha'''\beta' + \alpha''\beta''') - \alpha'\beta''',$$

onde linhas denotam derivadas em relação a θ .

Usando as relações acima, é possível simplificarmos a expressão para $\rho(\theta)$ em (4.4). Após uma intensa manipulação algébrica, que omitimos aqui, encontramos

$$\rho(\theta) = -\frac{4\beta'^2\alpha''^2 + \alpha'\beta'\alpha''\beta'' - 5\alpha'^2\beta''^2 - 3\alpha'\beta'^2\alpha''' + 3\alpha'^2\beta'\beta'''}{(\alpha'\beta')^3}. \quad (4.6)$$

Vale ressaltar que a expressão para $\rho(\theta)$, dada em (4.6), requer somente conhecimento das funções α e β e de suas três primeiras derivadas. Poderíamos, também, ter expresso $\rho(\theta)$ em (4.6) em termos da função α e de suas três primeiras derivadas, juntamente com a função ξ e suas quatro primeiras derivadas, embora seja muito mais simples deixar $\rho(\theta)$ como na equação (4.6). Quando α é uma função linear em θ , que corresponde à *família exponencial natural*, (4.6) reduz-se a $\rho(\theta) = (5\beta''^2 - 3\beta'\beta''')/\beta'^3$, que está de acordo com a equação (13) de Cordeiro (1983) (vide Seção 4.5). Os outros termos em (4.6) dependem do afastamento desta forma simples da família exponencial. Em particular, $\rho(\theta)$ é igual a zero para distribuições na família exponencial natural para a qual a função β satisfaz a equação diferencial $5\beta''^2 = 3\beta'\beta'''$. Neste caso, a estatística da razão de verossimilhança ω é distribuída como qui-quadrado com 1 grau de liberdade para uma segunda ordem de aproximação (e não primeira ordem). Assim, podemos

obter, facilmente, a correção de Bartlett para qualquer distribuição dada por (4.1), bastando fornecer as correspondentes funções α , β e suas derivadas envolvidas em (4.6).

O objetivo do desenvolvimento feito é obter uma forma simples para $\rho(\theta)$ e não explicar sua estrutura geral. A dificuldade principal em interpretar (4.6) é que os termos individuais não são invariantes sob reparametrização e, portanto, não têm interpretação geométrica que é independente do sistema de coordenadas escolhido. Na Seção 4.5, uma interpretação muito simples da correção de Bartlett é dada via uma reparametrização.

Na Seção 4.3 verificamos que para várias distribuições encontramos $\rho(\theta) = 2$. Usando a equação (4.6) podemos provar, facilmente, que $\rho(\theta) = 2$ se: (i) $\alpha(\theta)\xi(\theta) = c_1$ ou se (ii) $\alpha(\theta)$ é linear, digamos $\alpha(\theta) = c_1\theta + c_2$, e $\xi(\theta) = c_3/(\theta c_4)$, onde c_1, c_2, c_3 e c_4 são constantes arbitrárias. Estas condições (individualmente) são suficientes, mas não necessárias, para garantir $\rho(\theta) = 2$. Na Seção 4.3 consideramos várias distribuições para as quais as condições (i) e (ii) são satisfeitas.

Apresentamos, agora, uma aplicação importante da equação (4.6) ao teste da razão de verossimilhança ω de homogeneidade de parâmetros em p populações independentes tomadas de (4.1). Sejam Y_{ij} , $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, n_i$, variáveis aleatórias independentes com função densidade ou de probabilidade (4.1) com parâmetros θ_i , para $i = 1, 2, \dots, p$. A estatística da razão de verossimilhança para testar a hipótese nula $H_0 : \theta_1 = \theta_2 = \dots = \theta_p (= \theta)$, contra a hipótese alternativa H_1 : os θ 's não são todos iguais, é

$$\omega = 2 \sum_{i=1}^p \sum_{j=1}^{n_i} \{t(y_{ij}; \hat{\theta}_i) - t(y_{ij}; \tilde{\theta})\} ,$$

onde $\hat{\theta}_i$ é o estimador de máxima verossimilhança irrestrito de θ_i e $\tilde{\theta}$ é o estimador de máxima verossimilhança de θ sob H_0 . Temos, então, sob H_0 e ignorando termos de ordem n_i^{-2} , que

$$\begin{aligned} E(\omega) &= E \left\{ 2 \sum_{i=1}^p \sum_{j=1}^{n_i} [t(y_{ij}; \hat{\theta}_i) - t(y_{ij}; \theta)] \right\} - E \left\{ 2 \sum_{i=1}^p \sum_{j=1}^{n_i} [t(y_{ij}; \tilde{\theta}) - t(y_{ij}; \theta)] \right\} \\ &= p - 1 + \frac{\rho(\theta)}{12} \left[\sum_{i=1}^p \frac{1}{n_i} - \frac{1}{n_+} \right] , \end{aligned}$$

onde $\rho(\theta)$ é dado em (4.6) e $n_+ = \sum_{i=1}^p n_i$. Portanto, a correção de Bartlett é dada por

$$c = 1 + \frac{\rho(\theta)}{12(p-1)} \left[\sum_{i=1}^p \frac{1}{n_i} - \frac{1}{n_+} \right].$$

A última expressão é uma generalização da clássica correção de Bartlett para o teste de homogeneidade de variâncias em p populações normais com médias conhecidas. Segue do caso (xx, a) na Seção 4.3 que $\rho(\theta) = 4$. A aplicação apresentada acima pode ser vista como uma generalização deste resultado bastante conhecido (Bartlett, 1937).

4.3 ALGUNS CASOS ESPECIAIS

Nesta seção mostramos que a expressão de $\rho(\theta)$ em (4.6) pode ser usada a fim de obtermos fórmulas simples em forma fechada para a correção de Bartlett da estatística da razão de verossimilhança em muitas distribuições importantes. Consideramos trinta casos especiais e encontramos expressões em forma fechada para $\rho(\theta)$, obtidas usando o pacote computacional MATHEMATICA (Wolfram, 1991). O programa é bastante simples. Basta fornecermos as funções α e β e o programa realiza todo o resto, fornecendo no final uma expressão fechada para $\rho(\theta)$. Qualquer outro pacote de computação simbólica poderia ser usado com este objetivo.

Os casos considerados aqui, na verdade, cobrem mais de trinta distribuições, já que alguns deles são famílias de distribuições. Por exemplo, o sistema Burr de distribuições cobre dez distribuições (todas distribuições no sistema Burr, com exceção da Burr I e Burr IX). Várias outras distribuições em (4.1) poderiam, também, ser analisadas. A maior parte das distribuições que estudamos é bastante conhecida e tem um vasto campo de aplicação prática, tais como, engenharia, biologia, zoologia, economia e medicina, entre outros. Casos (i) a (x) envolvem variáveis aleatórias discretas enquanto variáveis aleatórias contínuas são consideradas nos casos (xi) a (xxx). Devemos destacar que embora as correções de Bartlett conduzam, usualmente, a uma melhora na taxa de convergência para uma distribuição qui-quadrado em modelos contínuos, não existe garantia total que isto também ocorra para modelos discretos (vide Frydenberg e

Jensen (1989)). Entretanto, os resultados de simulação, apresentados na Seção 4.6, mostram que, até mesmo, no caso discreto a correção de Bartlett produz uma melhora na realização do teste da razão de verossimilhança em amostras pequenas, comprovando a importância da equação (4.6), na prática, para algumas distribuições discretas.

Apresentamos, a seguir, algumas considerações e particularidades que caracterizam algumas das distribuições que estudamos, sem sermos, no entanto, exaustivos. O leitor que necessitar de maiores detalhes e propriedades de todas as distribuições que trabalhamos, pode consultar os nove volumes da “Encyclopedia of Statistical Sciences”(1982-1989), publicados pela John Wiley and Sons e editados por S. Kotz e N.L. Johnson. Alternativamente, pode consultar os livros clássicos seguintes, bem como, os artigos neles referidos: “Handbook of statistical distributions”, publicado pela Marcel Dekker, Inc., 1976, de J.K. Patel, C.H. Kapadia e D.B. Owen, “Discrete distributions” e “Continuous univariate distributions I and II”, estes três últimos publicados pela Houghton Mifflin Company, 1970, de autoria de N.L. Johnson e S. Kotz.

As distribuições em (i) – (iii) são bastante conhecidas na literatura estatística e, portanto, dispensam apresentações. A distribuição binomial negativa representa o número de ensaios onde não ocorre sucesso antes de ocorrer sucesso pela r -ésima vez. Apesar da binomial negativa ser uma distribuição discreta ela é bastante usada na análise de dados contínuos, cuja variância é superior à média, ou seja, quando se tem dados super-dispersos. Ela pode ser caracterizada como uma mistura de distribuições de Poisson e gama (Cordeiro, 1992). A análise de dados contínuos com super-dispersão, via a distribuição binomial negativa, é justificável se considerarmos somas de quantidades aleatórias ao longo de um intervalo de tempo, também aleatório, ao contrário do tempo ser fixado. Uma outra caracterização da binomial negativa, usada extensivamente na análise de dados biológicos, é ter variância quadrática, $V(Y) = \theta + \theta^2/\gamma$ e não proporcional à média. Quando temos uma distribuição binomial negativa $BN(\theta, \gamma)$, tal que $\theta\gamma = 1$, então temos caracterizada a distribuição geométrica, que pode ser pensada como repre-

sentando o número de ensaios antes do primeiro sucesso, em experimentos de Bernoulli, cuja probabilidade de sucesso é $\gamma/(1+\gamma)$. Esta distribuição é a única entre as discretas caracterizada pela ausência de memória.

Como sabemos, existe uma vasta literatura tratando das propriedades e utilizações práticas do modelo Poisson. Podemos citar muitos textos em probabilidade e estatística que fornecem boas introduções à distribuição de Poisson, como por exemplo, Feller (1957), Munroe (1951), Wilks (1962), Lindley (1965), entre outros, e do ponto de vista de física nuclear, Evans (1955) e Chandrasekhar (1943). Entretanto, não é muito freqüente destacar a possibilidade de modificar as probabilidades Poisson, pela omissão de algum ou alguns valores do “range” de possíveis valores, por exemplo, o valor zero. Haight (1967) comenta sobre algumas situações práticas em que é adequado adotar-se um modelo de Poisson truncado no zero. Segundo Haight (1967), um exemplo clássico da Poisson truncada resulta em conexão com o nascimento de crianças com defeito genético específico (por exemplo, albinismo) em famílias portadoras do gene apropriado. Se o caráter genético dos pais pode ser observado somente por meio do nascimento de tais crianças, todas as famílias com crianças normais ou com nenhuma criança, escaparão da contagem, até mesmo se elas puderem pertencer à população sendo amostrada. Neste caso, o suporte da distribuição será $y = 1, 2, \dots$ crianças defeituosas. O leitor que se interessar por outras aplicações da Poisson ou Poisson truncada pode consultar o livro texto de Haight (1967, Capítulo 7).

A série logarítmica tem inúmeras aplicações em termos de contagens. Pode, ainda, ser usada no contexto econômico, para representar a demanda de itens de certo tipo durante um período de tempo especificado. Para valores grandes dessa variável a forma da cauda desta distribuição é similar à da geométrica.

A família de distribuições em série de potências fornece uma formulação muito elegante e perceptiva de várias distribuições discretas clássicas que são utilizadas em pesquisa estatística

e ensino. Casos especiais da distribuição em série de potências são as distribuições binomial, Poisson, binomial negativa e a série logarítmica.

Considerem-se, agora, duas binomiais independentes, $X \sim B(m_1, \pi_1)$ e $Y \sim B(m_2, \pi_2)$, com probabilidades de sucesso distintas, π_1 e π_2 , respectivamente. A distribuição de X condicional ao número total de sucessos nas duas variáveis ser igual a r , isto é, $X + Y = r$, é denominada hipergeométrica não-central, sendo caracterizada pelos parâmetros m_1 , m_2 , r e $\theta = \pi_1(1 - \pi_2)/\{\pi_2(1 - \pi_1)\}$. Quando $\pi_1 = \pi_2$, então $\theta = 1$, e neste caso temos a distribuição hipergeométrica central.

A distribuição de Maxwell é utilizada na física para estudar a magnitude de velocidade. Suponhamos que X_1 , X_2 e X_3 representam os componentes de velocidade de uma molécula de gás e que X_1 , X_2 e X_3 sejam independentes, cada uma com distribuição normal $N(0, \sigma^2)$. Então, a variável $Y = (X_1^2 + X_2^2 + X_3^2)^{1/2}$, denominada magnitude da velocidade, tem uma distribuição de Maxwell.

A distribuição gama $G(\theta, k)$ é usada na análise de dados contínuos não-negativos, os quais apresentam uma variância crescente com a média e mais, fundamentalmente, quando o coeficiente de variação dos dados for aproximadamente constante. O caso $k = 1$, corresponde à distribuição exponencial, que desempenha papel importante na descrição de uma grande classe de fenômenos, particularmente, nos assuntos de teoria da confiabilidade e no estudo da lei de falhas. Quando $\theta = \frac{1}{2}$ e $k = n/2$ obtemos a distribuição χ_n^2 . Outro caso especial desta distribuição ocorre quando $\theta = 1$ e k é um número inteiro, que corresponde à distribuição de Erlang. Podemos, também, considerar a variável aleatória obtida tomando-se o logaritmo da variável aleatória com distribuição gama $G(\theta, k)$. Neste caso, temos a distribuição log-gama, amplamente utilizada em análise de sobrevivência.

O sistema Burr de distribuições foi construído em 1941 por Irving W. Burr. Uma vez que as funções de densidade correspondentes têm uma grande variedade de formas, este sistema é útil

para ajustar histogramas, particularmente, quando uma estrutura matemática simples para a função de distribuição acumulada ajustada é requerida. Outras aplicações incluem simulação, resposta quantal, aproximação de distribuições e desenvolvimento de cartas de controle não-normal. A motivação original para desenvolver o sistema Burr foi fornecer um método para ajustar funções de distribuição acumuladas a dados frequentistas. As funções de distribuição acumuladas $F(x)$ do sistema Burr satisfazem uma equação diferencial, análoga à equação diferencial que gera o sistema Pearson (vide Encyclopedia of Statistical Sciences, v.1, 1982, p.109, de N.L. Johnson e S. Kotz).

A distribuição de Rayleigh foi derivada, originalmente, por Lord Rayleigh (1919) num problema em acústica. Ela tem sido muito utilizada no campo da engenharia elétrica. Considere uma amostra aleatória de tamanho n , $Y_1, Y_2, \dots, Y_n$, de uma distribuição normal $N(0, \sigma^2)$. A função densidade de probabilidade da variável X , distância da origem ao ponto $(Y_1, \dots, Y_n)$, num espaço Euclidiano n -dimensional, ou seja, $X = (\sum_{i=1}^n Y_i^2)^{1/2}$, tem distribuição de Rayleigh. O caso especial tratado em (xvii) corresponde a $n = 2$. A distribuição de Maxwell (caso (xi)) é obtida fazendo $n = 3$.

Já a distribuição de Pareto $PA(\theta, k)$ tem bastante utilidade em modelagem de variáveis econômicas como, por exemplo, distribuições de renda, sendo o parâmetro k a moda da distribuição. Para θ e k fixos a densidade é uma função não-crescente de y . Sua distribuição acumulada é frequentemente usada para representar a chance da renda de um indivíduo ser limitada por y , quando a renda mínima possível é k .

A variável aleatória Y com distribuição de Weibull $W(\phi, \theta)$ é definida por $Y = \theta Z^{1/\phi}$, sendo Z distribuída exponencialmente com média unitária. Ela tem sido usada, principalmente, em confiabilidade de sistemas, análise de sobrevivência e, ainda, em controle de qualidade, quando as condições usuais de aleatoriedade, restritas à distribuição exponencial, não são verificadas. Uma outra aplicação desta distribuição ocorre em microeconomia, na teoria do consumidor,

para maximização da função de utilidades, dado que o máximo de uma variável aleatória Weibull é também uma variável aleatória com distribuição de Weibull. Quando $\phi = 2$, temos a distribuição de Rayleigh e para $\phi = 1$, a distribuição resultante é exponencial.

A distribuição de Laplace surge da diferença de duas variáveis aleatórias independentes com distribuição exponencial.

A distribuição do valor extremo aparece quase naturalmente no estudo do efeito do tamanho em resistência dos materiais, na ocorrência de enchentes e estiagens, na confiabilidade de sistemas obtidos a partir de um grande número de componentes e na avaliação dos níveis de poluição do ar.

A lognormal, chamada algumas vezes de antilognormal, é usada para estudar o diâmetro de pequenas partículas após um processo de trituração, o tamanho de um organismo sujeito a alguns pequenos impulsos e, também, a duração da vida de determinadas peças. Suponhamos que X é $N(\mu, \sigma^2)$, então $Y = e^X$ possui distribuição lognormal, ou seja, Y é lognormal se $\log Y$ é normal.

As aplicações do modelo normal inverso são o estudo do movimento Browniano de partículas, análise de regressão com dados consideravelmente assimétricos, testes de confiabilidade, análise sequencial e análogo de análise de variância para classificações encaixadas. Outras aplicações incluem modelagem de tempos, como: duração de greves, tempo de primeira passagem nos passeios aleatórios, tempo de sobrevivência, tempo gasto para injetar uma substância no sistema biológico, etc.

Propriedades importantes da distribuição de McCullagh são encontradas em seu artigo fundamental, McCullagh (1989), onde são abordados, também, aspectos sobre estimação e testes de hipóteses.

A distribuição de von Mises é apropriada para estudar dados direcionais no círculo unitário.

Sua densidade é simétrica em torno de $y = \mu$, sendo μ a moda da distribuição. Por outro lado, θ é um parâmetro de precisão no sentido de que quanto maior for θ , maior é a concentração da densidade ao redor de μ ; se $\theta = 0$, a distribuição é uniforme no intervalo unitário, enquanto que para valores grandes de θ , aproxima-se da normal. Propriedades e aplicações do modelo de von Mises, bem como um estudo sobre estimação e testes de hipóteses, são encontradas no livro de Mardia (1972). Não fazemos menção à distribuição Zeta (caso xxix) e à distribuição secante hiperbólica generalizada (caso xxx), porque há pouca informação disponível na literatura estatística tratando de seus usos em aplicações práticas.

Apresentamos, agora, a expressão de $\rho(\theta)$, dado em (4.6), juntamente com as funções $\alpha(\theta)$, $\xi(\theta)$, $d(y)$ e $v(y)$, o espaço paramétrico e o suporte de y , os quais juntos podem reproduzir a função densidade ou de probabilidade, usando a forma (4.1), das distribuições que consideramos:

(i) Bernoulli ($0 < \theta < 1$, $y = 0, 1$):

$$\alpha(\theta) = -\log\{\theta/(1-\theta)\}, \quad \xi(\theta) = (1-\theta)^{-1}, \quad d(y) = y, \quad v(y) = 0;$$

$$\rho(\theta) = \frac{2(1-\theta+\theta^2)}{\theta(1-\theta)}.$$

(ii) Binomial ($0 < \theta < 1$, $m \in \mathbb{N}$, m conhecido, $y = 0, 1, 2, \dots, m$):

$$\alpha(\theta) = -\log\{\theta/(1-\theta)\}, \quad \xi(\theta) = (1-\theta)^{-m}, \quad d(y) = y, \quad v(y) = \log\binom{m}{y};$$

$$\rho(\theta) = \frac{2(1-\theta+\theta^2)}{m\theta(1-\theta)}.$$

(iii) Geométrica ($0 < \theta < 1$, $y = 0, 1, 2, \dots$):

$$\alpha(\theta) = -\log(1-\theta), \quad \xi(\theta) = \theta^{-1}, \quad d(y) = y, \quad v(y) = 0;$$

$$\rho(\theta) = \frac{2(1-\theta+\theta^2)}{1-\theta}.$$

(iv) Binomial negativa ($0 < \theta < 1$, $\gamma > 0$ conhecido, $y = 0, 1, 2, \dots$):

$$\alpha(\theta) = -\log \theta, \quad \xi(\theta) = (1 - \theta)^{-\gamma}, \quad d(y) = y, \quad v(y) = \log \left(\frac{\gamma + y - 1}{y} \right);$$

$$\rho(\theta) = \frac{2(1 - \theta + \theta^2)}{\gamma\theta}.$$

(v) Poisson ($\theta > 0, y = 0, 1, 2, \dots$):

$$\alpha(\theta) = -\log \theta, \quad \xi(\theta) = \exp \theta, \quad d(y) = y, \quad v(y) = -\log(y!);$$

$$\rho(\theta) = 2\theta^{-1}.$$

(vi) Poisson truncada ($\theta > 0, y = 1, 2, \dots$):

$$\alpha(\theta) = -\log \theta, \quad \xi(\theta) = e^\theta(1 - e^{-\theta}), \quad d(y) = y, \quad v(y) = -\log(y!);$$

$$\begin{aligned} \rho(\theta) = & -\{2 + 6\theta + 16\theta^2 + 9\theta^3 + 2\theta^4 + e^\theta(-8 - 18\theta - 40\theta^2 - 9\theta^3 - 2\theta^4) \\ & + e^{2\theta}(12 + 18\theta + 32\theta^2 - 3\theta^3 + 2\theta^4) + e^{3\theta}(-8 - 6\theta - 8\theta^2 + 3\theta^3) + 2e^{4\theta}\} / \\ & \{\theta e^\theta(1 + \theta - e^\theta)^3\}. \end{aligned}$$

(vii) Série logarítmica ($0 < \theta < 1, y = 1, 2, \dots$):

$$\alpha(\theta) = -\log \theta, \quad \xi(\theta) = -\log(1 - \theta), \quad d(y) = y, \quad v(y) = -\log y;$$

$$\begin{aligned} \rho(\theta) = & \left\{ \theta[\theta + \log(1 - \theta)]^3 \right\}^{-1} \left\{ -2\theta^4 - 6\theta^3 \log(1 - \theta) - 8(\theta^2 + \theta^3)[\log(1 - \theta)]^2 \right. \\ & \left. - 3(2\theta + 2\theta^2 - \theta^3)[\log(1 - \theta)]^3 - 2(1 - \theta + \theta^2)[\log(1 - \theta)]^4 \right\}. \end{aligned}$$

(viii) Série de potências ($\theta > 0, a_y \geq 0, y = 0, 1, 2, \dots$):

$$\alpha(\theta) = -\log \theta, \quad \xi(\theta) = \sum_{y=0}^{\infty} a_y \theta^y, \quad d(y) = y, \quad v(y) = \log a_y;$$

$$\rho(\theta) = \frac{2g^2 + 6\theta gg' + 8\theta^2(3g'^2 - gg'') + 3\theta^3(4g'g'' - gg''') + \theta^4(5g''^2 - 3g'g''')}{\theta(g + \theta g')^3},$$

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

(ix) Zeta ($\theta > 0, y = 1, 2, 3, \dots$):

$$\alpha(\theta) = \theta + 1, \quad \xi(\theta) = \text{Zeta}(\theta + 1), \quad d(y) = \log y, \quad v(y) = 0;$$

$$\rho(\theta) = \frac{5g''^2 - 3g'g'''}{g'^3},$$

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

(x) Hipergeométrica não-central ($\theta > 0, m_1, m_2, r$ inteiros positivos conhecidos, $a = \max\{0, r - m_2\} \leq y \leq \min\{m_1, r\} = b$):

$$\alpha(\theta) = \theta, \quad \xi(\theta) = D_0(\theta), \quad d(y) = -y, \quad v(y) = \log \left\{ \binom{m_1}{y} \binom{m_2}{r-y} \right\};$$

$$\begin{aligned} \rho(\theta) = & \left\{ -2D_1^6 + 6D_0D_1^4D_2 - 8D_0^2D_1^3D_3 + 18D_0^3D_1D_2D_3 - 3D_0^3D_1^2D_4 - 9D_0^3D_2^3 \right. \\ & \left. + 3D_0^4D_2D_4 - 5D_0^4D_3^2 \right\} / \left\{ (D_1^2 - D_0D_2)^3 \right\}, \end{aligned}$$

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

(xi) Maxwell ($\theta > 0, y > 0$):

$$\alpha(\theta) = (2\theta^2)^{-1}, \quad \xi(\theta) = \theta^3, \quad d(y) = y^2, \quad v(y) = \log(y^2 \sqrt{2/\pi});$$

$$\rho(\theta) = \frac{4}{3}.$$

(xii) Erlang ($\theta > 0, y > 0$):

$$\alpha(\theta) = (1 - \theta), \quad \xi(\theta) = \Gamma(\theta), \quad d(y) = \log y, \quad v(y) = -y;$$

$$\rho(\theta) = \frac{5\psi''(\theta)^2 - 3\psi'(\theta)\psi'''(\theta)}{\psi'(\theta)^3},$$

onde $\Gamma(\cdot)$ é a função gama e $\psi(\cdot) = d \log \Gamma(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ é a função digama.

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

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

$$\rho(\theta) = 2\kappa^{-1};$$

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

$$\rho(k) = \frac{5\psi''(k)^2 - 3\psi'(k)\psi'''(k)}{\psi'(k)^3}.$$

(xiv) Log-gama ($\theta > 0, -\infty < \mu < \infty, -\infty < y < \infty$):

(a) μ conhecido: $\alpha(\theta) = \theta$, $\xi(\theta) = \Gamma(\theta)$, $d(y) = -(y - \mu)$, $v(y) = -\exp(y - \mu)$;

$$\rho(\theta) = \frac{5\psi''(\theta)^2 - 3\psi'(\theta)\psi'''(\theta)}{\psi'(\theta)^3}.$$

(b) θ conhecido: $\alpha(\mu) = \exp\{-\mu\}$, $\xi(\mu) = \exp\{\theta\mu\}$, $d(y) = \exp(y)$, $v(y) = \theta y - \log \Gamma(\theta)$;

$$\rho(\theta) = 2\theta^{-1}.$$

(xv) Sistema Burr de distribuições ($\theta > 0, b > 0$ conhecido, $y > 0$):

$\alpha(\theta) = \theta$, $\xi(\theta) = c(\theta)/\theta$, $d(y) = -\log G(y)$, $v(y) = \log\{|d \log G(y)/dy|\}$;

$$\rho(\theta) = 2,$$

onde as funções $c(\cdot)$ e $G(\cdot)$ assumem valores reais positivos. Escolhas diferentes de $c(\theta)$ e $G(y)$ conduzem a diferentes distribuições (vide Burr, 1942).

(xvi) Exponencial ($\theta > 0, y > 0$):

$\alpha(\theta) = \theta^{-1}$, $\xi(\theta) = \theta$, $d(y) = y$, $v(y) = 0$;

$$\rho(\theta) = 2.$$

(xvii) Rayleigh ($\theta > 0, y > 0$):

$\alpha(\theta) = \theta^{-2}$, $\xi(\theta) = \theta^2$, $d(y) = y^2$, $v(y) = \log(2y)$;

$$\rho(\theta) = 2.$$

(xviii) Pareto ($\theta > 0$, $k > 0$ conhecido, $y > k$):

$$\alpha(\theta) = \theta + 1, \quad \xi(\theta) = (\theta \kappa^\theta)^{-1}, \quad d(y) = \log y, \quad v(y) = 0;$$

$$\rho(\theta) = 2 .$$

(xix) Weibull ($\theta > 0$, $\phi > 0$ conhecido, $y > 0$):

$$\alpha(\theta) = \theta^{-\phi}, \quad \xi(\theta) = \theta^\phi, \quad d(y) = y^\phi, \quad v(y) = \log \phi + (\phi - 1) \log y;$$

$$\rho(\theta) = 2 .$$

(xx) Potência ($\theta > 0$, $\phi > 0$ conhecido, $0 < y < \phi$):

$$\alpha(\theta) = 1 - \theta, \quad \xi(\theta) = \theta^{-1} \phi^\theta, \quad d(y) = \log y, \quad v(y) = 0;$$

$$\rho(\theta) = 2 .$$

(xxi) Laplace ($\theta > 0$, $-\infty < k < \infty$ conhecido, $y > 0$):

$$\alpha(\theta) = \theta^{-1}, \quad \xi(\theta) = 2\theta, \quad d(y) = |y - k|, \quad v(y) = 0;$$

$$\rho(\theta) = 2 .$$

(xxii) Valor extremo ($-\infty < \theta < \infty$, $\phi > 0$ conhecido, $-\infty < y < \infty$):

$$\alpha(\theta) = \exp\{\theta/\phi\}, \quad \xi(\theta) = \phi \exp\{-\theta/\phi\}, \quad d(y) = \exp\{-y/\phi\}, \quad v(y) = -y/\phi;$$

$$\rho(\theta) = 2 .$$

(xxiii) Valor extremo truncado ($\theta > 0$, $y > 0$):

$$\alpha(\theta) = \theta^{-1}, \quad \xi(\theta) = \theta, \quad d(y) = \exp\{y\} - 1, \quad v(y) = y;$$

$$\rho(\theta) = 2 .$$

(xxiv) Lognormal ($\theta > 0$, $\mu > 0$ conhecido, $y > 0$):

$$\alpha(\theta) = \theta^{-2}, \quad \xi(\theta) = \theta, \quad d(y) = (\log y - \mu)^2/2, \quad v(y) = -\log y + \log(2\pi)/2;$$

$$\rho(\theta) = 4 .$$

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

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

$$\rho(\theta) = 4 .$$

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

$$\rho(\mu) = 0 .$$

(xxvi) Normal inversa ($\theta > 0$, $\mu > 0$, $y > 0$):

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

$$\rho(\theta) = 4 .$$

$$(b) \quad \theta \text{ conhecido: } \alpha(\mu) = \theta(2\mu^2)^{-1}, \quad \xi(\mu) = \exp\{-\theta/\mu\}, \quad d(y) = y, \quad v(y) = -\theta(2y)^{-1} + \log[\theta/(2\pi y^3)]/2;$$

$$\rho(\mu) = 0 .$$

(xxvii) Qui-quadrado ($\theta > 0$, $y > 0$):

$$\alpha(\theta) = (2 - \theta)/2, \quad \xi(\theta) = 2^{\theta/2}\Gamma(\theta/2), \quad d(y) = \log y, \quad v(y) = -y/2;$$

$$\rho(\theta) = \frac{5\psi''(\theta/2)^2 - 3\psi'(\theta/2)\psi'''(\theta/2)}{\psi'(\theta/2)^3} .$$

(xxviii) McCullagh ($\theta > -1/2$, $-1 \leq \mu \leq 1$ conhecido, $0 < y < 1$):

$$\alpha(\theta) = -\theta, \quad \xi(\theta) = 4^{-\theta} B(\theta + 1/2, 1/2), \quad d(y) = \log\{y(1-y)/[(1+\mu)^2 - 4\mu y]\}, \quad v(y) = -\log[y(1-y)]/2;$$

$$\rho(\theta) = \left\{ -5[\psi''(\theta+1) - \psi''(\theta + \frac{1}{2})]^2 + 3[\psi'(\theta+1) - \psi'(\theta + \frac{1}{2})][\psi'''(\theta+1) - \psi'''(\theta + \frac{1}{2})] \right\} / \left\{ [\psi'(\theta+1) - \psi'(\theta + \frac{1}{2})]^3 \right\},$$

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

(xxix) Von Mises ($\theta > 0$, $0 < \mu < 2\pi$ conhecido, $0 < y < 2\pi$):

$$\alpha(\theta) = -\theta, \quad \xi(\theta) = 2\pi I_0(\theta), \quad d(y) = \cos(y - \mu), \quad v(y) = 0;$$

$$\rho(\theta) = \frac{5r''(\theta)^2 - 3r'(\theta)r'''(\theta)}{r'(\theta)^3},$$

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

(xxx) Secante hiperbólica generalizada ($-\pi/2 \leq \theta \leq \pi/2$, $r > 0$ conhecido, $0 < y < 1$):

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

$$\rho(\theta) = \frac{2\{1 - 4(\cos\theta)^2\}}{r}.$$

É interessante notar que para algumas distribuições a correção não depende do valor especificado sob a hipótese nula, mas este não é sempre o caso. Em algumas situações, a correção de Bartlett, e assim a aproximação χ^2 de primeira ordem para o teste ω da razão de verossimilhança, depende do valor de θ especificado na hipótese nula H (vide Seção 4.4 para maiores detalhes). É de se destacar, também, que nossa expressão geral para $\rho(\theta)$, em (4.6), pode gerar fórmulas simples (por exemplo, $\rho(\theta) = 2$) e expressões complexas (por exemplo, $\rho(\theta)$ para as distribuições Poisson truncada, série logarítmica, hipergeométrica não-central e série de potências) para casos especiais diferentes. Nos casos em que a expressão para $\rho(\theta)$ é bastante complexa, é importante e útil obtermos expansões assintóticas que produzam aproximações

simples para $\rho(\theta)$ e não necessitam de avaliação de funções complicadas, tais como, funções poligamas e de Bessel ou função zeta. Começamos com a distribuição gama (θ conhecido). Para valores grandes de k , temos que (Abramowitz e Stegun, 1970; fórmula 6.4.12)

$$\psi'(k) = \frac{1}{k} + \frac{1}{2\kappa^2} + \frac{1}{6\kappa^3} - \frac{1}{30\kappa^5} + \frac{1}{42\kappa^7} + O(\kappa^{-9}) .$$

É possível usarmos esta expansão de $\psi'(\cdot)$ para simplificar a expressão de $\rho(k)$ para a distribuição gama. Após alguma manipulação algébrica, via o MATHEMATICA, obtemos, para valores grandes de k ,

$$\rho(k) = -\frac{1}{k} - \frac{1}{2\kappa^2} + \frac{1}{2\kappa^3} + O(\kappa^{-4}) .$$

Esta expressão é muito conveniente já que não requer a avaliação de funções poligamas. Ela também é válida para a distribuição log-gama (μ conhecido) e de Erlang, com k trocado por θ . Notamos, também, que um resultado similar para a distribuição qui-quadrado pode ser obtido, apenas, trocando k por $\theta/2$ na expressão acima, isto é, para a distribuição qui-quadrado com θ graus de liberdade resulta

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

que constitui uma aproximação para valores grandes de θ . Fazendo desenvolvimentos algébricos semelhantes e usando a relação

$$\psi'(\theta + 1) - \psi'(\theta + \frac{1}{2}) = 2\psi'(\theta) - 4\psi'(2\theta) - \frac{1}{\theta^2} ,$$

obtemos, para valores grandes de θ , a seguinte expansão assintótica para $\rho(\theta)$ considerando a distribuição de McCullagh:

$$\rho(\theta) = 4 + \frac{3}{2\theta^2} - \frac{3}{4\theta^3} + O(\theta^{-4}) .$$

Este resultado fornece uma boa aproximação para θ , quando θ é grande. Notamos, por exemplo, que para esta expressão $\lim_{\theta \rightarrow \infty} \rho(\theta) = 4$, que está de acordo com o gráfico mostrado na Figura 4.4 da Seção 4.4.

Consideramos, agora, a distribuição de von Mises, onde $\rho(\theta)$ é expresso em termos de funções de Bessel de primeira espécie. Para valores grandes de θ , podemos escrever $r(\theta)$ como (Abramowitz e Stegun, 1970, relações 9.6.26, 9.6.27, 9.7.1, 9.7.2)

$$r(\theta) = 1 - \frac{1}{2\theta} - \frac{1}{8\theta^2} - \frac{1}{8\theta^3} - \frac{25}{\theta^4} - \frac{13}{32\theta^5} + \dots$$

Usando este resultado conseguimos, para valores grandes de θ , a seguinte aproximação:

$$\rho(\theta) = 4 - \frac{9}{2\theta^2} - \frac{57}{2\theta^3} + O(\theta^{-4}) .$$

Para valores pequenos de θ podemos usar o fato que (Mardia, 1972, p.63)

$$r(\theta) = \frac{\theta}{2} \left\{ 1 - \frac{\theta^2}{8} + \frac{\theta^4}{48} - \dots \right\} ,$$

para obtermos a aproximação abaixo para $\rho(\theta)$,

$$\rho(\theta) = \frac{9}{2} + \frac{3\theta^2}{2} + O(\theta^4) .$$

Estas aproximações, ao contrário da expressão para $\rho(\theta)$ dada em (xxix), não envolvem funções de Bessel e podem ser avaliadas facilmente. Vejamos, agora, a distribuição em série logarítmica.

Para valores pequenos de θ , temos que

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

Usando este resultado, e depois de um trabalho algébrico bastante extenso e tedioso, que omitimos aqui, obtemos para a distribuição em série logarítmica, com θ pequeno, a fórmula

$$\rho(\theta) = \frac{4}{\theta} - \frac{10}{3} + \frac{7\theta}{9} + \frac{74\theta^2}{135} + \frac{67\theta^3}{162} + O(\theta^4) ,$$

que fornece boas aproximações para valores de θ até 0,3. Para a distribuição de Poisson truncada fazemos uso da expansão em série da função exponencial, ou seja, escrevemos $e^\theta = 1 + \theta + \theta^2/2 + \theta^3/6 + \theta^4/24 + \dots$ para obter a seguinte expansão para θ pequeno:

$$\rho(\theta) = \frac{4}{\theta} - \frac{2}{3} + \frac{10\theta}{9} - \frac{89\theta}{135} + O(\theta^3) .$$

Finalmente, consideramos a distribuição zeta. Para encontrar uma expressão aproximada para $\rho(\theta)$ em (ix), usamos a expansão da função zeta de Riemann, que denotamos por $\xi(\cdot)$, dada em Abramowitz e Stegun (1970, p. 807, relação 23.2.5), escrita por

$$\xi(\theta + 1) = \frac{1}{\theta} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma_n \theta^n ,$$

onde os coeficientes γ_n são calculados por

$$\gamma_n = \lim_{m \rightarrow \infty} \left\{ \sum_{k=1}^m \frac{(\ln k)^n}{k} - \frac{(\ln m)^{n+1}}{n+1} \right\} ,$$

sendo γ_0 a constante de Euler, isto é, $\gamma_0 \cong 0,5772$. A expansão acima implica que

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

Como $g(\theta) = d \log \xi(\theta + 1) / d\theta = \xi'(\theta + 1) / \xi(\theta + 1)$ segue, até ordem θ^3 , que

$$\begin{aligned} g(\theta) = & -\frac{1}{\theta} + \gamma_0 - (2\gamma_1 + \gamma_0^2)\theta + (\gamma_0^3 + 3\gamma_0\gamma_1 + \frac{3}{2}\gamma_2)\theta^2 \\ & - (\gamma_0^4 + 4\gamma_0^2\gamma_1 + 2\gamma_0\gamma_2 + 2\gamma_1^2 + \frac{2}{3}\gamma_3)\theta^3 + O(\theta^4) . \end{aligned}$$

Tomando as três primeiras derivadas da expansão acima para $g(\theta)$ e substituindo na expressão de $\rho(\theta)$ em (ix), obtemos, após intenso trabalho algébrico, que, para valores pequenos de θ ,

$$\begin{aligned} \rho(\theta) = & 2 + 24\theta^2(\gamma_0^2 + 2\gamma_1) - 44\theta^3(2\gamma_0^3 + 6\gamma_0\gamma_1 + 3\gamma_2) + 4\theta^4(69\gamma_0^4 \\ & + 276\gamma_0^2\gamma_1 + 171\gamma_1^2 + 105\gamma_0\gamma_2 + 35\gamma_3) + O(\theta^5) . \end{aligned}$$

É possível simplificar ainda mais a expressão acima, fazendo uma avaliação numérica dos γ 's.

Usando o pacote algébrico MAPLE V (Abell e Braselton, 1994) para fazer tal cálculo, obtemos

$$\rho(\theta) = 2 + \frac{42594}{9463}\theta^2 - \frac{137395}{30206}\theta^3 + O(\theta^4) .$$

Assim, quando $\theta \cong 0$, temos $\rho(\theta) \cong 2$. Uma simplificação de $\rho(\theta)$, para valores grandes de θ será objeto de estudo futuro, uma vez que a expansão apropriada da função zeta de Riemann, neste caso, tem uma expressão extremamente complicada, sendo função dos zeros da função

zeta. Entretanto, observando a Figura 4.7 na Seção 4.4, é possível concluirmos que quando $\theta \rightarrow \infty$, $\rho(\theta)$ diverge para ∞ .

A fim de avaliarmos o desempenho das expansões que usamos no desenvolvimento acima, calculamos os valores exatos e aproximados da quantidade $\rho(\theta)$, para alguns valores de θ . Os resultados encontrados estão nas Tabelas 4.1 a 4.6, correspondentes, respectivamente, às distribuições gama, McCullagh, von Mises, série logarítmica, Poisson truncada e zeta. Observando estas tabelas verificamos que as expressões aproximadas de $\rho(\theta)$ têm um desempenho bastante satisfatório, uma vez que seus valores estão muito próximos dos valores exatos de $\rho(\theta)$ em certas regiões do espaço paramétrico. Este desempenho é confirmado, também, graficamente (vide Seção 4.4).

Tabela 4.1: Valores exatos e aproximados de $\rho(k)$, k grande, para a distribuição gama (θ conhecido).

k	$\rho(k)$ exato	$\rho(k)$ aproximado
1,0	-0,70716	-1,00000
2,0	-0,51864	-0,56250
3,0	-0,35976	-0,37037
10,0	-0,10450	-0,10450

Tabela 4.2: Valores exatos e aproximados de $\rho(\theta)$, θ grande, para a distribuição de McCullagh.

θ	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado
1,0	4,35036	4,75000
2,0	4,19520	4,28125
3,0	4,11372	4,08333
4,0	4,00044	4,08203
5,0	4,04964	4,05400
10,0	4,01448	4,01425

Tabela 4.3: Valores exatos e aproximados de $\rho(\theta)$, θ pequeno e θ grande, para a distribuição de von Mises.

θ (pequeno)	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado	θ (grande)	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado
0,1	4,51500	4,5150	3,0	8,21232	2,444
0,5	4,87008	4,8750	5,0	3,66252	3,592
1,0	5,90856	6,0000	7,0	3,58740	3,825
1,5	7,32972	7,8750	9,0	3,83532	3,905
2,0	8,55756	10,500	10,0	3,88812	3,926

Tabela 4.4: Valores exatos e aproximados de $\rho(\theta)$, θ pequeno, para a distribuição em série logarítmica.

θ	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado
0,05	76,70700	76,70698
0,10	36,75000	36,75034
0,15	23,46396	23,46373
0,20	16,84812	16,84746
0,25	12,90348	12,90183
0,30	10,29744	10,29383

Tabela 4.5: Valores exatos e aproximados de $\rho(\theta)$, θ pequeno, para a distribuição de Poisson truncada.

θ	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado
0,05	79,38732	79,38724
0,10	39,43812	39,43785
0,15	26,15256	26,15183
0,20	19,53096	19,52919
0,30	12,94644	12,94067
0,40	9,68568	9,67230
0,50	7,74960	7,72407
0,70	5,56908	5,50236
0,90	4,37916	4,24378
1,00	3,96672	3,78519

Tabela 4.6: Valores exatos e aproximados de $\rho(\theta)$, θ pequeno, para a distribuição zeta.

θ	$\rho(\theta)$ exato	$\rho(\theta)$ aproximado
0,05	2,01072	2,01068
0,10	2,04096	2,04046
0,15	2,08824	2,08592
0,20	2,15088	2,14366
0,30	2,31576	2,28229
0,40	2,52696	2,37334
0,50	2,77920	2,55670

4.4 ANÁLISE GRÁFICA

Fica claro dos resultados da seção anterior que para várias distribuições ρ não é uma constante, mas uma função de θ , variando, portanto, com este parâmetro. Uma questão interessante é, então, como valores diferentes de θ , especificados na hipótese nula, afetam a correção de Bartlett e, assim, a aproximação de primeira ordem para a distribuição χ_1^2 ? Para esclarecer esta questão, fazemos o gráfico de $\rho(\theta)$ contra θ para algumas distribuições consideradas na seção anterior. Isto mostra como a correção muda com valores diferentes de θ . Apresentamos, aqui, gráficos de $\rho(\theta)$ versus θ para as seguintes distribuições: Poisson truncada, secante hiperbólica generalizada ($r = 1$), série logarítmica, McCullagh, Bernoulli, gama (com θ conhecido), zeta e von Mises. Estes gráficos são apresentados nas Figuras 4.1 a 4.8, respectivamente. Simultaneamente desenhemos nos gráficos as aproximações apresentadas na seção anterior, sendo que linha sólida corresponde a $\rho(\theta)$ e linha tracejada à aproximação.

Figura 4.1: Poisson truncada

Figura 4.2: Secante hiperbólica generalizada

Figura 4.3: Série logarítmica

Figura 4.4: McCullagh

Figura 4.5: Bernoulli

Figura 4.6: Gama

Figura 4.7: Zeta

Figura 4.8: Von Mises

Fica claro da Figura 4.1 que para a distribuição de Poisson truncada a correção torna-se muito grande para valores pequenos de θ e muito pequena para valores grandes de θ . A aproximação dada na seção anterior funciona bem para $\theta < 1$. Para a distribuição secante hiperbólica generalizada (xxx) observamos que a correção mostra um comportamento oscilatório, variando de -6 a 2 (Figura 4.2). No caso da distribuição em série logarítmica (Figura 4.3), a correção torna-se grande para valores de θ próximos a 0 ou 1 , e nossa aproximação funciona bem para valores de 0 até $0,7$. O gráfico de $\rho(\theta)$ versus θ para a distribuição de McCullagh (Figura 4.4) mostra um comportamento bastante interessante: $\rho(\theta)$ aumenta bem rápido para θ pequeno, e depois alcançando um pico em, aproximadamente, $4,4$, decresce continuamente aproximando-se de um valor assintótico quando θ cresce. Aqui, $\lim_{\theta \rightarrow \infty} \rho(\theta) = 4$. Nossa aproximação, neste caso, parece ter um bom desempenho para $\theta > 2$. Para a distribuição de Bernoulli (Figura 4.5), a correção tem forma de U , alcançando um mínimo para $\theta = 1/2$. No caso da distribuição gama como θ conhecido (Figura 4.6), a correção está próxima de 2 quando θ está próximo de zero, decresce rapidamente e torna-se negativa e, então, converge para zero quando θ cresce. Para este caso, nossa aproximação produz resultados satisfatórios para $k > 2$. O gráfico de $\rho(\theta)$ versus θ para a distribuição zeta é dada na Figura 4.7. É possível visualizar que $\rho(\theta)$ é aproximadamente igual a 2 para valores pequenos de θ , e que $\rho(\theta)$ diverge para ∞ quando $\theta \rightarrow \infty$. Está claro neste gráfico que a aproximação utilizada somente fornece bons resultados quando $\theta < 1$. Finalmente, considere a distribuição de von Mises (Figura 4.8). Para valores pequenos de θ , $\rho(\theta) \cong 4,5$, e pode ser mostrado que $\lim_{\theta \rightarrow 0} \rho(\theta) = 4,5$. Entretanto, é interessante notar que $\rho(\theta)$ cresce, substancialmente, para θ entre 1 e 3 . Neste caso, a aproximação para θ pequeno funciona bem para, digamos, $\theta < 1,75$, e a aproximação para θ grande produz resultados satisfatórios para $\theta > 8$. Quando $\rho(\theta) > 0$ e $\rho(\theta) < 0$ (vide Figuras 4.1 a 4.8) as taxas de rejeição de ω devem, respectivamente, superestimar e subestimar os níveis nominais baseados na χ^2_1 . Este fato está de acordo com os resultados de simulação obtidos na Seção 4.6.

4.5 FAMÍLIA EXPONENCIAL NATURAL

Nesta seção nossa atenção está voltada ao estudo da família exponencial natural, que é indexada pelo parâmetro natural α , na forma

$$\pi(y; \alpha) = \frac{1}{\delta(\alpha)} \exp\{-\alpha d(y) + v(y)\} , \quad (4.7)$$

onde $-d(y)$ é a estatística canônica. Condições semelhantes àsquelas assumidas para a função densidade ou de probabilidade em (4.1) são válidas aqui. $\delta(\alpha)$ é o gerador de cumulantes da função $-d(y)$, isto é, o r -ésimo momento de $-d(y)$ é $\kappa_r = d^r \log \delta(\alpha)/d\alpha^r$, para $r = 1, 2, \dots$. Também, $\kappa_{r+1} = d^r \beta/d\alpha^r$, para $r = 1, 2, \dots$, onde $\beta(\alpha) = \delta'(\alpha)/\delta(\alpha)$. Em particular, a média e variância de $-d(y)$ são $\kappa_1 = \beta$ e $\kappa_2 = d\beta/d\alpha = \beta'$. No que segue, reservamos as linhas para denotarem derivadas em relação a α . A função de variância β' não depende da particular parametrização usada, e é uma ferramenta poderosa para manipular correções de Bartlett nos modelos da família exponencial por duas razões. Primeiro, ela define a distribuição em (4.7). Segundo, é muito mais simples expressar a função de variância do que a função geradora $\delta(\alpha)$. O primeiro fato foi originalmente sugerido por Jørgensen (1987), que mostrou que a função de variância define unicamente, a menos de uma transformação linear, a distribuição em (4.7).

O critério da razão de verossimilhança para testar a hipótese nula $H : \theta = \theta^{(0)}$ é invariante para reparametrizações, e já que (4.7) define uma correspondência biunívoca entre α e θ , a equação (4.6) reduz-se a

$$\rho(\alpha) = 5\gamma_1^2 - 3\gamma_2 , \quad (4.8)$$

onde $\gamma_1^2 = \beta''^2/\beta'^3$ e $\gamma_2 = \beta'''/\beta'^2$ são o terceiro e quarto cumulantes padronizados de $-d(y)$. A expressão (4.8) foi, primeiramente, apresentada por Cordeiro (1983) e McCullagh e Cox (1986). Esta fórmula tem uma forma mais elegante do que (4.6) embora esta última seja mais conveniente, uma vez que fornece a correção de Bartlett em termos do parâmetro original θ . A Figura 4.9 apresenta o gráfico tri-dimensional de $\rho(\alpha)$ contra γ_1 e γ_2 . Notamos que o campo

Figura 4.9: Família exponencial natural

de variação de $\rho(\alpha)$ é limitado pelos valores γ_1 e γ_2 satisfazendo as relações $\gamma_1^2 + \gamma_2 + 2 \geq 0$ e $\gamma_2 - \gamma_1^2 \geq 1$, que implicam em $-3\gamma_2 \leq \rho(\alpha) \leq 2\gamma_1^2 - 3$. Está claro, ainda, desta figura que $\rho(\alpha)$ decresce quando γ_2 cresce, assumindo, assim, valores negativos para valores grandes de γ_2 quando γ_1 é pequeno. Também, $\rho(\alpha)$ mostra um comportamento quadrático como uma função de γ_1 para dado γ_2 .

A correção de Bartlett em (4.8) reduz-se a uma constante $k \neq 0$ quando a função geradora é dada por $\delta(\alpha) = c_1 \exp\{c_2\alpha\}(k\alpha + c_3)^{-2/k}$, onde os c_i 's são constantes arbitrárias. As funções $\delta(\alpha)$ de todos os casos especiais considerados na Seção 4.3, tendo correções não-nulas, verificam esta condição. Por outro lado, as correções de Bartlett desaparecem quando $\delta(\alpha) = \exp\{c_1 + c_2\alpha + c_3\alpha^2\}$ ou $\delta(\alpha) = \exp\{c_1 + c_2\alpha^{1/2} + c_3\alpha\}$, onde, novamente, os c_i 's são constantes arbitrárias com $c_3 \neq 0$. Dos 30 casos especiais considerados na Seção 4.3, somente dois casos produzem $\rho(\alpha) = 0$: o caso da normal com variância conhecida, para o qual a primeira equação está satisfeita, e o caso da normal inversa com índice θ conhecido, cuja função $\delta(\alpha)$ satisfaz a última equação.

O objetivo agora é fornecer correções de Bartlett em forma fechada para famílias especiais

Figura 4.10: Função de variância potência

de funções de variância. Começamos com a função de variância potência definida por $\beta' = \beta^p/c$, com c uma constante positiva, onde o domínio de β é $\mathbb{R}$ para $p = 0$ e $\mathbb{R}_+$ quando $p \neq 0$. A correção de Batlett para esta função de variância reduz-se a

$$\rho(\beta) = \frac{p(3-p)\beta^{p-2}}{c}. \quad (4.9)$$

A equação (4.9) cobre muitos casos importantes, incluindo as distribuições: normal com θ conhecido ($p = 0$), Poisson ($p = 1$), gama com $k = 1$ ($p = 2$) e normal inversa com θ conhecido ($p = 3$). No caso da Poisson temos $c = 1$, uma vez que a variância é igual à média. Outros valores de p definem um número de distribuições que foram classificadas por Jørgensen (1987). A função de variância potência com $p \leq 0$ corresponde a distribuições geradas por distribuições estáveis extremas cujo suporte é $\mathbb{R}$. Para $1 < p < 2$ e para $p > 2$, (4.9) corresponde a certas distribuições de Poisson compostas e a distribuições geradas por distribuições estáveis positivas, respectivamente. Em ambos os casos, as distribuições são contínuas e têm suporte em $\mathbb{R}_+$. Os casos restantes com potências em $0 < p < 1$ não correspondem a distribuições em (4.9). O gráfico tri-dimensional de $\rho(\beta)$ contra β e p é apresentado na Figura 4.10. A correção desaparece somente para $p = 0$ e $p = 3$. Está claro, também, que $\rho(\beta)$ assume valores negativos grandes

quando β é pequeno e p é negativo.

Consideramos, agora, a família de funções de variância definida como funções polinomiais de ordem menor ou igual a 3, digamos, $\beta' = c_0 + c_1 + c_2\beta^2 + c_3\beta^3$. Para o caso onde $c_3 = 0$, Morris (1982) define seis tipos de família exponencial natural. Letac e Mora (1990) estendem a classificação de Morris e mostram que existem somente seis tipos de distribuições com a forma (4.7), cuja função de variância é polinomial de grau 3 em β . A fórmula geral para $\rho(\beta)$ para testar $H : \theta = \theta^{(0)}$, para o caso de função de variância cúbica é

$$\rho(\beta) = 2c_2 + \frac{2}{\beta'} \{c_1 - 4c_0c_2 - c_3\beta(9c_0 + 3c_1\beta + c_2\beta^2)\} . \quad (4.10)$$

As correções para todos 12 casos discutidos em Morris (1982) e Letac e Mora (1990) podem ser obtidas da equação (4.10), e estão apresentadas na Tabela 4.7. Para funções de variância quadrática, a correção é obtida considerando $c_3 = 0$.

As correções na Tabela 4.7 para as distribuições normal, Poisson, binomial, gama e secante hiperbólica generalizada concordam com os resultados na Seção 4.3. Duas das seis distribuições na Tabela 4.7 com função de variância cúbica são contínuas: a distribuição normal inversa e a distribuição de Ressel. As quatro distribuições restantes estão concentradas no conjunto de inteiros não-negativos $\mathbb{N}$: a distribuição de Abel, Takács e as duas distribuições arco-seno (estrito e estendido).

Tentamos, também, obter correções de Bartlett para uma classe mais ampla de funções de variância como, por exemplo, a função de variância $\beta' = P + Q\sqrt{R}$, onde P , Q e R são polinômios de grau menor ou igual a 3, 2 e 2, respectivamente. Funções de variância escritas por esta última forma incluem as funções de variância para a classe de Babel, para a qual $\beta' = b\Delta + (a\beta + c)\sqrt{\Delta}$, onde Δ é um polinômio de grau menor ou igual a 2, que não é um quadrado perfeito e a , b e c são três números reais. As expressões em forma fechada para as correções de Bartlett para estas funções de variância são obtidas via MATHEMATICA, mas não apresentamos aqui por serem muito extensas.

Tabela 4.7: Correções de Bartlett para funções de variância cúbica.

	c_0	c_1	c_2	c_3	“Range” de β	$\rho(\beta)$
normal (xxv, b)	θ	0	0	0	$\mathbb{R}$	0
Poisson (v)	0	1	0	0	$\mathbb{R}_+$	$\frac{2}{\beta}$
binomial (ii)	0	1	$-\frac{1}{m}$	0	$(0, m)$	$\frac{2(m^2 - m\beta + \beta^2)}{m\beta(m - \beta)}$
binomial negativa (iv)	0	1	$\frac{1}{\gamma}$	0	$\mathbb{R}_+$	$\frac{\beta^2 + \beta\gamma + \gamma^2}{\beta\gamma(\beta + \gamma)}$
gama (xiii, a)	0	0	$\frac{1}{k}$	0	$\mathbb{R}_+$	$\frac{2}{k}$
secante hiperbólica generalizada (xxx)	r	0	$\frac{1}{r}$	0	$\mathbb{R}$	$\frac{2(\beta^2 - 3r^2)}{r(\beta^2 + r^2)}$
Abel (p conhecido)	0	1	$\frac{2}{p}$	$\frac{1}{p^2}$	$\mathbb{R}_+$	$\frac{2}{\beta}$
Takács ($p, a > 0$ conhecido)	0	1	$\frac{2a+1}{ap}$	$\frac{a+1}{ap^2}$	$\mathbb{R}_+$	(*)
arcoseno estrito (p conhecido)	0	1	0	$\frac{1}{p^2}$	$\mathbb{R}_+$	$-\frac{2(3\beta^2 - p^2)}{\beta(\beta^2 + p^2)}$
arcoseno estendido ($p, a > 0$ conhecido)	0	1	$\frac{2}{ap}$	$\frac{1+a^2}{a^2p^2}$	$\mathbb{R}_+$	(**)
Ressel ($p > 0$)	0	0	$\frac{1}{p}$	$\frac{1}{p^2}$	$\mathbb{R}_+$	$\frac{2}{\beta+p}$
normal inversa (xxvi, b)	0	0	0	$-\frac{1}{\theta^2}$	$\mathbb{R}_+$	0

$$\text{Nota: } (*) = \frac{2\{a^2p^2 + ap(1 + 2a)\beta + (1 + a + a^2)\beta^2\}}{a\beta(p + \beta)(ap + \beta + a\beta)}$$

$$(**) = \frac{2(a^2p^2 + 2ap\beta + (1 - 3a^2)\beta^2)}{\beta\{a^2p^2 + 2ap\beta + (1 + a^2)\beta^2\}}$$

4.6 ALGUNS ESTUDOS DE SIMULAÇÃO

Apresentamos, agora, alguns estudos de simulação que visam a comprovar a importância das correções desenvolvidas. Para tal, estudamos as distribuições geométrica, binomial, Poisson truncada, série logarítmica, binomial negativa, qui-quadrado e Poisson. Em cada uma destas distribuições 10.000 amostras foram geradas para diversas combinações, variando o tamanho da amostra n , o valor do parâmetro $\theta^{(0)}$ especificado na hipótese nula H e o nível de significância nominal α ($\alpha = 5\%$ e 10%). Para cada amostra gerada, calculamos a estatística da razão de verossimilhança usual ω , e a correspondente corrigida $\omega^* = \omega/c$, sendo $c = 1 + \rho(\theta^{(0)})/12n$, e, usando os 10.000 valores simulados de ω e ω^* , calculamos as taxas de rejeição de ω e ω^* , ou seja, as proporções $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ das vezes em que os valores simulados de

ω e ω^* excedem x_α , onde x_α é o $(1 - \alpha)$ -ésimo percentil da distribuição qui-quadrado com 1 grau de liberdade. Todas as simulações foram processadas através de sub-rotinas do sistema computacional NAG-FORTRAN.

4.6.1 DISTRIBUIÇÃO GEOMÉTRICA

Para a distribuição geométrica, caso (iii) da Seção 4.3, temos que o estimador de máxima verossimilhança é $\hat{\theta} = (1 + \bar{y})^{-1}$, onde $\bar{y}$ é a média amostral, a estatística da razão de verossimilhança é expressa por $\omega = 2n\{\bar{y} \log \bar{y} - (1 + \bar{y}) \log(1 + \bar{y}) - \log \theta^{(0)} - \bar{y} \log(1 - \theta^{(0)})\}$ e a correspondente estatística corrigida é $\omega^* = \omega/c$, com $c = 1 + (1 - \theta^{(0)} + \theta^{(0)^2})/[6n(1 - \theta^{(0)})]$. Amostras foram geradas para cada combinação de valores de $\theta^{(0)}$ ($\theta^{(0)} = 0,1; 0,4; 0,6$; e $0,8$) e do tamanho da amostra n ($n = 5, 10$ e 20). As taxas de rejeição de ω e ω^* foram calculadas, baseadas em 10.000 simulações, considerando níveis de significância nominais $\alpha = 5\%$ e 10% . Os resultados encontrados estão na Tabela 4.8.

Tabela 4.8: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste de hipótese do parâmetro θ da distribuição geométrica.

Nível nominal α	n	$\theta^{(0)} = 0,1$		$\theta^{(0)} = 0,4$		$\theta^{(0)} = 0,6$		$\theta^{(0)} = 0,8$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	5,68	5,17	5,61	5,13	5,78	5,23	6,73	5,32
	10	5,21	5,12	5,26	5,08	5,61	5,19	5,94	5,21
	20	5,10	5,02	5,14	5,05	5,17	5,06	5,41	5,10
10%	5	11,04	10,31	11,27	10,43	12,55	10,44	12,67	10,91
	10	10,43	10,14	10,62	10,23	10,93	10,35	11,36	10,47
	20	10,27	10,08	10,41	10,11	10,57	10,10	10,78	10,23

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

Observando esta tabela verifica-se que, em todas as combinações consideradas para a distribuição geométrica, as taxas de rejeição de ω e ω^* são superiores aos níveis nominais fixados. No entanto, todas as taxas de rejeição de ω^* estão mais próximas dos níveis nominais do que aquelas correspondentes de ω , notadamente, para valores pequenos de $\theta^{(0)}$ (0,1 e 0,4) e tamanho de amostra n pequeno ($n = 5$). Também, verificamos que a aproximação de ω por χ^2 piora quando θ está mais próximo de 1, implicando, neste caso, que a correção torna-se mais necessária. Evidentemente, conforme esperado, para θ fixo, a diferença entre a estatística corrigida e a não corrigida diminui quando o tamanho da amostra aumenta.

4.6.2 DISTRIBUIÇÃO BINOMIAL

Para o estudo da distribuição binomial, consideramos $Y_1, \dots, Y_n$ uma amostra aleatória de tamanho n da distribuição binomial com parâmetros m (conhecido) e θ , onde θ representa a probabilidade de sucesso nas m repetições do experimento binomial e é suposta constante nas n retiradas aleatórias. Esquemáticamente, temos

Y_i = número de sucessos ($Y_i = 0, 1, \dots, m$) em m repetições do experimento binomial, $i = 1, \dots, n$;

$\bar{Y}$ = média de sucessos na amostra $Y_1, Y_2, \dots, Y_n$.

Então, o estimador de máxima verossimilhança de θ é $\hat{\theta} = \bar{y}/m$ com $\bar{y}$ sendo o valor observado de $\bar{Y}$, e a estatística da razão de verossimilhança é escrita por

$$\omega = 2n\bar{y} \log \left\{ \frac{\bar{y}(1 - \theta^{(0)})}{(m - \bar{y})\theta^{(0)}} \right\} + 2nm \log \left\{ \frac{m - \bar{y}}{m(1 - \theta^{(0)})} \right\},$$

onde $\theta^{(0)}$ é o valor de θ especificado na hipótese nula. A correção de Bartlett c para a estatística da razão de máxima verossimilhança ω , que produz a estatística corrigida $\omega^* = \omega/c$, é dada,

sob a hipótese nula $H : \theta = \theta^{(0)}$, por

$$c = 1 + \frac{(1 - \theta^{(0)} + \theta^{(0)2})}{6nm\theta^{(0)}(1 - \theta^{(0)})}.$$

Observe-se que a correção c é simétrica em relação à reta $\theta^{(0)} = 0.5$. Foram simuladas 10.000 amostras para cada combinação da binomial $(m, \theta^{(0)})$, para $m = 5$ e $\theta^{(0)} = 0,1; 0,2; 0,3; 0,4$ e $0,5$, com o tamanho de amostra n ($n = 1, 2$ e 4) e, em cada amostra, calculamos as estatísticas ω e ω^* . Com os 10.000 valores de ω e ω^* encontramos as taxas de rejeição de ω e ω^* , ou seja, as proporções $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente, para os níveis de significância nominais α de 5% e 10%. Na Tabela 4.9 estão os resultados obtidos.

Tabela 4.9: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste de hipótese do parâmetro θ da distribuição binomial $(5, \theta)$.

Nível nominal α	n	$\theta^{(0)} = 0,1$		$\theta^{(0)} = 0,2$		$\theta^{(0)} = 0,3$		$\theta^{(0)} = 0,4$		$\theta^{(0)} = 0,5$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	1	8,79	6,85	7,02	5,27	6,48	5,19	6,25	5,21	6,14	5,17
	2	6,94	5,87	6,49	5,14	5,90	5,06	5,61	5,10	5,47	4,98
	4	6,56	5,73	6,07	5,13	5,88	5,10	5,30	4,97	5,13	4,99
10%	1	13,89	12,23	12,94	10,99	12,13	10,43	11,79	10,08	11,67	10,12
	2	12,64	11,37	11,50	10,23	11,08	10,19	10,87	10,11	10,39	9,91
	4	11,31	10,48	10,67	9,85	10,42	9,79	10,38	9,94	10,23	9,94

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

É possível concluir desta tabela que o desempenho da estatística da razão de verossimilhança corrigida ω^* é sempre superior ao da não corrigida ω , no sentido de produzir taxas de rejeição mais próximas dos níveis nominais. Para o maior tamanho de amostra considerado ($n = 4$), praticamente as duas estatísticas, ω e ω^* se equivalem em termos de níveis nominais. Aqui,

ao contrário da distribuição geométrica, observa-se uma piora na aproximação de ω por χ^2 quando a probabilidade de sucesso θ (ou de fracasso $1 - \theta$) é reduzida, implicando, assim, uma necessidade maior do uso da correção de Bartlett. Este fato pode, também, ser constatado através da Figura 4.5 ($m = 1$), que mostra que $\rho(\theta)$ aumenta quando θ aproxima-se de 0 ou 1.

4.6.3 DISTRIBUIÇÃO DE POISSON TRUNCADA

A estatística da razão de verossimilhança para o teste de hipótese do parâmetro θ ($\theta > 0$) da distribuição de Poisson truncada na origem, pode ser expressa por

$$\omega = 2n\bar{y} \log \left(\frac{\hat{\theta}}{\theta^{(0)}} \right) + 2n \log \left[\frac{e^{\theta^{(0)}} (1 - e^{-\theta^{(0)}})}{e^{\hat{\theta}} (1 - e^{-\hat{\theta}})} \right],$$

onde $\theta^{(0)}$ é o valor de θ especificado na hipótese nula e $\hat{\theta}$ é a estimativa de máxima verossimilhança para θ , que é obtida resolvendo-se, iterativamente, a equação não-linear $\hat{\theta}/(1 - e^{-\hat{\theta}}) = \bar{y}$. Isto é conseguido através da subrotina C05ADF do programa NAG. O fator de correção c , que aperfeiçoa a estatística ω por $\omega^* = \omega/c$, é calculado de $c = 1 + \rho(\theta^{(0)})/12n$, onde $\rho(\theta^{(0)})$ é dado no caso (vi) da Seção 4.3. Para cada combinação de valores de $\theta^{(0)}$ ($\theta^{(0)} = 0,5, 1, 2, 5$ e 10) e de tamanho de amostra n ($n = 5, 10$ e 20), geramos 10.000 amostras e calculamos para cada amostra as estatísticas ω e ω^* . Dos 10.000 valores simulados de ω e ω^* , encontramos a taxa de rejeição de ω , $P(\omega \geq x_\alpha)$, e a taxa de rejeição de ω^* , $P(\omega^* \geq x_\alpha)$, para valores de α iguais a 5% e 10%. Os resultados são apresentados na Tabela 4.10.

Observando os valores desta tabela, concluímos que existe evidência da superioridade da aproximação por χ^2 , da estatística aperfeiçoada ω^* sobre a estatística da razão de verossimilhança usual ω , já que, para todas as situações consideradas, as taxas de rejeição de ω^* , $P(\omega^* \geq x_\alpha)$, estão mais próximas dos níveis nominais do que aquelas de ω , $P(\omega \geq x_\alpha)$. Para valores pequenos de θ a aproximação de ω por χ^2 é pior, indicando que, nestes casos, a correção de Bartlett é mais necessária. A Figura 4.1 confirma esta conclusão, já que revela que $\rho(\theta)$

Tabela 4.10: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste de hipótese do parâmetro θ da distribuição de Poisson truncada.

Nível nominal α	n	$\theta^{(0)} = 0,5$		$\theta^{(0)} = 1$		$\theta^{(0)} = 2$		$\theta^{(0)} = 5$		$\theta^{(0)} = 10$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	6,52	5,47	5,81	5,22	5,41	5,15	5,13	4,97	5,06	4,98
	10	5,78	5,21	5,37	5,13	5,22	5,08	5,07	5,04	5,05	5,02
	20	5,38	4,94	5,20	5,09	5,11	5,06	5,04	5,03	5,02	5,00
10%	5	12,32	10,27	11,17	10,13	10,74	10,09	10,21	10,17	10,14	9,93
	10	11,24	10,30	10,63	10,08	10,32	10,14	10,13	10,06	10,11	9,96
	20	10,63	10,14	10,35	10,11	10,19	10,06	10,09	9,95	10,05	9,99

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

aumenta quando θ aproxima-se de zero. Evidentemente, para θ fixado, conforme cresce o tamanho total da amostra as proporções $P(\omega^* \geq x_\alpha)$ e $P(\omega \geq x_\alpha)$ vão aproximando-se e pouca diferença faz em optar-se por um teste ou outro.

4.6.4 DISTRIBUIÇÃO EM SÉRIE LOGARÍTMICA

O estimador de máxima verossimilhança para o parâmetro θ da distribuição em série logarítmica é obtido resolvendo-se a equação não linear

$$\bar{y} = \frac{-\hat{\theta}}{(1 - \hat{\theta}) \log(1 - \hat{\theta})}. \quad (4.11)$$

Para obtê-lo podemos usar a solução aproximada (Birch, 1963), válida quando $\bar{y} < 25$, dada por

$$\hat{\theta} \doteq 1 - \left[1 + \left\{ \left(\frac{5}{3} - \frac{1}{16} \log \bar{y} \right) (\bar{y} - 1) + 2 \right\} \log \bar{y} \right]^{-1}. \quad (4.12)$$

Como $E(Y)$ é uma função crescente em θ , a equação (4.11) tem solução única.

A estatística da razão de verossimilhança é, então, expressa por

$$\omega = 2n \log \left(\frac{\log(1 - \theta^{(0)})}{\log(1 - \hat{\theta})} \right) + n\bar{y} \log \left(\frac{\hat{\theta}}{\theta^{(0)}} \right)$$

com $\hat{\theta}$ obtido, aproximadamente, usando (4.12) e $\theta^{(0)}$ é o valor de θ especificado na hipótese nula. A estatística da razão de verossimilhança aperfeiçoada é, então, $\omega^* = \omega/c$, onde $c = 1 + \rho(\theta^{(0)})/12n$ e $\rho(\theta)$ é dado no caso (vii) da Seção 4.3.

Geramos 10.000 amostras para cada combinação de valores de $\theta^{(0)}$ ($\theta^{(0)} = 0,1; 0,2; 0,4; 0,6$ e $0,8$) e de tamanhos de amostra n ($n = 5, 10$ e 20). Para cada amostra gerada encontramos ω e ω^* e através dos 10.000 valores de ω e ω^* obtemos as taxas de rejeição $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, para níveis de significância nominais de 5% e 10%. A Tabela 4.11 apresenta os resultados encontrados.

Tabela 4.11: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste de hipótese do parâmetro θ da distribuição em série logarítmica.

Nível nominal α	n	$\theta^{(0)} = 0,1$		$\theta^{(0)} = 0,2$		$\theta^{(0)} = 0,4$		$\theta^{(0)} = 0,6$		$\theta^{(0)} = 0,8$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	11,78	8,95	8,34	6,37	6,38	5,37	5,80	5,22	5,62	5,14
	10	8,63	7,17	6,47	5,67	5,49	5,17	5,37	5,16	5,30	5,10
	20	6,80	5,95	5,79	5,31	5,28	5,09	5,21	5,07	5,14	4,98
10%	5	19,91	14,72	14,48	11,69	12,02	10,10	11,22	10,17	10,97	10,31
	10	15,22	11,88	12,39	11,34	11,10	10,21	10,88	10,08	10,51	10,10
	20	12,38	11,01	11,25	10,38	10,63	10,05	10,34	9,98	10,28	9,97

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

Uma análise da Tabela 4.11 mostra que, em todos os casos considerados, o desempenho da estatística corrigida ω^* é superior ao da estatística não corrigida ω , no sentido de produzir

taxas de rejeição mais próximas dos níveis nominais. Essa superioridade é mais acentuada em amostras de tamanho pequeno ($n = 5$ ou 10) e valores pequenos de $\theta^{(0)}$, situação em que a aproximação de ω por χ^2 é pior, tornando mais necessária a correção de Bartlett. Notemos que a Figura 4.3 confirma esta conclusão, pois fornece os maiores valores de $\rho(\theta)$ quando os valores de θ aproximam-se de zero. À medida que cresce o tamanho da amostra, para um θ fixado, ambas as taxas de rejeição tendem a ser mais próximas do correspondente nível nominal, exceto, novamente, quando θ é muito reduzido, que ainda favorece o teste que utiliza a estatística corrigida ω^* .

4.6.5 DISTRIBUIÇÃO BINOMIAL NEGATIVA

Através das funções α , β , d e v da Seção 4.3, caso (iv), chegamos à função de probabilidade da distribuição binomial negativa,

$$P(Y = y) = \binom{\gamma + y - 1}{y} \theta^y (1 - \theta)^\gamma, \quad \gamma > 0, \quad y = 0, 1, 2, \dots$$

Se γ é um número natural, esta distribuição pode ser interpretada como o número de fracassos ocorridos em ensaios independentes até se ter γ sucessos, onde o evento sucesso tem probabilidade constante $1 - \theta$. Podemos usar este fato para definir $Y \sim BN(\gamma, 1 - \theta)$.

O estimador de máxima verossimilhança do parâmetro θ ($0 < \theta < 1$) da distribuição binomial negativa é dado por $\hat{\theta} = \bar{y}/(\gamma + \bar{y})$, onde $\bar{y}$ é a média amostral. Pode-se dizer que este estimador é esperado, pois $1 - \hat{\theta} = \gamma/(\gamma + \bar{y})$ é o quociente entre o número de sucessos e o número médio de ensaios, este último dado pela soma do número de sucessos especificados (γ) com o número médio de fracassos $\bar{y}$. Temos aqui, $E(Y) = \gamma\theta/(1 - \theta)$ e $\text{Var}(Y) = \gamma\theta/(1 - \theta)^2$. A estatística da razão de verossimilhança para o teste da hipótese nula $H : \theta = \theta^{(0)}$ é calculada por

$$\omega = 2n \log \left\{ \left[\frac{\bar{y}}{\theta^{(0)}(\gamma + \bar{y})} \right]^{\bar{y}} \left[\frac{\gamma}{(1 - \theta^{(0)})(\gamma + \bar{y})} \right]^\gamma \right\}.$$

Assim, a estatística da razão de verossimilhança aperfeiçoada é $\omega^* = \omega/c(\theta^{(0)})$, onde $c(\theta) = 1 + (1 - \theta + \theta^2)/6n\gamma\theta$.

Através do pacote computacional GLIM versão 3.77, processado no computador VAX, geramos, para cada combinação de tamanho de amostra n ($n = 5$ e 10), valor da probabilidade de fracasso $\theta^{(0)}$ ($\theta^{(0)} = 0,1; 0,3; 0,5; 0,7$ e $0,9$) e para cada valor de γ ($\gamma = 1,4$ e 10), 10.000 amostras. Para cada conjunto de 10.000 amostras simuladas calculamos as taxas de rejeição de ω e ω^* , ou seja, $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente, onde x_α é o $(1 - \alpha)$ -percentil da distribuição χ_1^2 , com nível de significância nominal α , igual a 5% e 10%. A Tabela 4.12 mostra os resultados desta simulação.

Tabela 4.12: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste de hipótese do parâmetro θ da distribuição binomial negativa.

Nível nominal α	n	γ	$\theta^{(0)} = 0,1$		$\theta^{(0)} = 0,3$		$\theta^{(0)} = 0,5$		$\theta^{(0)} = 0,7$		$\theta^{(0)} = 0,9$	
			(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	1	8,33	6,17	6,51	5,81	5,85	5,23	5,69	5,18	5,42	5,11
		4	6,50	5,64	5,79	5,40	5,47	5,14	5,40	5,12	5,33	5,09
		10	5,78	5,16	5,46	5,08	5,32	5,07	5,25	5,08	5,21	5,04
	10	1	6,54	5,76	5,97	5,36	5,81	5,29	5,61	5,25	5,31	5,15
		4	5,60	5,27	5,39	5,20	5,35	5,14	5,33	5,14	5,25	5,10
		10	5,43	5,15	5,23	5,07	5,21	5,10	5,18	5,06	5,14	5,03
	5	1	14,55	11,29	13,02	11,35	12,35	11,20	12,16	10,81	11,97	11,01
		4	12,34	10,77	11,38	10,42	10,95	10,43	11,21	10,72	11,62	10,81
		10	11,80	10,53	11,11	10,30	10,81	10,48	10,76	10,48	10,49	10,33
10%	10	1	13,81	11,87	12,61	11,01	11,47	10,85	11,64	10,89	10,93	10,41
		4	12,01	11,01	11,27	10,59	10,78	10,49	10,79	10,93	10,61	10,30
		10	10,97	10,48	10,59	10,33	10,44	10,28	10,53	10,25	10,28	10,05

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

Da Tabela 4.12 fica claro que, para todos os casos simulados, a taxa de rejeição de ω^* , $P(\omega^* \geq x_\alpha)$, está mais próxima dos níveis nominais (5% e 10%) do que a taxa de rejeição de ω , $P(\omega \geq x_\alpha)$, configurando um desempenho melhor da estatística corrigida ω^* . Verifica-se, também, que a aproximação de ω por χ^2 piora quando a probabilidade de fracasso θ é reduzida (próxima de zero). Neste caso, a correção de Bartlett torna-se mais necessária. Como deveria ser esperado, para γ e θ fixos, a aproximação χ^2 para ambas estatísticas (ω e ω^*) melhora quando n cresce, o mesmo ocorrendo para n e θ fixos e γ crescendo.

4.6.6 DISTRIBUIÇÃO QUI-QUADRADO

O estimador de máxima verossimilhança para θ é obtido iterativamente da equação

$$\psi\left(\frac{\hat{\theta}}{2}\right) = \log\left(\frac{\bar{y}_g}{2}\right),$$

onde $\bar{y}_g$ é a média geométrica das observações $y_1, y_2, \dots, y_n$, ou seja, $\bar{y}_g = (\prod_{i=1}^n y_i)^{1/n}$. Para valores de $\hat{\theta}$ grandes, pode-se usar a aproximação $\psi(\frac{\theta}{2}) = \log(\frac{\theta-1}{2})$. A estatística da razão de verossimilhança para testar as hipóteses $H : \theta = \theta^{(0)}$ versus $A : \theta \neq \theta^{(0)}$ é dada por

$$\omega = n(\hat{\theta} - \theta^{(0)}) \log\left(\frac{\bar{y}_g}{2}\right) + 2n \log\left(\frac{\Gamma(\theta^{(0)}/2)}{\Gamma(\hat{\theta}/2)}\right).$$

A estatística modificada é $\omega^* = \omega/c$, onde

$$c = c(\theta^{(0)}) = 1 + \frac{5\psi''(\theta^{(0)}/2) - 3\psi'(\theta^{(0)}/2)\psi'''(\theta^{(0)}/2)}{12n\psi'(\theta^{(0)}/2)^3}.$$

O simulador usado aqui baseia-se na distribuição gama. Se $G(\mu, r)$ é a distribuição gama de média $\mu > 0$ e índice $r > 0$ vale a seguinte identidade: $\chi_\theta^2 \stackrel{d}{=} G(1, \frac{\theta}{2})$. Este gerador foi usado por Cordeiro (1995) para simular dados com distribuição gama. A grande vantagem deste simulador é que podemos gerar variáveis qui-quadrado com graus de liberdade sendo números reais positivos. Nosso estudo é, então, conduzido da seguinte forma: Para cada combinação de valores de $\theta^{(0)}$ ($\theta^{(0)} = 0, 1; 0, 2; 0, 5; 1; 2; 5; 10$) e de tamanho de amostra n ($n = 5, 10, 20$)

geramos 10.000 amostras. Para estas amostras calculamos as estatísticas ω e ω^* e a proporção em que estas estatísticas excedem o valor x_α ($\alpha = 5\%$ e 10%), onde x_α é o $(1 - \alpha)$ -ésimo percentil da distribuição χ^2_1 , $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente. Os resultados são apresentados na Tabela 4.13.

Tabela 4.13: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para testar $H : \theta = \theta^{(0)}$ na distribuição χ^2_θ .

Nível nominal α	n	$\theta^{(0)} = 0,1$		$\theta^{(0)} = 0,2$		$\theta^{(0)} = 0,5$		$\theta^{(0)} = 1$		$\theta^{(0)} = 2$		$\theta^{(0)} = 5$		$\theta^{(0)} = 10$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	5,48	5,21	5,35	5,18	5,17	5,08	4,95	4,97	4,83	4,92	4,90	4,95	4,92	4,94
	10	5,26	5,10	5,21	5,13	5,09	5,03	4,96	4,99	4,89	4,97	4,93	4,99	4,97	4,98
	20	5,12	5,06	5,10	5,07	5,05	4,99	4,98	5,01	4,96	5,02	4,96	5,01	4,98	5,00
10%	5	12,06	11,24	11,47	11,05	10,85	10,48	9,34	9,70	9,26	9,61	9,51	9,68	9,63	9,78
	10	11,45	10,81	10,81	10,85	10,58	10,25	9,48	9,74	9,30	9,59	9,69	9,81	9,81	9,89
	20	11,03	10,56	10,66	10,69	10,37	10,10	9,71	9,88	9,45	9,90	9,85	9,94	9,93	9,97

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

Os números na Tabela 4.13 evidenciam alguns pontos: (1) para n fixado, a aproximação de ω por χ^2 é ruim para valores pequenos do parâmetro θ , para $\alpha = 5\%$ ou 10% . Nestes casos, torna-se mais necessária a correção de Bartlett. Verifica-se que para $\theta^{(0)} \geq 1$ a aproximação por χ^2_1 para ω já é adequada e a correção não é necessária; (2) para θ fixado, como esperado, a aproximação por χ^2 , para ambas estatísticas (ω e ω^*), melhora quando n cresce; (3) em algumas situações as taxas de rejeição de ω e ω^* são subestimadas ($\theta \geq 1$) e (4) para todas as situações consideradas o desempenho da estatística ω^* é melhor do que o da ω , já que todas as taxas de rejeição de ω^* estão mais próximas dos níveis nominais, do que as de ω .

4.6.7 DISTRIBUIÇÃO DE POISSON

Para uma amostra aleatória $Y_1, Y_2, \dots, Y_n$ da distribuição de Poisson de média θ , o estimador de máxima verossimilhança para θ é a média amostral $\bar{y}$. A estatística da razão de verossimilhança para testar a hipótese nula $H : \theta = \theta^{(0)}$ é dada por

$$\omega = 2n \left\{ \bar{y} \log \left(\frac{\bar{y}}{\theta^{(0)}} \right) + \theta^{(0)} - \bar{y} \right\} .$$

A estatística ω é uma forma especial do desvio de Poisson estudado, exhaustivamente, através de simulações Monte-Carlo, por Cordeiro (1995).

A idéia da geração da Poisson Y é simples. Dada a taxa de Poisson θ , calcula-se a distribuição acumulada, através de

$$F_Y(y) = P(Y \leq y) = e^{-\theta} \sum_{i=0}^y \frac{\theta^i}{i!} = P(\chi_{2(1+y)}^2 > 2\theta) ,$$

para $y = 0, 1, 2, \dots, N-1, N$, onde N é um número arbitrário suficientemente grande (digamos $N = 50\theta$). Escolhem-se números aleatórios δ_i em $(0,1)$. Para cada δ_i gera-se a variável aleatória Y_i de Poisson, como aquele número em $\{0, 1, \dots, N\}$, tal que $F(y_i) \leq \delta_i < F(y_i + 1)$. A macro do GLIM 3.77, que gera Poisson desta maneira, faz parte da biblioteca de macros da versão GLIM 3.77 para o computador VAX. Simulamos 10.000 amostras para cada uma das variações de $\theta^{(0)} = 0,2; 0,5; 1,0; 1,5$ e $5,0$, $n = 5, 10$ e 20 , $\alpha = 5$ e 10% e calculamos as estatísticas ω e ω^* . A partir dos 10.000 valores simulados de ω e ω^* obtemos as respectivas taxas de rejeição $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, onde x_α é o $(1 - \alpha)$ -ésimo percentil da distribuição χ_1^2 . A Tabela 4.14 fornece os resultados.

Tabela 4.14: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste do parâmetro θ da Poisson.

Nível nominal α	n	$\theta^{(0)} = 0,2$		$\theta^{(0)} = 0,5$		$\theta^{(0)} = 1,0$		$\theta^{(0)} = 1,5$		$\theta^{(0)} = 5,0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5%	5	7,02	5,63	6,75	5,43	5,44	5,36	5,35	5,22	5,26	5,15
	10	6,31	5,35	5,87	5,27	5,36	5,21	5,24	5,15	5,12	5,09
	20	5,78	5,20	5,47	5,10	5,23	5,08	5,13	5,08	5,06	5,02
10%	5	13,27	11,81	12,34	11,41	11,41	10,91	11,02	10,80	10,78	10,64
	10	12,55	10,77	11,87	10,53	10,83	10,47	10,66	10,41	10,45	10,31
	20	11,34	10,68	10,90	10,40	10,45	10,26	10,38	10,21	10,11	10,08

Nota: Os números (1) e (2) correspondem a $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente.

É evidente da Tabela 4.14 o desempenho superior da estatística corrigida ω^* , em relação à estatística não corrigida ω , particularmente, para tamanhos de amostra pequenos ($n = 5$ e 10) e valores pequenos de θ ($\theta = 0,2$ e $0,5$). Nestas situações observa-se, também, as piores aproximações por χ^2 da estatística ω , tornando-se mais necessário o uso da correção de Bartlett, ou seja, da estatística corrigida ω^* . Para θ fixado, à medida que n cresce a correção de Bartlett diminui, mostrando, como esperado, que as duas estatísticas ω e ω^* têm boa aproximação por χ^2_1 .

CAPÍTULO 5

CORREÇÕES DE BARTLETT EM MODELOS BIPARAMÉTRICOS DA FAMÍLIA EXPONENCIAL

5.1 INTRODUÇÃO

No Capítulo 4 desenvolvemos o aperfeiçoamento do teste da razão de verossimilhança para distribuições uniparamétricas na família exponencial. A expressão resultante, nesta situação, é bastante simples e de fácil aplicação, envolvendo somente operações triviais com certas funções e suas derivadas. Nosso objetivo, agora, é ampliar o desenvolvimento anterior para o caso em que temos dois parâmetros, ou seja, desenvolvemos o aperfeiçoamento da estatística da razão de verossimilhança para as distribuições biparamétricas pertencentes à família exponencial, cuja função de probabilidade ou função densidade pode ser escrita por

$$\pi(y; \theta, \phi) = \exp\{\alpha_1(\theta, \phi)d_1(y) + \alpha_2(\theta, \phi)d_2(y) - \rho(\theta, \phi) + v(y)\} , \quad (5.1)$$

onde $(\theta, \phi)^\top$ é o vetor de parâmetros, $\alpha_1(\cdot, \cdot)$, $\alpha_2(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, $d_1(\cdot)$, $d_2(\cdot)$ e $v(\cdot)$ são funções conhecidas e considera-se que o suporte de $\pi(y; \theta, \phi)$ não depende de θ e ϕ . Assumimos, também, que as funções α_1 , α_2 e ρ têm as quatro primeiras derivadas contínuas em relação aos componentes do vetor de parâmetros $(\theta, \phi)^\top$.

Seja $Y_1, \dots, Y_n$ um conjunto de n variáveis aleatórias independentes e identicamente distribuídas com função densidade (ou de probabilidade) dada por (5.1). Denotamos por y_l o valor observado de Y_l e por $y = (y_1, \dots, y_n)^\top$ o vetor das n observações. O logaritmo da função de verossimilhança total, para os parâmetros desconhecidos θ e ϕ , dado y , é

$$\ell = \ell(\theta, \phi) = \sum_{l=1}^n t(y_l; \theta, \phi), \quad (5.2)$$

sendo $t(y_l; \theta, \phi) = \alpha_1 d_1(y_l) + \alpha_2 d_2(y_l) - \rho + v(y_l)$, onde $\alpha_1 = \alpha_1(\theta, \phi)$, $\alpha_2 = \alpha_2(\theta, \phi)$ e $\rho = \rho(\theta, \phi)$. Assumimos que ℓ é regular com respeito às derivadas em relação aos componentes de $(\theta, \phi)^\top$ até quarta ordem. Seja $\alpha_1^{(i,j)} = \partial^{i+j} \alpha_1 / \partial \theta^i \partial \phi^j$, $i, j = 1, 2, 3$. Definição análoga é usada para $\alpha_2^{(i,j)}$ e $\rho^{(i,j)}$. Para a função (5.2), usando as condições de regularidade, valem as seguintes relações:

$$\begin{aligned} \alpha_1^{(1,0)} \beta_1 + \alpha_2^{(1,0)} \beta_2 &= \rho^{(1,0)}, \\ \alpha_1^{(0,1)} \beta_1 + \alpha_2^{(0,1)} \beta_2 &= \rho^{(0,1)}, \end{aligned} \quad (5.3)$$

onde $\beta_r = E\{d_r(y)\}$, $r = 1, 2$. Usando as relações em (5.3), e definindo $\beta_r^{(i,j)}$ de maneira análoga a $\alpha_1^{(i,j)}$, obtemos

$$\begin{aligned} \kappa_{\theta\theta} &= E \left[\frac{\partial^2 \ell}{\partial \theta^2} \right] = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)}, \\ \kappa_{\phi\phi} &= E \left[\frac{\partial^2 \ell}{\partial \phi^2} \right] = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)}, \end{aligned} \quad (5.4)$$

as informações de Fisher para θ e ϕ , respectivamente, e, assim, $\kappa^{\theta\theta} = \kappa_{\theta\theta}^{-1}$ e $\kappa^{\phi\phi} = \kappa_{\phi\phi}^{-1}$. É possível mostrarmos, também, facilmente, que

$$\kappa_{\theta\phi} = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)}, \quad \kappa_{\phi\theta} = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,0)}. \quad (5.5)$$

Podemos resolver o sistema dado pelas equações em (5.3), para β_1 e β_2 , resultando

$$\begin{aligned}\beta_1 &= \frac{\rho^{(0,1)}\alpha_2^{(1,0)} - \rho^{(1,0)}\alpha_2^{(0,1)}}{\alpha_1^{(0,1)}\alpha_2^{(1,0)} - \alpha_1^{(1,0)}\alpha_2^{(0,1)}} , \\ \beta_2 &= \frac{\rho^{(0,1)}\alpha_1^{(1,0)} - \rho^{(1,0)}\alpha_1^{(0,1)}}{\alpha_2^{(0,1)}\alpha_1^{(1,0)} - \alpha_2^{(1,0)}\alpha_1^{(0,1)}} .\end{aligned}\tag{5.6}$$

Este capítulo segue estruturado da seguinte forma: na Seção 5.2 apresentamos testes da razão de verossimilhança aperfeiçoados nos modelos biparamétricos pertencentes à família exponencial, cujos parâmetros são globalmente ortogonais. A expressão da correção de Bartlett resultante é bastante simples, dependendo de derivadas das funções $\alpha_1, \alpha_2, \beta_1$ e β_2 , em relação a θ e a ϕ . Na Seção 5.3, pesquisamos duas situações particulares destes modelos, que produzem simplificações no cálculo da correção. Na Seção 5.4, aplicamos os resultados das seções anteriores em alguns modelos desta classe encontrados na literatura, cujos resultados constituem um cheque parcial das nossas expressões. A construção de parâmetros globalmente ortogonais é apresentada na Seção 5.5, segundo formulação de Cox e Reid (1987), bem como, aplicações desta metodologia em alguns exemplos. Finalizando, na Seção 5.6, apresentamos um pequeno estudo de simulação para o modelo gama discutido na Seção 5.4.3, com o intuito de verificar o desempenho da correção de Bartlett.

5.2 OBTENÇÃO DA CORREÇÃO PARA PARÂMETROS ORTOGONAIS

Sem perda de generalidade, vamos assumir que o vetor de parâmetros é $(\theta, \phi)^\top$, sendo θ o parâmetro de interesse e ϕ o parâmetro de perturbação, ou seja, se ϕ é o parâmetro de interesse basta considerarmos, nas deduções a seguir, $(\phi, \theta)^\top$ como o vetor de parâmetros desconhecidos. Assumimos, ainda, que os parâmetros θ e ϕ são globalmente ortogonais, isto é, $\kappa_{\theta\phi} = \kappa_{\phi\theta} = 0$. Quando esta condição não está satisfeita podemos, através de reparametrização (vide Seção 5.5), obter parâmetros ortogonais e aplicar nossos resultados.

Nosso objetivo é, então, testar a hipótese nula $H : \theta = \theta^{(0)}$ contra a alternativa $A : \theta \neq \theta^{(0)}$, onde $\theta^{(0)}$ é um valor especificado e ϕ é o parâmetro de incômodo. Denotamos por $(\hat{\theta}, \hat{\phi})^\top$ o estimador de máxima verossimilhança irrestrito de $(\theta, \phi)^\top$ e por $\tilde{\phi}$ o estimador de máxima verossimilhança de ϕ , sob H . A estatística da razão de verossimilhança para o teste de H é dada por

$$\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta^{(0)}, \tilde{\phi})] . \quad (5.7)$$

Assintoticamente, ω tem distribuição χ_1^2 , sob a hipótese nula H . Reescrevendo a estatística ω em (5.7) como

$$\omega = 2\{[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta^{(0)}, \phi)] - [\ell(\theta^{(0)}, \tilde{\phi}) - \ell(\theta^{(0)}, \phi)]\} , \quad (5.8)$$

sendo $\ell(\theta^{(0)}, \phi)$ o logaritmo da função de verossimilhança avaliada nos parâmetros verdadeiros sob H , temos, segundo Lawley (1956), que

$$2E\{\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta^{(0)}, \phi)\} = 2 + \varepsilon_2 + O(n^{-2}) , \quad (5.9)$$

onde o termo ε_2 , de ordem n^{-1} , é dado por

$$\varepsilon_2 = \sum_{\theta, \phi} '(\ell_{rstu} - \ell_{rstuvw}) , \quad (5.10)$$

com cada um dos índices r, s, t, u, v e w sendo igual a θ ou ϕ , $\ell_{rstu} = \kappa^{rs} \kappa^{tu} (\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rs}^{(tu)})$, $\ell_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} [\kappa_{rtv} (\frac{1}{6} \kappa_{suw} - \kappa_{sw}^{(u)}) + \kappa_{rtu} (\frac{1}{4} \kappa_{svw} - \kappa_{sw}^{(v)}) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)}]$ e Σ' denotando o somatório sobre θ e ϕ . Como no Capítulo 2, a quantidade ε_2 em (5.10) pode ser decomposta em

$$\varepsilon_2 = A_\theta + B_\phi + C_{\theta, \phi} , \quad (5.11)$$

sendo A_θ a parcela do somatório que envolve apenas θ , B_ϕ a parcela do somatório que envolve apenas ϕ e $C_{\theta, \phi}$ a parcela mista do somatório, ou seja, envolvendo simultaneamente os parâmetros θ e ϕ . Assim, $A_\theta = \ell_{\theta\theta\theta\theta} - \ell_{\theta\theta\theta\theta\theta}$, $B_\phi = \ell_{\phi\phi\phi\phi} - \ell_{\phi\phi\phi\phi\phi}$ e $C_{\theta, \phi} = \sum'_{\theta, \phi} (\ell_{rstu} - \ell_{rstuvw})$.

Lawley mostrou, também, que

$$2E\{\ell(\theta^{(0)}, \tilde{\phi}) - \ell(\theta^{(0)}, \phi)\} = 1 + \varepsilon_1 + O(n^{-2}) , \quad (5.12)$$

com ε_1 obtido a partir de (5.10) observando, agora, que o somatório Σ' daquela expressão restringe-se apenas ao parâmetro ϕ , ou seja,

$$\varepsilon_1 = \ell_{\phi\phi\phi\phi} - \ell_{\phi\phi\phi\phi\phi} = B_\phi . \quad (5.13)$$

De (5.9) e (5.12) segue que, sob a hipótese nula H , o valor esperado da estatística da razão de verossimilhança ω em (5.7) tem a forma

$$E(\omega) = 1 + \varepsilon_2 - \varepsilon_1 + O(n^{-2}) = 1 + A_\theta + C_{\theta,\phi} + O(n^{-2}) . \quad (5.14)$$

Podemos melhorar a aproximação da estatística ω pela distribuição χ_1^2 substituindo ω pela estatística $\omega^* = c^{-1}\omega$, onde o fator de correção de Bartlett c é determinado pela relação $c = E(\omega)$, sendo que $E(\omega)$ é calculado até ordem n^{-1} . Assim,

$$c = 1 + A_\theta + C_{\theta,\phi} , \quad (5.15)$$

e a estatística corrigida $\omega^* = c^{-1}\omega$ tem distribuição χ_1^2 até ordem n^{-1} , sob H .

O cálculo das quantidades A_θ e $C_{\theta,\phi}$, em (5.15), envolve o cálculo de cumulantes mistos até quarta ordem, que por ser bastante trabalhoso, é apresentado no Apêndice 4. Devido à ortogonalidade global entre θ e ϕ alguns cumulantes mistos se anulam, produzindo algumas simplificações na obtenção de A_θ e $C_{\theta,\phi}$. Usando as definições de ℓ_{rstu} e ℓ_{rstuvw} e substituindo os cumulantes envolvidos em seus cálculos, encontramos

$$\begin{aligned} A_\theta = & \frac{1}{4n \left(\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \right)^2} \sum_{r=1}^2 (\alpha_r^{(3,0)} \beta_r^{(1,0)} + \alpha_r^{(2,0)} \beta_r^{(2,0)} - \alpha_r^{(1,0)} \beta_r^{(3,0)}) \\ & - \frac{1}{3n \left(\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \right)^3} \sum_{r=1}^2 \sum_{s=1}^2 (a_r a_s + a_r b_s - \frac{5}{4} b_r b_s) , \end{aligned} \quad (5.16)$$

onde $a_r = \alpha_r^{(2,0)} \beta_r^{(1,0)}$, $b_r = \alpha_r^{(1,0)} \beta_r^{(2,0)}$, $r = 1, 2$, e

$$C_{\theta,\phi} = C_{1\theta,\phi} - C_{2\theta,\phi} , \quad (5.17)$$

sendo

$$C_{1_{\theta,\phi}} = \sum_{\theta,\phi} 'l_{rstu} = \frac{1}{2n \left(\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \right) \left(\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)} \right)} \sum_{r=1}^2 (\alpha_r^{(1,2)} \beta_r^{(1,0)} + \alpha_r^{(2,1)} \beta_r^{(0,1)} + 3\alpha_r^{(1,1)} \beta_r^{(1,1)}) \quad (5.17')$$

e

$$\begin{aligned} C_{2_{\theta,\phi}} = \sum_{\theta,\phi} 'l_{rstuvw} = & \frac{1}{4n \left(\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \right) \left(\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)} \right)} \left\{ \frac{1}{\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)}} \right. \\ & \left[\left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(1,0)} \right)^2 + 4 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(1,0)} \right) \left(\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,1)} \right) \right. \\ & + 2 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(0,1)} \right) \left(\sum_{r=1}^2 (2\alpha_r^{(2,0)} \beta_r^{(1,0)} + \alpha_r^{(1,0)} \beta_r^{(2,0)}) \right) \\ & - 4 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(0,1)} \right) \left(\sum_{r=1}^2 \alpha_r^{(2,0)} \beta_r^{(1,0)} \right) \Big] \\ & + \frac{1}{\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)}} \left[\left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(0,1)} \right)^2 \right. \\ & + 4 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(0,1)} \right) \left(\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,1)} \right) \\ & + 2 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(1,0)} \right) \left(\sum_{r=1}^2 (2\alpha_r^{(0,2)} \beta_r^{(0,1)} + \alpha_r^{(0,1)} \beta_r^{(0,2)}) \right) \\ & \left. \left. - 4 \left(\sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(1,0)} \right) \left(\sum_{r=1}^2 \alpha_r^{(0,2)} \beta_r^{(0,1)} \right) \right] \right\}. \quad (5.17'') \end{aligned}$$

Observando as expressões (5.16) - (5.17) verificamos que, para o teste da hipótese nula $H : \theta = \theta^{(0)}$, a correção da estatística da razão de verossimilhança depende apenas das três primeiras derivadas de α_1 , α_2 , β_1 e β_2 em relação a θ e das duas primeiras derivadas de α_1 , α_2 , β_1 e β_2 em relação a ϕ e de derivadas mistas das funções α_1 e α_2 até terceira ordem e β_1 e β_2 até segunda ordem, em relação aos dois parâmetros θ e ϕ . Conclusão análoga pode ser tirada para testar a hipótese nula $H : \phi = \phi^{(0)}$, sendo θ o parâmetro de incômodo, apenas trocando θ por ϕ e vice-versa.

A grande vantagem da correção (5.15) é que esta só depende de derivadas das funções α_1 , α_2 , β_1 e β_2 , dispensando, então, o cálculo trabalhoso de cumulantes de derivadas do logaritmo da função de verossimilhança.

5.3 DOIS CASOS ESPECIAIS

Investigamos, agora, duas situações particulares de modelos biparamétricos pertencentes à família exponencial, que produzem simplificações no cálculo da correção de Bartlett e que são de interesse prático.

Caso 1: $\alpha_1(\theta, \phi) = c_1(\theta)$ e $\alpha_2(\theta, \phi) = c_2(\phi)$.

Consideramos, primeiramente, que nosso interesse é o teste de hipótese sobre o parâmetro θ , sendo ϕ o parâmetro de perturbação. Temos que a função α_1 depende somente do parâmetro θ e a função α_2 somente do parâmetro ϕ . Para esta situação vale,

$$\begin{aligned}\alpha_1^{(0,j)} &= \partial^j \alpha_1(\theta, \phi) / \partial \phi^j = 0, \quad \alpha_2^{(i,0)} = \partial^i \alpha_2(\theta, \phi) / \partial \theta^i = 0, \text{ e} \\ \alpha_r^{(i,j)} &= \partial^{i+j} \alpha_r(\theta, \phi) / \partial \theta^i \partial \phi^j = 0, \text{ para } r = 1, 2 \text{ e } i, j = 1, 2, \dots\end{aligned}$$

Assim,

$$\beta_1 = \frac{\rho^{(1,0)}}{\alpha_1^{(1,0)}} \quad \text{e} \quad \beta_2 = \frac{\rho^{(0,1)}}{\alpha_2^{(0,1)}}. \quad (5.18)$$

Da ortogonalidade global entre θ e ϕ não é difícil mostrarmos que $\rho^{(i,j)} = 0$, resultando $\beta_1^{(0,j)} = \beta_2^{(i,0)} = 0$ e $\beta_r^{(i,j)} = 0$, para $r = 1, 2$ e $i, j = 1, 2, 3, \dots$. Logo, as quantidades A_θ e $C_{\theta,\phi}$, que determinam o fator de correção c da estatística da razão de verossimilhança em (5.15) - (5.17), são dadas por

$$\begin{aligned}A_\theta &= \frac{1}{4n(\alpha_1^{(1,0)}\beta_1^{(1,0)})^2} [\alpha_1^{(3,0)}\beta_1^{(1,0)} + \alpha_1^{(2,0)}\beta_1^{(2,0)} - \alpha_1^{(1,0)}\beta_1^{(3,0)}] \\ &\quad - \frac{1}{3n(\alpha_1^{(1,0)}\beta_1^{(1,0)})^3} \left[(\alpha_1^{(2,0)}\beta_1^{(1,0)})^2 - \frac{5}{4} (\alpha_1^{(1,0)}\beta_1^{(2,0)})^2 + \alpha_1^{(2,0)}\beta_1^{(1,0)}\alpha_1^{(1,0)}\beta_1^{(2,0)} \right],\end{aligned}$$

$$C_{\theta,\phi} = 0. \quad (5.19)$$

Destes últimos resultados é possível concluirmos que, nos modelos exponenciais biparamétricos em que $\alpha_1(\theta, \phi) = c_1(\theta)$ e $\alpha_2(\theta, \phi) = c_2(\phi)$, a correção da estatística da razão de verossimilhança ω em (5.7), para o teste de hipótese sobre o parâmetro θ , depende somente das três primeiras derivadas de α_1 e β_1 em relação a θ , sendo dada por $c = 1 + A_\theta$, com A_θ em (5.19).

Se o interesse é testar o parâmetro ϕ , sendo θ o parâmetro de incômodo, consideramos $(\phi, \theta)^\top$ como o vetor paramétrico e, assim, temos

$$\begin{aligned} \alpha_1^{(i,0)} &= \partial^i \alpha_1(\theta, \phi) / \partial \phi^i = 0, \quad \alpha_2^{(0,j)} = \partial^j \alpha_2(\theta, \phi) / \partial \theta^j = 0 \quad \text{e} \\ \alpha_r^{(i,j)} &= \partial^{i+j} \alpha_r(\theta, \phi) / \partial \phi^i \partial \theta^j = 0, \quad \text{para } r = 1, 2 \text{ e } i, j = 1, 2, 3, \dots \end{aligned}$$

resultando

$$\beta_1 = \frac{\rho^{(0,1)}}{\alpha_1^{(0,1)}} \quad \text{e} \quad \beta_2 = \frac{\rho^{(1,0)}}{\alpha_2^{(1,0)}}.$$

Estes resultados para β_1 e β_2 são os mesmos daqueles em (5.18), já que lá o vetor de parâmetros é $(\theta, \phi)^\top$. É fácil mostrarmos que, devido à ortogonalidade global entre θ e ϕ , $\rho^{(i,j)} = 0$, produzindo $\beta_1^{(i,0)} = \beta_2^{(0,j)} = 0$ e $\beta_r^{(i,j)} = 0$, para $r = 1, 2$ e $i, j = 1, 2, 3, \dots$. Assim, de (5.16) - (5.17), temos

$$\begin{aligned} A_\phi &= \frac{1}{4n \left(\alpha_2^{(1,0)} \beta_2^{(1,0)} \right)^2} \left[\alpha_2^{(3,0)} \beta_2^{(1,0)} + \alpha_2^{(2,0)} \beta_2^{(2,0)} - \alpha_2^{(1,0)} \beta_2^{(3,0)} \right] \\ &\quad - \frac{1}{3n \left(\alpha_2^{(1,0)} \beta_2^{(1,0)} \right)^3} \left[\left(\alpha_2^{(2,0)} \beta_2^{(1,0)} \right)^2 - \frac{5}{4} \left(\alpha_2^{(1,0)} \beta_2^{(2,0)} \right)^2 + \alpha_2^{(2,0)} \beta_2^{(1,0)} \alpha_2^{(1,0)} \beta_2^{(2,0)} \right] \quad (5.20) \end{aligned}$$

e

$$C_{\theta,\phi} = 0.$$

Logo, $c = 1 + A_\phi$. Notamos, agora, que para o teste do parâmetro ϕ a correção depende, apenas, das três primeiras derivadas de α_2 e β_2 em relação a ϕ .

Caso 2: $\alpha_1(\theta, \phi) = c_1(\theta, \phi)$ e $\alpha_2(\theta, \phi) = c_2(\phi)$.

Neste caso, temos a função α_2 dependente somente do parâmetro ϕ , que implica $\alpha_2^{(i,0)} = 0 = \alpha_2^{(i,j)}$, para $i, j = 1, 2, \dots$ resultando

$$\beta_1 = \frac{\rho^{(1,0)}}{\alpha_1^{(1,0)}} , \quad \beta_2 = \frac{\rho^{(0,1)}\alpha_1^{(1,0)} - \rho^{(1,0)}\alpha_1^{(0,1)}}{\alpha_2^{(0,1)}\alpha_1^{(1,0)}} . \quad (5.21)$$

Como θ e ϕ são globalmente ortogonais, temos $\kappa_{\theta\phi} = -n \sum_{r=1}^2 \alpha_r^{(1,0)}\beta_r^{(0,1)} = -n\alpha_1^{(1,0)}\beta_1^{(0,1)} = 0$ e $\kappa_{\phi\theta} = -n \left(\alpha_1^{(0,1)}\beta_1^{(1,0)} + \alpha_2^{(0,1)}\beta_2^{(1,0)} \right) = 0$. Logo, para testes de hipóteses sobre o parâmetro θ , sendo ϕ o parâmetro de incômodo, as quantidades A_θ e $C_{\theta,\phi}$ em (5.15) - (5.17), respectivamente, que determinam o fator de correção c da estatística da razão de verossimilhança, são calculadas por

$$A_\theta = \frac{1}{4n \left(\alpha_1^{(1,0)}\beta_1^{(1,0)} \right)^2} \left[\alpha_1^{(3,0)}\beta_1^{(1,0)} + \alpha_1^{(2,0)}\beta_1^{(2,0)} - \alpha_1^{(1,0)}\beta_1^{(3,0)} \right] \\ - \frac{1}{3n \left(\alpha_1^{(1,0)}\beta_1^{(1,0)} \right)^3} \left[\left(\alpha_1^{(2,0)}\beta_1^{(1,0)} \right)^2 - \frac{5}{4} \left(\alpha_1^{(1,0)}\beta_1^{(2,0)} \right)^2 + \alpha_1^{(2,0)}\beta_1^{(1,0)}\alpha_1^{(1,0)}\beta_1^{(2,0)} \right], \quad (5.22)$$

$$C_{\theta,\phi} = C_{1_{\theta,\phi}} - C_{2_{\theta,\phi}} , \quad (5.23)$$

com

$$C_{1_{\theta,\phi}} = \frac{1}{2n\alpha_1^{(1,0)}\beta_1^{(1,0)} \sum_{r=1}^2 \alpha_r^{(0,1)}\beta_r^{(0,1)}} \left[\alpha_1^{(1,2)}\beta_1^{(1,0)} + \alpha_1^{(2,1)}\beta_1^{(0,1)} + 3\alpha_1^{(1,1)}\beta_1^{(1,1)} \right], \quad (5.23')$$

$$C_{2_{\theta,\phi}} = \frac{1}{4n\alpha_1^{(1,0)}\beta_1^{(1,0)} \sum_{r=1}^2 \alpha_r^{(0,1)}\beta_r^{(0,1)}} \left\{ \frac{1}{\alpha_1^{(1,0)}\beta_1^{(1,0)}} \left[\left(\alpha_1^{(1,1)}\beta_1^{(1,0)} \right)^2 + 4\alpha_1^{(1,1)}\beta_1^{(1,0)}\alpha_1^{(1,0)}\beta_1^{(1,1)} \right. \right. \\ \left. \left. + 2\alpha_1^{(1,1)}\beta_1^{(0,1)} \left(2\alpha_1^{(2,0)}\beta_1^{(1,0)} + \alpha_1^{(1,0)}\beta_1^{(2,0)} \right) - 4\alpha_1^{(1,1)}\beta_1^{(0,1)}\alpha_1^{(2,0)}\beta_1^{(1,0)} \right] \right. \\ \left. + \frac{1}{\sum_{r=1}^2 \alpha_r^{(0,1)}\beta_r^{(0,1)}} \left[\left(\alpha_1^{(1,1)}\beta_1^{(0,1)} \right)^2 + 4\alpha_1^{(1,1)}\beta_1^{(0,1)} \sum_{r=1}^2 \alpha_r^{(0,1)}\beta_r^{(1,1)} \right. \right. \\ \left. \left. + 2\alpha_1^{(1,1)}\beta_1^{(1,0)} \sum_{r=1}^2 \left(2\alpha_r^{(0,2)}\beta_r^{(0,1)} + \alpha_r^{(0,1)}\beta_r^{(0,2)} \right) - 4\alpha_1^{(1,1)}\beta_1^{(1,0)} \sum_{r=1}^2 \alpha_r^{(0,2)}\beta_r^{(0,1)} \right] \right\} \quad (5.23'')$$

Se estamos interessados em testar o parâmetro ϕ , sendo θ o parâmetro de perturbação, devemos considerar $(\phi, \theta)^\top$ como o vetor paramétrico, resultando $\alpha_2^{(0,j)} = 0 = \alpha_2^{(i,j)}$, para $i, j = 1, 2, 3, \dots$. Assim, temos

$$\beta_1 = \frac{\rho^{(0,1)}}{\alpha_1^{(0,1)}} \quad \text{e} \quad \beta_2 = \frac{\rho^{(1,0)}\alpha_1^{(0,1)} - \rho^{(0,1)}\alpha_1^{(1,0)}}{\alpha_2^{(1,0)}\alpha_1^{(0,1)}}. \quad (5.24)$$

As expressões resultantes para β_1 e β_2 aqui são as mesmas daquelas calculadas por (5.21), sendo que nesta última o vetor de parâmetros é $(\theta, \phi)^\top$. A correção da estatística da razão de verossimilhança é $c = 1 + A_\phi + C_{\theta, \phi}$, onde A_ϕ é obtida por (5.16) e $C_{\theta, \phi}$ por (5.17), sendo que, em ambas expressões, as derivadas envolvidas são em relação ao vetor $(\phi, \theta)^\top$. Nesta situação temos,

$$\begin{aligned} C_{1_{\theta, \phi}} &= \frac{1}{2n\alpha_1^{(0,1)}\beta_1^{(0,1)}\sum_{r=1}^2\alpha_r^{(1,0)}\beta_r^{(1,0)}} \left[\alpha_1^{(1,2)}\beta_1^{(1,0)} + \alpha_1^{(2,1)}\beta_1^{(1,0)} + 3\alpha_1^{(1,1)}\beta_1^{(1,1)} \right], \\ C_{2_{\theta, \phi}} &= \frac{1}{4n\alpha_1^{(0,1)}\beta_1^{(0,1)}\sum_{r=1}^2\alpha_r^{(1,0)}\beta_r^{(1,0)}} \left\{ \frac{1}{\sum_{r=1}^2\alpha_r^{(1,0)}\beta_r^{(1,0)}} \left[\left(\alpha_1^{(1,1)}\beta_1^{(1,0)} \right)^2 \right. \right. \\ &\quad + 4\alpha_1^{(1,1)}\beta_1^{(1,0)}\sum_{r=1}^2\alpha_r^{(1,0)}\beta_r^{(1,1)} + 2\alpha_1^{(1,1)}\beta_1^{(0,1)}\sum_{r=1}^2 \left(2\alpha_r^{(2,0)}\beta_r^{(1,0)} + \alpha_r^{(1,0)}\beta_r^{(2,0)} \right) \\ &\quad \left. - 4\alpha_1^{(1,1)}\beta_1^{(0,1)}\sum_{r=1}^2\alpha_r^{(2,0)}\beta_r^{(1,0)} \right] + \frac{1}{\alpha_1^{(0,1)}\beta_1^{(0,1)}} \left[\left(\alpha_1^{(1,1)}\beta_1^{(0,1)} \right)^2 \right. \\ &\quad + 4\alpha_1^{(1,1)}\beta_1^{(0,1)}\alpha_1^{(0,1)}\beta_1^{(1,1)} + 2\alpha_1^{(1,1)}\beta_1^{(1,0)} \left(2\alpha_1^{(0,2)}\beta_1^{(0,1)} + \alpha_1^{(0,1)}\beta_1^{(0,2)} \right) \\ &\quad \left. \left. - 4\alpha_1^{(1,1)}\beta_1^{(1,0)}\alpha_1^{(0,2)}\beta_1^{(0,1)} \right] \right\}. \quad (5.25) \end{aligned}$$

5.4 APLICAÇÕES

Apresentamos, nesta seção, alguns modelos encontrados na literatura, que pertencem à família exponencial biparamétrica, cujos parâmetros são ortogonais. Obtemos para estes modelos o valor da correção da estatística da razão de verossimilhança para testes de hipóteses sobre os parâmetros, baseado nos resultados apresentados nas duas seções anteriores.

5.4.1 DISTRIBUIÇÃO NORMAL $N(\theta, \phi)$

Para a distribuição normal podemos adotar,

$$\alpha_1(\theta, \phi) = \frac{\theta}{\phi}, \quad \alpha_2(\theta, \phi) = -\frac{1}{2\phi}, \quad d_1(y) = y, \quad d_2(y) = y^2,$$

$$\rho(\theta, \phi) = \frac{\theta^2}{2\phi} + \frac{1}{2} \log \phi \quad \text{e} \quad v(y) = -\frac{1}{2} \log 2\pi.$$

Como é bem conhecido, os estimadores de máxima verossimilhança para θ e ϕ são, respectivamente,

$$\hat{\theta} = \bar{y} \quad \text{e} \quad \hat{\phi} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (5.26)$$

Para o teste da hipótese $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$, a estatística da razão de verossimilhança é calculada por

$$\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta^{(0)}, \tilde{\phi})] = n \log \left\{ \frac{\sum_{i=1}^n (y_i - \theta^{(0)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right\}, \quad (5.27)$$

sendo $\tilde{\phi} = \sum_{i=1}^n (y_i - \theta^{(0)})^2 / n$. O fator c que corrige a estatística ω , produzindo a estatística melhorada $\omega^* = c^{-1}\omega$, pode ser calculado usando os resultados da Seção 5.3 (caso 2), em que a função α_2 depende apenas do parâmetro de perturbação ϕ . As derivadas envolvidas neste cálculo estão apresentadas no Apêndice 4. Por (5.21) temos

$$\beta_1 = \theta \quad \text{e} \quad \beta_2 = \theta^2 + \phi$$

e $c = 1 + A_\theta + C_{\theta, \phi}$, onde A_θ e $C_{\theta, \phi}$ são obtidas, respectivamente, de (5.22) e (5.23), resultando $A_\theta = 0$, $C_{\theta, \phi} = C_{1_{\theta, \phi}} - C_{2_{\theta, \phi}} = 3/2n$, já que $C_{1_{\theta, \phi}} = 2/n$ e $C_{2_{\theta, \phi}} = 1/2n$. Logo,

$$c = 1 + \frac{3}{2n},$$

conforme já sabemos (vide Seção 2.4.1 (i)).

Se o interesse é testar $H : \phi = \phi^{(0)}$ contra $A : \phi \neq \phi^{(0)}$, sendo $\phi^{(0)}$ um valor pré-fixado, a estatística da razão de verossimilhança é calculada por

$$\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\tilde{\theta}, \phi^{(0)})] = -n \left[\log \left(\frac{\hat{\phi}}{\phi^{(0)}} \right) - \frac{\hat{\phi}}{\phi^{(0)}} + 1 \right]. \quad (5.28)$$

A correção da estatística ω em (5.28) é dada por $c = 1 + A_\phi + C_{\theta,\phi}$, onde A_ϕ é obtida de (5.16) e $C_{\theta,\phi}$ de (5.17), sendo que as derivadas envolvidas nestas expressões são em relação ao vetor de parâmetros $(\phi, \theta)^\top$. Então, temos $A_\phi = 1/3n$ e $C_{\theta,\phi} = 3/2n$, resultando

$$c = 1 + \frac{11}{6n} ,$$

conforme esperado (vide Seção 2.4.1 (ii)).

5.4.2 DISTRIBUIÇÃO NORMAL INVERSA $NI(\theta, \phi)$

Para esta distribuição de média $\theta > 0$ e parâmetro de precisão $\phi > 0$, podemos escrever $\alpha_1(\theta, \phi) = -\phi/2\theta^2$, $\alpha_2(\theta, \phi) = -\phi/2$, $\rho(\theta, \phi) = -(\phi/\theta + \log(\phi)/2)$, $d_1(y) = y$, $d_2(y) = 1/y$ e $v(y) = -\frac{1}{2} \log(2\pi y^3)$. Utilizando (5.21) ou (5.24) encontramos $\beta_1 = \theta$ e $\beta_2 = \theta^{-1} + \phi^{-1}$. Como a função α_2 depende apenas de ϕ , podemos aplicar os resultados da Seção 5.3 para obtermos a correção da estatística da razão de verossimilhança.

As estimativas de máxima verossimilhança para θ e ϕ são dadas, respectivamente, por

$$\hat{\theta} = \bar{y} \quad \text{e} \quad \hat{\phi} = \left(\frac{1}{\bar{y}_h} - \frac{1}{\bar{y}} \right)^{-1} , \quad (5.29)$$

sendo $\bar{y}_h$ a média harmônica de $y_1, \dots, y_n$, ou seja, $\bar{y}_h^{-1} = \frac{1}{n} \sum_{i=1}^n y_i^{-1}$. Para testar as hipóteses $H : \theta = \theta^{(0)}$ versus $A : \theta \neq \theta^{(0)}$, a estatística da razão de verossimilhança é calculada por

$$\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta^{(0)}, \tilde{\phi})] , \quad (5.30)$$

onde $\tilde{\phi}$ é o estimador de máxima verossimilhança de ϕ sob a hipótese nula. Não é difícil provarmos que $\tilde{\phi} = \left[\frac{1}{n\theta^{(0)2}} \sum_{i=1}^n \frac{(y_i - \theta^{(0)})^2}{y_i} \right]^{-1}$. Assim, (5.30) reduz-se a $\omega = n \log(\hat{\phi}/\tilde{\phi})$. Utilizando (5.22) e (5.23) encontramos $A_\theta = 0 = C_{1,\theta,\phi}$ e $C_{2,\theta,\phi} = -3/2n$, resultando $c = 1 + 3/2n$, conforme já esperado, e $\omega^* = c^{-1}\omega$ tem distribuição assintótica χ_1^2 até ordem n^{-1} .

Se temos interesse em testar a hipótese nula $H : \phi = \phi^{(0)}$ contra a alternativa bilateral, a

estatística da razão de verossimilhança é obtida de

$$\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\tilde{\theta}, \phi^{(0)})] , \quad (5.31)$$

onde $\tilde{\theta}$ é o estimador de máxima verossimilhança para θ sob H , sendo $\tilde{\theta} = \hat{\theta} = \bar{y}$. Escrevemos, então,

$$\omega = n \log \left(\frac{\hat{\phi}}{\phi^{(0)}} \right) - \frac{(\hat{\phi} - \phi^{(0)})}{\bar{y}^2} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{y_i} . \quad (5.32)$$

A obtenção das quantidades envolvidas no cálculo da correção da estatística ω é conseguida através das expressões (5.16) e (5.25), com derivadas tomadas em relação ao vetor $(\phi, \theta)^\top$, fornecendo $A_\phi = 1/3n$, $C_{1\theta, \phi} = 0$ e $C_{2\theta, \phi} = -3/2n$, que resulta $c = 1 + 11/6n$, conforme esperado.

5.4.3 DISTRIBUIÇÃO GAMA $G(\theta, \phi)$

Consideramos a distribuição gama com parametrização tal que a sua média é $\theta > 0$ e a sua variância é $\frac{\theta^2}{\phi}$. Esta parametrização faz com que os parâmetros θ e ϕ sejam ortogonais. A sua função densidade, neste caso, pode ser expressa por

$$\pi(y; \theta, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi}{\theta} \right)^\phi y^{\phi-1} \exp \left(-\frac{\phi}{\theta} y \right) , \quad y > 0, \quad \phi > 0 \text{ e } \theta > 0.$$

Podemos adotar $\alpha_1(\theta, \phi) = -\phi/\theta$, $\alpha_2(\theta, \phi) = \phi$, $d_1(y) = y$, $d_2(y) = \log(y)$, $\rho(\theta, \phi) = \phi \log \theta - \phi \log \phi + \log \Gamma(\phi)$ e $v(y) = -\log y$. Através da expressão (5.21) obtemos $\beta_1 = \theta$ e $\beta_2 = \log \theta + \psi(\phi) - \log \phi$, sendo $\psi(\cdot)$ a função digama, isto é, $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$.

O estimador de máxima verossimilhança para θ é $\hat{\theta} = \bar{y}$ e para ϕ , pode ser obtido resolvendo-se a seguinte equação:

$$n^{-1} \sum_{i=1}^n \log y_i - \log \bar{y} = \psi(\hat{\phi}) - \log \hat{\phi} ,$$

que pode ser reescrita como

$$\log \hat{\phi} - \psi(\hat{\phi}) = \log \left(\frac{\bar{y}}{(\prod_{i=1}^n y_i)^{1/n}} \right) . \quad (5.33)$$

Note-se que $\bar{y}_g = (\prod_{i=1}^n y_i)^{1/n}$ é a média geométrica dos dados $y_1, \dots, y_n$. Seja $g = \log(\bar{y}/\bar{y}_g)$ o logaritmo da razão entre a média aritmética e a média geométrica das observações $y_1, \dots, y_n$. Segundo Johnson e Kotz (1970a), o valor de $\hat{\phi}$ pode ser determinado por interpolação inversa numa tabela da função $\{\log \phi - \psi(\phi)\}$, a partir do valor de g . Uma tabela destas é fornecida por Masuyama e Kuroiwa (1952). Chapman (1956) apresenta uma tabela com alguns resultados desta interpolação inversa, isto é, valores de $\hat{\phi}$, correspondentes a uns poucos valores da razão g .

Para testar a hipótese nula $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$, a estatística da razão de verossimilhança é calculada por (5.30), onde $\tilde{\phi}$ é o estimador de máxima verossimilhança para ϕ , sob a hipótese nula H , ou seja, $\tilde{\phi}$ é solução da equação não-linear

$$\log \tilde{\phi} - \psi(\tilde{\phi}) = -\log \left(\frac{\bar{y}_g}{\theta^{(0)}} \right) + \frac{\bar{y}}{\theta^{(0)}} - 1. \quad (5.34)$$

A equação (5.34) pode ser resolvida para $\tilde{\phi}$ de forma análoga à solução $\hat{\phi}$ de (5.33). Observar que (5.34) reduz-se a (5.33) quando $\theta^{(0)}$ é substituído por $\bar{y}$. Assim, encontramos

$$\omega = 2n \left\{ (\hat{\phi} - \tilde{\phi}) \log \bar{y}_g - \left(\hat{\phi} - \frac{\tilde{\phi} \bar{y}}{\theta^{(0)}} \right) + \log \left[\left(\frac{\hat{\phi}}{\bar{y}} \right)^{\hat{\phi}} \frac{1}{\Gamma(\hat{\phi})} \right] - \log \left[\left(\frac{\tilde{\phi}}{\theta^{(0)}} \right)^{\tilde{\phi}} \frac{1}{\Gamma(\tilde{\phi})} \right] \right\}. \quad (5.35)$$

Como antes, a correção da estatística ω é dada pela equação $c = 1 + A_\theta + C_{\theta, \phi}$, com A_θ e $C_{\theta, \phi}$ dados por (5.22) e (5.23), respectivamente, já que α_2 depende apenas de ϕ . Substituindo as derivadas envolvidas nestas expressões (vide Apêndice 4), encontramos $A_\theta = 1/6n\phi$ e $C_{\theta, \phi} = C_{1, \theta, \phi} - C_{2, \theta, \phi}$, com $C_{1, \theta, \phi} = 0$ e $C_{2, \theta, \phi} = \frac{1}{4n\phi(\phi\psi' - 1)^2}(2\phi^2\psi'' + \phi\psi' + 1)$. Logo,

$$\begin{aligned} c &= 1 + \frac{1}{6n\phi} - \frac{1}{4n\phi(\phi\psi' - 1)^2}(2\phi^2\psi'' + \phi\psi' + 1) \\ &= 1 + \frac{1}{12n\phi} \left[\frac{-3(2\phi^2\psi'' + \phi\psi' + 1)}{(\phi\psi' - 1)^2} + 2 \right], \end{aligned} \quad (5.36)$$

onde ψ' , ψ'' e ψ''' são as três primeiras derivadas da função $\psi(\cdot)$. Como c depende de ϕ que é desconhecido, devemos avaliar c em $\tilde{\phi}$, a estimativa de máxima verossimilhança de ϕ sob H , que é a solução da equação (5.34); isto não afeta a ordem de aproximação de $E(\omega)$, ou seja, $\tilde{\omega}^* = \tilde{c}^{-1}\omega$ tem distribuição χ_1^2 até ordem n^{-1} , com $\tilde{c} = c(\tilde{\phi})$.

Se desejamos testar $H : \phi = \phi^{(0)}$ contra $A : \phi \neq \phi^{(0)}$, a estatística da razão de verossimilhança ω é dada pela expressão em (5.31), sendo $\tilde{\theta} = \hat{\theta} = \bar{y}$ a estimativa de máxima verossimilhança de θ segundo H e $\hat{\phi}$ a solução da equação não-linear em (5.33). Substituindo-se as estimativas em (5.31), vem

$$\omega = 2n \left\{ (\hat{\phi} - \phi^{(0)}) [\log(\bar{y}_g/\bar{y}) - 1] + \log \left(\frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \right) - \log \left(\frac{\phi^{(0)\phi^{(0)}}}{\Gamma(\phi^{(0)})} \right) \right\}, \quad (5.37)$$

e a estatística aperfeiçoada $\omega^* = c^{-1}\omega$ tem distribuição assintótica χ_1^2 até ordem n^{-1} sob H , sendo $c = 1 + A_\phi + C_{\theta,\phi}$. As quantidades A_ϕ e $C_{\theta,\phi}$ são obtidas, respectivamente, através de (5.16) e (5.25) resultando

$$\begin{aligned} A_\phi &= \frac{1}{12n\phi} \left[\frac{5(\phi^2\psi'' + 1)^2}{(\phi\psi' - 1)^3} - \frac{3(\phi^3\psi'' - 2)}{(\phi\psi' - 1)^2} \right], \\ C_{1\theta,\phi} &= 0, \\ C_{2\theta,\phi} &= \frac{1}{4n\phi(\phi\psi' - 1)^2} [2\phi^2\psi'' + \phi\psi' + 1]. \end{aligned} \quad (5.38)$$

Assim, temos

$$c = 1 + \frac{1}{12n\phi^{(0)}} \left[\frac{5(\phi^{(0)2}\psi_0'' + 1)^2}{(\phi^{(0)}\psi_0' - 1)^3} - \frac{3\phi^{(0)2}(\phi^{(0)}\psi_0''' + 2\psi_0'')}{(\phi^{(0)}\psi_0' - 1)^2} - \frac{3}{\phi^{(0)}\psi_0' - 1} \right], \quad (5.39)$$

sendo ψ_0' , ψ_0'' , ψ_0''' as três primeiras derivadas da função $\psi(\cdot)$, calculadas em $\phi = \phi^{(0)}$.

É importante notar que o cálculo das correções em (5.36) e (5.39) envolve a avaliação das três primeiras derivadas da função digama, que, para efeitos práticos, pode ser difícil de serem calculadas. Assim, é bastante útil obtermos expansões assintóticas que produzam aproximações simples para estas fórmulas. Como no Capítulo 4, podemos expressar a correção de Bartlett, para os modelos biparamétricos pertencentes à família exponencial, na forma

$$c = 1 + \frac{\rho(\theta, \phi)}{12n}.$$

No caso da expressão (5.36), escrevemos

$$\rho_1(\theta, \phi) = \rho_1(\phi) = \frac{1}{\phi} \left[\frac{-3(2\phi^2\psi'' + \phi\psi' + 1)}{(\phi\psi' - 1)^2} + 2 \right],$$

e para a expressão (5.39),

$$\rho_2(\theta, \phi) = \rho_2(\phi) = \frac{1}{\phi} \left[\frac{5(\phi^2 \psi'' + 1)^2}{(\phi \psi - 1)^3} - \frac{3\phi^2(\phi \psi''' + 2\psi'')}{(\phi \psi' - 1)^2} - \frac{3}{\phi \psi' - 1} \right].$$

Para conseguirmos uma aproximação simples para as expressões de ρ_1 e ρ_2 acima, utilizamos a expansão de $\psi'(x)$ para valores grandes de x , introduzida no Capítulo 4, isto é,

$$\psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \frac{1}{42x^7} + O(x^{-9}).$$

Após alguma manipulação algébrica, via MATHEMATICA, obtemos, para valores grandes de ϕ ,

$$\rho_1(\phi) = 18 - \frac{2}{3\phi^2} - \frac{8}{15\phi^3} + O(\phi^{-4}) \quad (5.39')$$

e

$$\rho_2(\phi) = 22 - \frac{2}{\phi} + \frac{14}{27\phi^3} + O(\phi^{-4}). \quad (5.39'')$$

A fim de verificarmos o desempenho das expansões utilizadas no desenvolvimento acima, calculamos os valores exatos e aproximados de ρ_1 e ρ_2 em (5.39') e (5.39''), respectivamente, para alguns valores de ϕ . Os resultados encontrados estão na Tabela 5.1

Tabela 5.1: Valores exatos e aproximados de $\rho_1(\phi)$ e $\rho_2(\phi)$, ϕ grande, para a distribuição gama.

ϕ	$\rho_1(\phi)$		$\rho_2(\phi)$	
	exata	aproximada	exata	aproximada
0,8	17,5606	15,9167	19,4041	20,5127
1,0	17,6029	16,8000	19,9376	20,5185
1,5	17,7443	17,5457	20,6560	20,8203
2,0	17,8353	17,7667	21,0035	21,0648
5,0	17,9709	17,9691	21,6024	21,6041
10,0	17,9929	17,9928	21,8004	21,8005
20,0	17,9983	17,9983	21,9001	21,9001

Observando a Tabela 5.1 concluímos que as expressões aproximadas para $\rho_1(\phi)$ e $\rho_2(\phi)$, em (5.39') e (5.39''), têm um desempenho bastante satisfatório, uma vez que seus valores estão

muito próximos dos correspondentes valores exatos. Além disto, conforme esperado destas fórmulas, à medida que ϕ cresce, ρ_1 e ρ_2 convergem para 18 e 22, respectivamente.

5.4.4 DISTRIBUIÇÃO LOG-GAMA $LG(\theta, \phi)$

A função densidade da distribuição log-gama, com parâmetros θ e ϕ ortogonais, é definida por

$$\pi(y; \theta, \phi) = \frac{\phi^\phi}{\Gamma(\phi)} \exp\{\phi[y - \theta - \exp(y - \theta)]\}, \quad \phi > 0, \quad \theta, y \in \mathbb{R}.$$

Podemos escrever, então, $\alpha_1(\theta, \phi) = -\phi\theta^{-\theta}$, $\alpha_2(\theta, \phi) = \phi$, $\rho(\theta, \phi) = \phi\theta - \phi \log \phi + \log[\Gamma(\phi)]$, $d_1(y) = \exp(y)$, $d_2(y) = y$ e $v(y) = 0$. Através de (5.21) ou (5.24) encontramos $\beta_1 = \exp(\theta)$ e $\beta_2 = \theta + \psi(\phi) - \log \phi$. A estimativa de máxima verossimilhança irrestrita para θ é $\hat{\theta} = \log\left(\frac{\sum_{l=1}^n \exp y_l}{n}\right)$ e para ϕ pode ser obtida da equação não-linear

$$\log \hat{\phi} - \psi(\hat{\phi}) = \log \left[\frac{\sum_{l=1}^n (\exp y_l - \exp \bar{y})}{n} \right].$$

Para testar $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$ a estatística da razão de verossimilhança pode ser calculada por (5.30) e, para este problema, tem a forma

$$\omega = 2n \left\{ \log \left(\frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \right) - \log \left(\frac{\tilde{\phi}^{\tilde{\phi}}}{\Gamma(\tilde{\phi})} \right) + \hat{\phi}(\bar{y} - \hat{\theta} - 1) - \tilde{\theta}(\bar{y} - \theta^{(0)} - \frac{e^{-\theta^{(0)}}}{n} \sum_{l=1}^n \exp(y_l)) \right\}, \quad (5.40)$$

onde $\tilde{\phi}$ é a estimativa de máxima verossimilhança para ϕ , sob a hipótese nula H , e pode ser obtida resolvendo-se a equação não-linear

$$\log \tilde{\phi} - \psi(\tilde{\phi}) = \theta^{(0)} - \bar{y} + \frac{e^{-\theta^{(0)}}}{n} \sum_{l=1}^n \exp(y_l) - 1. \quad (5.41)$$

A correção de Bartlett associada à estatística ω em (5.40) é obtida de (5.15), onde os termos A_θ e $C_{\theta, \phi}$ são calculados, respectivamente, por (5.22) e (5.23). Substituindo-se as derivadas envolvidas nestas expressões, encontramos que o fator de correção c , neste caso, é idêntico àquele do modelo gama, dado por (5.36) e, assim, $\omega^* = c^{-1}\omega$ tem distribuição assintótica χ_1^2 ,

até ordem n^{-1} . A justificativa deste fato deve-se à correção de Bartlett ser invariante em relação à transformação dos dados, mantida a mesma hipótese nula.

Se o interesse, agora, é o teste sobre ϕ , ou seja, $H : \phi = \phi^{(0)}$ contra $A : \phi \neq \phi^{(0)}$, a estatística da razão de verossimilhança é calculada por $\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\tilde{\theta}, \phi^{(0)})]$, onde $\tilde{\theta}$ é a estimativa de máxima verossimilhança para θ sob a hipótese nula. Como $\tilde{\theta} = \hat{\theta} = \log(\sum_{l=1}^n \exp(y_l)/n)$, temos

$$\omega = 2n \left\{ \log \left(\frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \right) - \log \left(\frac{\phi^{(0)\phi^{(0)}}}{\Gamma(\phi^{(0)})} \right) + (\hat{\phi} - \phi^{(0)})(\bar{y} - \hat{\theta} - 1) \right\}. \quad (5.42)$$

O fator de correção c , que aperfeiçoa a estatística ω em (5.42), vem de (5.15), com os termos A_ϕ e $C_{\theta, \phi} = C_{1, \theta, \phi} - C_{2, \theta, \phi}$ calculados através de (5.16) e (5.25), respectivamente, com as derivadas tomadas em relação ao vetor de parâmetros $(\phi, \theta)^\top$. O fator c resultante, neste caso, coincide com aquele do modelo gama para esta hipótese, dado por (5.39), conforme esperado, pela invariância da correção de Bartlett em relação à transformação dos dados, mantida a mesma hipótese nula. Assim, a estatística aperfeiçoada por este fator, $\omega^* = c^{-1}\omega$, tem distribuição χ_1^2 até ordem n^{-1} , sob a hipótese nula.

5.4.5 DISTRIBUIÇÃO GAMA INVERSA $GI(\theta, \phi)$

Finalizando, consideramos a distribuição gama inversa, cuja função densidade pode ser escrita por

$$\pi(y; \theta, \phi) = \frac{(\theta\phi)^\phi}{\Gamma(\phi)} \left(\frac{1}{y} \right)^{\phi+1} \exp \left(\frac{-\theta\phi}{y} \right), \quad y > 0, \quad \theta \text{ e } \phi > 0. \quad (5.43)$$

Esta parametrização faz com que θ e ϕ sejam globalmente ortogonais (vide Seção 5.5). Assim, podemos adotar $\alpha_1(\theta, \phi) = -\theta\phi$, $\alpha_2 = (\theta, \phi) = -\phi$, $\rho(\theta, \phi) = -\phi \log(\theta\phi) + \log \Gamma(\phi)$, $d_1(y) = 1/y$, $d_2(y) = \log y$ e $v(y) = -\log y$. De (5.21) segue que $\beta_1 = 1/\theta$ e $\beta_2 = \log \theta + \log \phi - \psi(\phi)$. O estimador de máxima verossimilhança para θ é

$$\hat{\theta} = \left(\frac{1}{n} \sum_{l=1}^n \frac{1}{y_l} \right)^{-1} = \bar{y}_h \quad (5.44)$$

e para ϕ é $\hat{\phi}$, solução da equação não-linear

$$\log \hat{\phi} - \psi(\hat{\phi}) = \log(\bar{y}_h \bar{y}_g) , \quad (5.45)$$

onde $\bar{y}_h = n / \sum_{l=1}^n y_l^{-1}$ é a média harmônica e $\bar{y}_g = (\prod_{l=1}^n y_l)^{1/n}$ é a média geométrica dos dados $y_1, \dots, y_n$.

Para testar $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$ a estatística da razão de verossimilhança, calculada por (5.30), tem a forma

$$\omega = 2n \left\{ \log \left(\frac{(\hat{\theta} \hat{\phi})^{\hat{\phi}}}{\Gamma(\hat{\phi})} \right) - \log \left(\frac{(\theta^{(0)} \tilde{\phi})^{\tilde{\phi}}}{\Gamma(\tilde{\phi})} \right) + (\hat{\phi} - \tilde{\phi}) \log \bar{y}_g - \left(\hat{\phi} - \frac{\theta^{(0)}}{\bar{y}_h} \tilde{\phi} \right) \right\} , \quad (5.46)$$

onde $\tilde{\phi}$ é o estimador de máxima verossimilhança para ϕ , segundo a hipótese nula. Não é difícil mostrarmos que $\tilde{\phi}$ é solução da equação não-linear,

$$\log \tilde{\phi} - \psi(\tilde{\phi}) = -\log(\theta^{(0)} / \bar{y}_g) + \frac{\theta^{(0)}}{\bar{y}_h} - 1 . \quad (5.47)$$

A correção de Bartlett c , que aperfeiçoa a estatística ω em (5.46), segue de (5.15), onde os termos A_θ e $C_{\theta, \phi} = C_{1_{\theta, \phi}} - C_{2_{\theta, \phi}}$ são calculados usando (5.22) e (5.23), respectivamente. Substituindo as derivadas envolvidas nestas expressões (vide Apêndice 4) encontramos que o fator c , neste problema, é o mesmo do modelo gama para o teste sobre o parâmetro θ . Logo, $\omega^* = c^{-1} \omega$ tem distribuição χ_1^2 até ordem n^{-1} sob a hipótese nula, onde c é calculado através de (5.36).

Para testar $H : \phi = \phi^{(0)}$ contra $A : \phi \neq \phi^{(0)}$ a estatística da razão de verossimilhança, calculada por (5.31), pode ser expressa por

$$\omega = 2n \left\{ \log \left(\frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \right) - \log \left(\frac{\phi^{(0) \phi^{(0)}}}{\Gamma(\phi^{(0)})} \right) + (\hat{\phi} - \phi^{(0)}) (\log(\bar{y}_h \bar{y}_g) - 1) \right\} . \quad (5.48)$$

A estatística melhorada $\omega^* = c^{-1} \omega$ tem distribuição χ_1^2 até ordem n^{-1} , segundo a hipótese nula, e o fator c é calculado por (5.15), sendo que os termos A_ϕ e $C_{\theta, \phi} = C_{1_{\theta, \phi}} - C_{2_{\theta, \phi}}$ envolvidos vêm de (5.16) e (5.25), respectivamente. Após substituição das derivadas (vide Apêndice 4), encontramos que o fator c , que corrige ω em (5.48), coincide com aquele do modelo gama para o teste do parâmetro ϕ , e é dado por (5.39).

5.5 CONSTRUÇÃO DE PARÂMETROS GLOBALMENTE ORTOGONAIS

Genericamente, consideremos um problema regular cujo logaritmo da função de verossimilhança ℓ é função de dois parâmetros unidimensionais α e β . Dizemos que α e β são globalmente ortogonais se $E[\partial^2 \ell / \partial \alpha \partial \beta] = 0$.

Seja $f(y; \theta, \phi)$ uma função densidade que pertence à família exponencial biparamétrica definida por (5.1) e $\ell(\theta, \phi)$ o logaritmo da função de verossimilhança definido por (5.2). Neste caso, $\kappa_{\theta\phi} = E[\partial^2 \ell / \partial \theta \partial \phi] = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)}$ e $\kappa_{\phi\theta} = E[\partial^2 \ell / \partial \phi \partial \theta] = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,0)}$. Se θ e ϕ não são parâmetros globalmente ortogonais, isto é, se as equações $\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)} = 0$ e $\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,0)} = 0$ não são satisfeitas, então, podemos reparametrizar o modelo segundo Cox e Reid (1987). Se ϕ for o parâmetro de interesse e θ o parâmetro de perturbação, o objetivo é determinar uma transformação dos parâmetros $(\theta, \phi)^\top$ para $(\mu, \phi)^\top$ de tal forma que μ e ϕ sejam globalmente ortogonais. Cox e Reid (1987) mostram que μ , globalmente ortogonal a ϕ , é dado por qualquer solução do sistema de equações diferenciais parciais,

$$E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) \frac{\partial \theta}{\partial \phi} = -E \left(\frac{\partial^2 \ell}{\partial \theta \partial \phi} \right), \quad (5.49)$$

isto é, $\frac{\partial \theta}{\partial \phi} = -\frac{\kappa_{\theta\phi}}{\kappa_{\theta\theta}}$. Na família exponencial biparamétrica temos, por (5.4) e (5.5), que

$$\kappa_{\theta\theta} = E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \quad \text{e} \quad \kappa_{\theta\phi} = E \left(\frac{\partial^2 \ell}{\partial \theta \partial \phi} \right) = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)}.$$

Assim, encontrar a solução da equação (5.49) é encontrar a solução da equação diferencial parcial

$$\frac{\partial \theta}{\partial \phi} = -\frac{\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)}}{\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)}}. \quad (5.50)$$

A solução de (5.50) fornece a nova parametrização e, portanto, recaímos no caso de parâmetros globalmente ortogonais, em que podemos aplicar os resultados das seções anteriores.

É importante destacarmos que a equação (5.50) não tem solução geral e, assim, deve ser

resolvida para cada distribuição que pertence à família exponencial biparamétrica, em que as equações

$$\kappa_{\theta\phi} = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)} = 0 \quad (5.51)$$

e

$$\kappa_{\phi\theta} = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,0)} = 0 \quad (5.52)$$

não são satisfeitas. De forma análoga, se o parâmetro de interesse é θ , sendo, portanto, ϕ o parâmetro de incômodo, devemos determinar uma transformação dos parâmetros $(\theta, \phi)^\top$ para $(\theta, \gamma)^\top$, onde θ e γ são globalmente ortogonais, e γ é determinado pela solução do sistema de equações diferenciais parciais

$$\frac{\partial\phi}{\partial\theta} = -\frac{\kappa_{\theta\phi}}{\kappa_{\phi\phi}} = -\frac{\sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)}}{\sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)}}. \quad (5.53)$$

Apresentamos, a seguir, alguns exemplos que ilustram a aplicação das equações (5.50) e (5.53).

Exemplo 5.1 Consideremos n observações independentes $y_1, \dots, y_n$ de uma distribuição $N(\theta, \theta^2\phi^2)$. Este modelo normal tem coeficiente de variação igual a ϕ . Sua função densidade é expressa por

$$\pi(y; \theta, \phi) = \frac{1}{\theta\phi\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\theta^2\phi^2} (y - \theta)^2 \right\}. \quad (5.54)$$

O logaritmo da função de verossimilhança total de θ e ϕ é

$$\ell = \ell(\theta, \phi) = -\frac{1}{2\theta^2\phi^2} \sum_{l=1}^n y_l^2 + \frac{1}{\theta\phi^2} \sum_{l=1}^n y_l - \frac{n}{2\phi^2} - n \log(\theta\phi) - \frac{n}{2} \log(2\pi).$$

Para esta função temos $\alpha_1 = -(2\theta^2\phi^2)^{-1}$, $\alpha_2 = (\theta\phi^2)^{-1}$ e $\rho = (2\phi^2)^{-1} + \log(\theta\phi)$. Por (5.6) obtemos $\beta_1 = \theta^2(\phi^2 + 1)$ e $\beta^2 = \theta$. Através de (5.4) e (5.5) encontramos

$$E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = \frac{-n(2\phi^2 + 1)}{\theta^2\phi^2}, \quad E \left(\frac{\partial^2 \ell}{\partial \phi^2} \right) = -\frac{2n}{\phi^2} \quad \text{e} \quad E \left(\frac{\partial^2 \ell}{\partial \theta \partial \phi} \right) = \frac{-2n}{\theta\phi}.$$

As estimativas de máxima verossimilhança para os parâmetros θ e ϕ da densidade (5.54) são obtidas resolvendo-se o sistema de equações

$$\begin{aligned}\frac{\partial \ell}{\partial \theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2 \phi^2} \left\{ \sum_{i=1}^n \frac{(y_i - \theta)^2}{\theta} + \sum_{i=1}^n (y_i - \theta) \right\} = 0, \\ \frac{\partial \ell}{\partial \phi} &= -\frac{n}{\phi} + \frac{1}{\theta^2 \phi^3} \sum_{i=1}^n (y_i - \theta)^2 = 0.\end{aligned}$$

É fácil mostrar que as soluções são dadas por

$$\hat{\theta} = \bar{y} \quad \text{e} \quad \hat{\phi} = \frac{s}{\bar{y}} \quad (y \neq 0),$$

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

Se estamos interessados em testar $\phi = \phi^{(0)}$, então, por (5.49) ou (5.50), temos a seguinte equação diferencial:

$$\frac{1}{2\theta} \frac{\partial \theta}{\partial \phi} = -\frac{\phi}{2\phi^2 + 1}.$$

Uma possível solução desta equação é $a(\mu) = \theta(2\phi^2 + 1)^{1/2}$, onde $a(\cdot)$ é uma função arbitrária.

Fazendo $a(\mu) = \mu$, temos os parâmetros globalmente ortogonais ϕ e $\mu = \theta(2\phi^2 + 1)^{1/2}$.

Podemos checar a ortogonalidade da reparametrização $\mu = \theta(2\phi^2 + 1)^{1/2}$ e ϕ , considerando a nova forma da função densidade em (5.54), que agora é expressa por

$$\begin{aligned}\pi^*(y; \mu, \phi) &= \frac{(2\phi^2 + 1)^{1/2}}{\phi \mu \sqrt{2\pi}} \exp \left\{ -\frac{(2\phi^2 + 1)}{2\mu^2 \phi} \left(y - \frac{\mu}{(2\phi^2 + 1)^{1/2}} \right)^2 \right\} \\ &= \exp \left\{ \frac{-y^2(2\phi^2 + 1)}{2\mu^2 \phi^2} + \frac{y(2\phi^2 + 1)^{1/2}}{\mu \phi^2} - \frac{1}{2\phi^2} - \log \left[\frac{\mu \phi}{(2\phi^2 + 1)^{1/2}} \right] - \frac{1}{2} \log(2\pi) \right\},\end{aligned}\tag{5.54'}$$

sendo, portanto, $\alpha_1 = \alpha_1(\mu, \phi) = -(2\phi^2 + 1)/2\mu^2 \phi^2$, $\alpha_2 = \alpha_2(\mu, \phi) = (2\phi^2 + 1)^{1/2}/\mu \phi^2$ e $\rho(\mu, \phi) = 1/2\phi^2 + \log [\mu \phi / (2\phi^2 + 1)^{1/2}]$. Através de (5.6) encontramos $\beta_1 = \mu^2(\phi^2 + 1)/(2\phi^2 + 1)$ e $\beta_2 = \mu/(2\phi^2 + 1)^{1/2}$. Para estas funções, substituindo as derivadas apropriadas, tomadas em relação ao vetor $(\mu, \phi)^\top$, nas expressões em (5.5), encontramos $\kappa_{\mu\phi} = \kappa_{\phi\mu} = 0$.

Para calcular a estatística da razão de verossimilhança ω , associada ao teste da hipótese nula $H : \phi = \phi^{(0)}$, necessitamos encontrar a estimativa de máxima verossimilhança de θ , segundo

esta hipótese, para a densidade (5.54). A estimativa de θ , neste caso, é solução da equação

$$-\frac{n}{\theta} + \frac{1}{\phi^{(0)^2}} \left\{ \frac{1}{\theta^3} \sum_{i=1}^n (y_i - \theta)^2 + \frac{1}{\theta^2} \sum_{i=1}^n (y_i - \theta) \right\} = 0 ,$$

que pode ser reescrita por

$$\sum_{i=1}^n (y_i - \theta)^2 + \theta \sum_{i=1}^n (y_i - \theta) = n\theta^2 \phi^{(0)^2}$$

ou, ainda, por

$$n\phi^{(0)^2} \theta^2 + n\bar{y}\theta - \sum_{i=1}^n y_i^2 = 0 .$$

Seja m_2 o segundo momento ordinário amostral, isto é, $m_2 = \sum_{i=1}^n y_i^2 / n$. Então, temos

$$\phi^{(0)^2} \tilde{\theta}^2 + \bar{y}\tilde{\theta} - m_2 = 0 ,$$

que resulta

$$\tilde{\theta} = \frac{-\bar{y} \pm \sqrt{\bar{y}^2 + 4\phi^{(0)^2} m_2}}{2\phi^{(0)^2}} .$$

Mostramos, a seguir, qual solução $\tilde{\theta}$ (note que $m_2 > 0$) deve ser considerada, tal que maximiza o logaritmo da função de verossimilhança segundo H , ou seja, minimiza a razão de máxima verossimilhança ω . Seja $\hat{\ell}_A$ o logaritmo da verossimilhança calculada em $\hat{\theta}$ e $\hat{\phi}$ e $\hat{\ell}_H$ o logaritmo da verossimilhança calculada em $\tilde{\theta}$ e $\phi^{(0)}$, ou seja, $\hat{\ell}_A = \ell(\hat{\theta}, \hat{\phi})$ e $\hat{\ell}_H = \ell(\tilde{\theta}, \phi^{(0)})$. Temos,

$$\begin{aligned} \hat{\ell}_A &= -\frac{n}{2} \log s^2 - \frac{n}{2} \quad \text{e} \\ \hat{\ell}_H &= -\frac{n}{2} \log(\phi^{(0)^2} \tilde{\theta}^2) - \frac{1}{2\phi^{(0)^2} \tilde{\theta}^2} \sum_{i=1}^n (y_i - \tilde{\theta})^2 . \end{aligned}$$

Então, a estatística da razão de verossimilhança, $\omega = 2(\hat{\ell}_A - \hat{\ell}_H)$, é escrita por

$$\omega = n \left\{ \log(\phi^{(0)^2} \tilde{\theta}^2) - \log s^2 - 1 + \frac{\sum_{i=1}^n (y_i - \tilde{\theta})^2}{n\phi^{(0)^2} \tilde{\theta}^2} \right\} .$$

Como $\tilde{\theta}$ anula a derivada do logaritmo da função de verossimilhança em relação a θ sob a hipótese nula, então $\tilde{\theta}$ satisfaz

$$\sum_{i=1}^n (y_i - \tilde{\theta})^2 + \tilde{\theta} \sum_{i=1}^n (y_i - \tilde{\theta}) = n\tilde{\theta}^2 \phi^{(0)^2} ,$$

que implica

$$\frac{\sum_{i=1}^n (y_i - \tilde{\theta})^2}{n\tilde{\theta}^2\phi^{(0)^2}} = 1 - \frac{\bar{y}}{\phi^{(0)^2}\tilde{\theta}} + \frac{1}{\phi^{(0)^2}}.$$

Logo,

$$\omega = n \left\{ \log(\phi^{(0)^2}\tilde{\theta}^2) - \log s^2 + \frac{1}{\phi^{(0)^2}} - \frac{\bar{y}}{\phi^{(0)^2}\tilde{\theta}} \right\}.$$

Somando e subtraindo $\log \bar{y}^2$, podemos escrever

$$\omega = n \left\{ \log \left(\frac{\phi^{(0)^2}\bar{y}^2}{s^2} \right) + \frac{1}{\phi^{(0)^2}} + \log \left(\frac{\tilde{\theta}}{\bar{y}} \right)^2 - \frac{\bar{y}}{\phi^{(0)^2}\tilde{\theta}} \right\}.$$

Seja $A = \log(\phi^{(0)^2}\bar{y}^2/s^2) + 1/\phi^{(0)^2}$ e $B = \log(\tilde{\theta}/\bar{y})^2 - \bar{y}/\phi^{(0)^2}\tilde{\theta}$. Então, $\omega = n\{A + B\}$.

O termo A não depende da solução $\tilde{\theta}$. Entretanto, B depende. Assim, queremos $\tilde{\theta}$ tal que B seja o menor que corresponde ao maior valor da log-verossimilhança segundo H . Tem-se $B = \log \delta^2 - \frac{1}{\phi^{(0)^2}}\delta^{-1}$, onde $\delta = \tilde{\theta}/\bar{y}$. Então, $\delta = \left(-1 \pm \sqrt{1 + 4\phi^{(0)^2}m_2/\bar{y}^2}\right)/2\phi^{(0)^2}$, onde o radicando é sempre positivo, mais precisamente, maior do que a unidade. Da análise da função $\log \delta^2 - \frac{1}{\phi^{(0)^2}}\delta^{-1}$, concluímos que $\delta > 0$ corresponde ao menor valor para B , que corresponde a tomarmos $\tilde{\theta}$ de mesmo sinal que $\bar{y}$. Logo, $\delta > 0$ implica nas soluções

$$\tilde{\theta} = \frac{\sqrt{\bar{y}^2 + 4\phi^{(0)^2}m_2} - \bar{y}}{2\phi^{(0)^2}}, \text{ quando } \bar{y} > 0$$

e

$$\tilde{\theta} = \frac{-\sqrt{\bar{y}^2 + 4\phi^{(0)^2}m_2} - \bar{y}}{2\phi^{(0)^2}}, \text{ quando } \bar{y} < 0.$$

Assim, esses valores de $\tilde{\theta}$ são usados para calcular a estatística da razão de máxima verossimilhança $\omega = 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\tilde{\theta}, \phi^{(0)})]$. A estatística ω é então corrigida por $\omega^* = c^{-1}\omega$ que tem, sob a hipótese nula, distribuição χ_1^2 até ordem n^{-1} . A correção de Bartlett c é calculada por $c = 1 + A_\phi + C_{\theta, \phi}$, onde as parcelas A_ϕ e $C_{\theta, \phi}$ são obtidas de (5.16) e (5.17), respectivamente com derivadas das funções α_1 , α_2 , β_1 e β_2 , correspondentes à densidade (5.54'), tomadas em relação ao vetor $(\phi, \mu)^T$. Utilizando o pacote MATHEMATICA encontramos

$$A_\phi = \frac{1 + 6\phi^2 + 18\phi^4 + 15\phi^6}{3n(1 + 2\phi^2)^3}$$

e

$$C_{\theta, \phi} = \frac{3 + 8\phi^2 + 7\phi^4}{2n(1 + 2\phi^2)^2} ,$$

resultando

$$c = 1 + \frac{1}{6n} \left[\frac{3(3 + 8\phi^2 + 7\phi^4)}{(1 + 2\phi^2)^2} + \frac{2(1 + 6\phi^2 + 18\phi^4 + 15\phi^6)}{(1 + 2\phi^2)^3} \right] .$$

Se, entretanto, θ é o parâmetro de interesse, então devemos encontrar a transformação de $(\theta, \phi)^\top$ para $(\theta, \gamma)^\top$, tal que θ e γ sejam ortogonais. A nova parametrização pode ser obtida através de (5.53), que resulta na equação diferencial

$$\frac{1}{\phi} \frac{\partial \phi}{\partial \theta} = \frac{1}{\theta} .$$

Uma possível solução da equação acima é $a(\gamma) = \theta\phi$, onde $a(\cdot)$ é uma função arbitrária. Fazendo $a(\gamma) = \gamma$ temos os parâmetros globalmente ortogonais θ e $\gamma = \theta\phi$. A ortogonalidade entre θ e γ pode ser checada escrevendo a função densidade em (5.54) na nova parametrização. Temos, então,

$$\begin{aligned} \pi^*(y; \theta, \gamma) &= \frac{1}{\sqrt{2\pi\gamma}} \exp \left\{ -\frac{1}{2\gamma^2} (y - \theta)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2\gamma^2} y^2 + \frac{\theta}{\gamma^2} y - \log(\sqrt{2\pi\gamma}) - \frac{\theta^2}{2\gamma^2} \right\} . \end{aligned} \quad (5.54'')$$

Esta forma acarreta $\alpha_1 = -1/2\gamma^2$, $\alpha_2 = \theta/\gamma^2$ e $\rho = \log(\sqrt{2\pi\gamma}) + \theta^2/(2\gamma^2)$, que através de (5.6) produz $\beta_1 = \gamma^2 + \theta^2$ e $\beta_2 = \theta$. Utilizando estas funções nas expressões em (5.6), com derivadas tomadas em relação ao vetor $(\theta, \gamma)^\top$, facilmente mostramos que $\kappa_{\theta\gamma} = \kappa_{\gamma\theta} = 0$.

Para o teste do parâmetro θ a estatística da razão de verossimilhança é calculada por (5.27), e a correspondente correção de Bartlett é idêntica àquela da Seção 5.4.1, isto é, $c = 1 + 3/2n$. Este fato é esperado, pois para testar a média θ da $N(\theta, \phi)$ ou da $N(\theta, \theta^2\phi^2)$ as estatísticas da razão de verossimilhança são iguais, uma vez que esta estatística é invariante sob reparametrização e, assim, as correções de Bartlett para esta hipótese são, também, idênticas.

Exemplo 5.2 Seja $y_1, \dots, y_n$ uma amostra aleatória da distribuição normal inversa $NI(\theta, \theta^2 \phi^2)$, sendo $\theta > 0$ a sua média e $\phi > 0$ o coeficiente de variação, cuja função densidade pode ser escrita por

$$\pi(y; \theta, \phi) = \exp \left\{ -\frac{(y - \theta)^2}{2y\theta\phi^2} + \frac{1}{2} \log \left(\frac{\theta}{\phi^2} \right) - \frac{1}{2} \log(2\pi y^3) \right\}. \quad (5.55)$$

O logaritmo da função de verossimilhança total de θ e ϕ é

$$\ell = \ell(\theta, \phi) = -\frac{1}{2\theta\phi^2} \sum_{l=1}^n y_l - \frac{\theta}{2\phi^2} \sum_{l=1}^n \frac{1}{y_l} + \frac{n}{\phi^2} + \frac{n}{2} \log \left(\frac{\theta}{\phi^2} \right) - \frac{1}{2} \sum_{l=1}^n \log(2\pi y_l^3).$$

Neste caso, temos $\alpha_1 = -1/(2\theta\phi^2)$, $\alpha_2 = -\theta/(2\phi^2)$ e $\rho = -1/\phi^2 - \frac{1}{2} \log(\theta/\phi^2)$. Por (5.6) obtemos $\beta_1 = \theta$ e $\beta_2 = (1 + \phi^2)/\theta$. De (5.4) e (5.5) vem

$$E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = \frac{-n(\phi^2 + 2)}{2\theta^2 \phi^2}, \quad E \left(\frac{\partial^2 \ell}{\partial \phi^2} \right) = -\frac{2n}{\phi^2} \quad \text{e} \quad E \left(\frac{\partial^2 \ell}{\partial \theta \partial \phi} \right) = \frac{n}{\theta \phi}.$$

As estimativas de máxima verossimilhança para os parâmetros θ e ϕ da densidade (5.55) são obtidas resolvendo-se o seguinte sistema de equações:

$$\begin{aligned} \frac{\partial \ell}{\partial \theta} &= -\frac{1}{2\phi^2} \sum_{i=1}^n \left[-\frac{(y_i - \theta)^2}{y_i \theta^2} - 2 \frac{(y_i - \theta)}{y_i \theta} \right] + \frac{n}{2\theta} = 0, \\ \frac{\partial \ell}{\partial \phi} &= \frac{1}{\phi^3 \theta} \sum_{i=1}^n \frac{(y_i - \theta)^2}{y_i} - \frac{n}{\phi} = 0. \end{aligned}$$

É fácil verificarmos que as soluções desse sistema de equações são

$$\hat{\theta} = \bar{y} \quad \text{e} \quad \hat{\phi}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\bar{y} y_i}.$$

A expressão para $\hat{\phi}^2$ pode ser reescrita numa forma mais simples por

$$\hat{\phi}^2 = \frac{\bar{y}}{\bar{y}_h} - 1.$$

Se estamos interessados no parâmetro ϕ , a transformação do vetor $(\theta, \phi)^\top$ para o vetor $(\mu, \phi)^\top$, tal que μ e ϕ sejam ortogonais, pode ser obtida pela solução de (5.49) ou (5.50), que resulta na equação diferencial

$$\frac{\partial \theta}{\partial \phi} = \frac{2\theta\phi}{\phi^2 + 2}.$$

Uma solução possível para esta equação diferencial é $a(\mu) = \theta/(\phi^2 + 2)$, com $a(\cdot)$ uma função arbitrária. Fazendo $a(\mu) = \mu$, resulta nos parâmetros ortogonais $\mu = \theta/(\phi^2 + 2)$ e ϕ . Para verificação da ortogonalidade entre μ e ϕ , reescrevemos a função densidade em (5.55) na nova parametrização. Temos,

$$\pi^*(y; \mu, \phi) = \exp \left\{ \frac{-[y - \mu(\phi^2 + 2)]^2}{2y\mu\phi^2(\phi^2 + 2)} + \frac{1}{2} \log \left[\frac{\mu(\phi^2 + 2)}{\phi^2} \right] - \frac{1}{2} \log(2\pi y^3) \right\}, \quad (5.55')$$

sendo, então, $\alpha_1 = \alpha_1(\mu, \phi) = -1/(2\phi^2\mu(\phi^2 + 2))$, $\alpha_2 = \alpha_2(\mu, \phi) = -\mu(\phi^2 + 2)/2\phi^2$ e $\rho(\mu, \phi) = -1/\phi^2 + \frac{1}{2} \log \phi^2 - \frac{1}{2} \log[\mu(\phi^2 + 2)]$, que, por (5.6) produz $\beta_1 = \mu(\phi^2 + 2)$ e $\beta_2 = (\phi^2 + 1)/[\mu(\phi^2 + 2)]$. Com estas funções, assim definidas, é fácil mostrarmos que $\kappa_{\mu\phi} = \kappa_{\phi\mu} = 0$. Para calcular a estatística da razão de verossimilhança ω para o teste de $H : \phi = \phi^{(0)}$, precisamos encontrar a estimativa de máxima verossimilhança de θ , restrita a $\phi = \phi^{(0)}$. Essa estimativa $\tilde{\theta}$ é solução da equação

$$\frac{\sum_{i=1}^n (y_i^2 - \theta^2)}{y_i} + \frac{n}{\theta} = 0,$$

ou, equivalentemente, da equação

$$\bar{y}_h^{-1} \theta^2 - \phi^{(0)2} \theta - \bar{y} = 0.$$

As raízes da equação anterior são dadas por

$$\tilde{\theta}_1 = \frac{\phi^{(0)2} + \sqrt{\phi^{(0)4} + 4\bar{y}\bar{y}_h^{-1}}}{2\bar{y}_h^{-1}} \quad \text{e} \quad \tilde{\theta}_2 = \frac{\phi^{(0)2} - \sqrt{\phi^{(0)4} + 4\bar{y}\bar{y}_h^{-1}}}{2\bar{y}_h^{-1}}.$$

Como $\theta > 0$, então só interessa a solução

$$\tilde{\theta} = \frac{\bar{y}_h}{2} \left\{ \phi^{(0)2} + \sqrt{\phi^{(0)4} + 4\bar{y}/\bar{y}_h} \right\}.$$

Assim, a estatística da razão de verossimilhança para este teste é dada por $\omega = 2[\hat{\ell}_A - \hat{\ell}_H]$, sendo

$$\hat{\ell}_A = \ell(\hat{\theta}, \hat{\phi}) = -\frac{1}{2\hat{\theta}\hat{\phi}^2} \sum_{i=1}^n \frac{(y_i - \hat{\theta})^2}{y_i} + \frac{n}{2} \log \left(\frac{\hat{\theta}}{\hat{\phi}^2} \right) - \frac{1}{2} \sum_{i=1}^n \log(2\pi y_i^3)$$

e

$$\hat{\ell}_H = \ell(\tilde{\theta}, \phi^{(0)}) = -\frac{1}{2\tilde{\theta}\phi^{(0)2}} \sum_{i=1}^n \frac{(y_i - \tilde{\theta})^2}{y_i} + \frac{n}{2} \log \left(\frac{\tilde{\theta}}{\phi^{(0)2}} \right) - \frac{1}{2} \sum_{i=1}^n \log(2\pi y_i^3).$$

Logo,

$$\begin{aligned}\omega &= \frac{1}{\tilde{\theta}\phi^{(0)^2}} \sum_{i=1}^n \frac{(y_i - \tilde{\theta})^2}{y_i} - \frac{1}{\hat{\theta}\hat{\phi}^2} \sum_{i=1}^n \frac{(y_i - \hat{\theta})^2}{y_i} + n \log \left(\frac{\hat{\theta}\phi^{(0)^2}}{\tilde{\theta}\hat{\phi}^2} \right) \\ &= \frac{1}{\tilde{\theta}\phi^{(0)^2}} \sum_{i=1}^n \frac{(y_i - \tilde{\theta})^2}{y_i} - \frac{1}{\bar{y} \left(\frac{\bar{y}}{\bar{y}_h} - 1 \right)} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{y_i} + n \log \left(\frac{\bar{y}\phi^{(0)^2}}{\tilde{\theta} \left(\frac{\bar{y}}{\bar{y}_h} - 1 \right)} \right) .\end{aligned}$$

Mas,

$$\frac{1}{\bar{y} \left(\frac{\bar{y}}{\bar{y}_h} - 1 \right)} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{y_i} = \frac{1}{\bar{y} \left(\frac{\bar{y}}{\bar{y}_h} - 1 \right)} n \bar{y} \left(\frac{\bar{y}}{\bar{y}_h} - 1 \right) = n .$$

Finalmente, a estatística da razão de verossimilhança é calculada por

$$\omega = \frac{1}{\tilde{\theta}\phi^{(0)^2}} \sum_{i=1}^n \frac{(y_i - \tilde{\theta})^2}{y_i} + n \left\{ \log \left[\frac{\bar{y}\bar{y}_h\phi^{(0)^2}}{\tilde{\theta}(\bar{y} - \bar{y}_h)} \right] - 1 \right\} ,$$

onde $\tilde{\theta} = \left(\phi^{(0)^2} + \sqrt{\phi^{(0)^4} + 4\bar{y}/\bar{y}_h} \right) \bar{y}_h/2$. A estatística ω , acima, pode ser aperfeiçoada por $\omega^* = c^{-1}\omega$, que tem sob a hipótese nula e até ordem n^{-1} , distribuição χ_1^2 , onde $c = 1 + A_\phi + C_{\theta,\phi}$. As quantidades A_ϕ e $C_{\theta,\phi}$ são obtidas, respectivamente, de (5.16) e (5.17), com as derivadas das funções $\alpha_1, \alpha_2, \beta_1, \beta_2$ correspondentes à densidade (5.55'), tomadas em relação ao vetor $(\phi, \mu)^\top$. Utilizando o pacote MATHEMATICA, encontramos

$$A_\phi = \frac{8 + 12\phi^2 - 3\phi^4 - 6\phi^6}{3n(2 + \phi^2)^3}$$

e

$$C_{\theta,\phi} = \frac{12 + 14\phi^2 + 7\phi^4}{2n(2 + \phi^2)^2} ,$$

que resulta

$$c = 1 + \frac{1}{6n} \left[\frac{2(8 + 12\phi^2 - 3\phi^4 - 6\phi^6)}{(2 + \phi^2)^3} + \frac{3(12 + 14\phi^2 + 7\phi^4)}{(2 + \phi^2)^2} \right] .$$

Se temos, agora, interesse no parâmetro θ , sendo ϕ o parâmetro de perturbação, devemos obter a nova parametrização $(\theta, \gamma)^\top$, com θ e γ ortogonais, através da solução da equação diferencial (5.53), que é expressa por

$$\left(\frac{1}{\phi} \right) \frac{\partial \phi}{\partial \theta} = \frac{1}{2\theta} .$$

Uma possível solução desta equação é $a(\gamma) = \phi\theta^{-1/2}$, onde $a(\cdot)$ é uma função arbitrária. Considerando $a(\gamma) = \gamma$ temos os parâmetros globalmente ortogonais θ e $\gamma = \phi\theta^{-1/2}$. Com esta nova parametrização, escrevemos a função densidade em (5.55) por

$$\pi^*(y; \theta, \gamma) = \exp \left\{ -\frac{y}{2\theta^2\gamma^2} - \frac{1}{2y\gamma^2} + \frac{1}{\theta\gamma^2} - \frac{1}{2} \log \gamma^2 - \frac{1}{2} \log(2\pi y^3) \right\}, \quad (5.55'')$$

resultando $\alpha_1 = \alpha_1(\theta, \gamma) = -1/(2\theta^2\gamma^2)$, $\alpha_2 = \alpha_2(\theta, \gamma) = -1/(2\gamma^2)$, $\rho = \rho(\theta, \gamma) = (\log \gamma^2)/2 - 1/(\theta\gamma^2)$. Através de (5.21) encontramos $\beta_1 = \theta$ e $\beta_2 = (1 + \theta\gamma^2)/\theta$. Utilizando as fórmulas (5.51) e (5.52), com estas funções, e derivadas tomadas em relação ao vetor $(\theta, \gamma)^T$, é fácil verificarmos que $\kappa_{\theta\gamma} = \kappa_{\gamma\theta} = 0$, o que demonstra a ortogonalidade global entre θ e γ .

Para este teste, a correção de Bartlett é idêntica àquela da Seção 5.4.2, isto é, $c = 1 + 3/2n$, pela mesma justificativa apresentada para o teste do parâmetro θ na $N(\theta, \theta^2\phi^2)$, do Exemplo 5.1.

Exemplo 5.3 Consideremos uma amostra aleatória $y_1, \dots, y_n$ de uma distribuição gama (θ, ϕ) com média $\theta\phi$, sendo θ o parâmetro de escala e ϕ o parâmetro de forma, cuja função densidade é

$$\pi(y; \theta, \phi) = \frac{1}{\Gamma(\phi)} \theta^{-\phi} y^{\phi-1} \exp \left(\frac{-y}{\theta} \right), \quad y, \theta \text{ e } \phi > 0. \quad (5.56)$$

A função de log-verossimilhança total de θ e ϕ é

$$\ell = \ell(\theta, \phi) = -\frac{1}{\theta} \sum_{l=1}^n y_l + (\phi - 1) \sum_{l=1}^n \log y_l - n\phi \log \theta - n \log \Gamma(\phi),$$

sendo $\alpha_1 = -\theta^{-1}$, $\alpha_2 = \phi - 1$, $\rho = \phi \log \theta + \log \Gamma(\phi)$. Como $\alpha_1 = \alpha_1(\theta)$ e $\alpha_2 = \alpha_2(\phi)$, podemos utilizar (5.18) para obtermos $\beta_1 = \theta\phi$ e $\beta_2 = \log \theta + \psi(\phi)$. De (5.4) e (5.5) vem,

$$E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = \frac{-n\phi}{\theta^2}, \quad E \left(\frac{\partial^2 \ell}{\partial \phi^2} \right) = -n\psi'(\phi) \quad \text{e} \quad E \left(\frac{\partial^2 \ell}{\partial \theta \partial \phi} \right) = -\frac{n}{\theta}.$$

As estimativas de máxima verossimilhança de θ e ϕ , para a densidade (5.56), são, respectivamente, $\hat{\theta} = \bar{y}/\hat{\phi}$ e $\hat{\phi}$ como solução da equação não linear $\log \hat{\phi} - \psi(\hat{\phi}) = \log(\bar{y}/\bar{y}_g)$, esta última, como esperado, idêntica àquela da distribuição gama estudada na Seção 5.4.3.

Se temos interesse no parâmetro ϕ , obtemos a nova parametrização $(\mu, \phi)^\top$, com μ e ϕ ortogonais, solucionando-se a equação diferencial em (5.49), que é escrita por

$$\left(-\frac{1}{\theta}\right) \frac{\partial \theta}{\partial \phi} = \frac{1}{\phi}$$

Uma possível solução desta equação é $a(\mu) = \theta\phi$, onde $a(\cdot)$ é uma função arbitrária. Tomando $a(\mu) = \mu$, resulta $\mu = \theta\phi$ e ϕ parâmetros globalmente ortogonais. Com esta nova parametrização, a função densidade em (5.56) é escrita por

$$\begin{aligned} \pi^*(y; \mu, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^\phi y^{\phi-1} \exp\left(-\frac{\phi}{\mu}y\right) \\ &= \exp\left\{-\frac{\phi}{\mu}y + (\phi-1)\log y - \phi\log \mu + \phi\log \phi - \log \Gamma(\phi)\right\}, \end{aligned} \quad (5.56')$$

sendo $\alpha_1(\mu, \phi) = -\phi/\mu$, $\alpha_2(\mu, \phi) = \phi-1$, $\rho(\mu, \phi) = \phi\log \mu - \phi\log \phi + \log \Gamma(\phi)$. Como $\alpha_2(\mu, \phi) = \alpha_2(\phi)$, obtemos de (5.21) $\beta_1 = \mu$ e $\beta_2 = \log \mu - \log \phi + \psi(\phi)$. Substituindo as derivadas apropriadas em (5.5), encontramos $\kappa_{\mu\phi} = \kappa_{\phi\mu} = 0$. O teste do parâmetro ϕ , aqui, é idêntico ao teste de ϕ , com média desconhecida, discutido na Seção 5.4.3. Conforme mencionamos nos exemplos anteriores, este fato era esperado, pois em ambas situações, a hipótese nula é a mesma e o parâmetro de perturbação está irrestrito.

Quando o interesse é o parâmetro θ , então a reparametrização $(\theta, \gamma)^\top$, com θ e γ ortogonais, pode ser obtida solucionando-se a equação diferencial (5.53), expressa por

$$\psi'(\phi) \frac{\partial \phi}{\partial \theta} = -\frac{1}{\theta}.$$

Uma possível solução desta equação é $a(\gamma) = \psi(\phi) + \log \theta$, onde $a(\cdot)$ é uma função arbitrária. Tomando $a(\gamma) = \gamma$, resulta θ e $\gamma = \psi(\phi) + \log \theta$ parâmetros globalmente ortogonais. Com esta nova reparametrização, a função densidade em (5.56) é escrita por

$$\pi^*(y; \theta, \gamma) = \exp\left\{-\frac{y}{\theta} + h(\theta, \gamma)\log y - h(\theta, \gamma)\log \theta - \log y - \log \Gamma(h(\theta, \gamma))\right\}, \quad (5.56'')$$

onde $h(\theta, \gamma) = \psi^{-1}(\gamma - \log \theta)$. A densidade (5.56'') conduz às funções $\alpha_1 = \alpha_1(\theta, \gamma) = -1/\theta$, $\alpha_2 = \alpha_2(\theta, \gamma) = h(\theta, \gamma) - 1$ e $\rho = \rho(\theta, \gamma) = h(\theta, \gamma)\log \theta + \log \Gamma(h(\theta, \gamma))$. Utilizando estas funções nas expressões em (5.6), com derivadas tomadas em relação ao vetor $(\theta, \gamma)^\top$,

encontramos $\beta_1 = \theta \{h(\theta, \gamma) - h^{(1,0)}(\theta, \gamma)[\Gamma^{(1,0)}(h(\theta, \gamma)) - \Gamma^{(0,1)}(h(\theta, \gamma))]/\Gamma(h(\theta, \gamma))\}$ e $\beta_2 = \log \theta + \Gamma^{(0,1)}(h(\theta, \gamma))/\Gamma(h(\theta, \gamma))$, onde os sobrescritos (i, j) indicam derivadas de ordens i e j em relação a θ e γ , respectivamente. A ortogonalidade entre os parâmetros θ e γ é demonstrada através das expressões (5.51) e (5.52), com as funções acima e suas derivadas tomadas em relação ao vetor de parâmetros $(\theta, \gamma)^\top$. O procedimento para obtermos a estatística da razão de verossimilhança ω para este teste, e a correspondente correção de Bartlett é análogo aos exemplos anteriores. No entanto, omitimos aqui porque suas expressões finais são extremamente complicadas.

Exemplo 5.4 Seja $y_1, \dots, y_n$ uma amostra aleatória de tamanho n da distribuição beta (θ, ϕ) , cuja função densidade é escrita por

$$\begin{aligned} \pi(y; \theta, \phi) &= \frac{1}{B(\theta, \phi)} y^{(\theta-1)} (1-y)^{(\phi-1)}, \quad 0 < y < 1, \quad \theta, \phi > 0 \\ &= \exp\{\theta \log y + \phi \log(1-y) - \log B(\theta, \phi) - \log[y(1-y)]\}, \end{aligned} \quad (5.57)$$

onde $B(\cdot, \cdot)$ é a função matemática beta, sendo calculada por $B(z, \omega) = \int_0^1 t^{z-1} (1-t)^{\omega-1} dt$. Temos, então, $\alpha_1(\theta, \phi) = \theta$, $\alpha_2(\theta, \phi) = \phi$, $\rho(\theta, \phi) = \log B(\theta, \phi) = \log [\Gamma(\theta)\Gamma(\phi)/\Gamma(\theta + \phi)]$, $d_1(y) = y$, $d_2(y) = 1 - y$ e $v(y) = \log[y(1-y)]$. Para esta situação podemos aplicar os resultados da Seção 5.3 (caso 1). De (5.18) vem $\beta_1 = \psi(\theta) - \psi(\theta + \phi)$ e $\beta_2 = \psi(\phi) - \psi(\theta + \phi)$. Como $\alpha_1^{(0,1)} = 0 = \alpha_2^{(1,0)}$ e $\alpha_1^{(1,0)} = 1 = \alpha_2^{(0,1)}$ resulta, por (5.4), $\kappa_{\theta\theta} = n(\psi^{(1,0)}(\theta + \phi) - \psi'(\theta))$ e $\kappa_{\phi\phi} = n(\psi^{(0,1)}(\theta + \phi) - \psi'(\phi))$ e de (5.5) vem $\kappa_{\theta\phi} = n\psi^{(0,1)}(\theta + \phi) \neq 0$ e $\kappa_{\phi\theta} = n\psi^{(1,0)}(\theta + \phi) \neq 0$, implicando que θ e ϕ não são ortogonais. A definição de $\psi^{(i,j)}$ é análoga à de $\alpha_1^{(i,j)}$. Segundo Cox e Reid (1987), para famílias exponenciais regulares, é possível mostrar que $\mu = \beta_1 = \psi(\theta) - \psi(\theta + \phi)$ é ortogonal a ϕ quando temos interesse no parâmetro ϕ , e $\gamma = \beta_2 = \psi(\phi) - \psi(\theta + \phi)$ é ortogonal a θ , se nosso interesse é o parâmetro θ . Destes resultados, observamos que na parametrização $(\mu, \phi)^\top$ não é possível expressarmos θ como função de μ e ϕ , da mesma forma que não é possível expressarmos ϕ como função de θ e γ , na parametrização $(\theta, \gamma)^\top$. Assim, embora possamos encontrar os parâmetros ortogonais, associados à densidade (5.57), este caso não

tem interesse prático, porque não conseguimos obter as derivadas segundo a reparametrização ortogonal, quando θ ou ϕ é o parâmetro de interesse.

5.6 UM ESTUDO DE SIMULAÇÃO

Finalizando este capítulo, apresentamos um estudo de simulação com o intuito de comparar o desempenho da estatística da razão de verossimilhança usual com o da estatística modificada. Estudamos a distribuição gama com parâmetros θ e ϕ ortogonais, ou seja, com média $\theta > 0$ e variância θ^2/ϕ , $\phi > 0$ (vide Seção 5.4.3). Sua função densidade tem a forma

$$\pi(y; \theta, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi}{\theta}\right)^\phi y^{\phi-1} \exp\left(-\frac{\phi}{\theta}y\right), \quad y, \theta, \phi > 0. \quad (5.58)$$

Consideramos, primeiramente, o teste da média, isto é, $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$, com ϕ parâmetro de perturbação. A estimativa de máxima verossimilhança irrestrita de ϕ é $\hat{\phi}$ satisfazendo a equação (5.33). Seja $\tilde{\phi}$ a estimativa de máxima verossimilhança restrita de ϕ a $H : \theta = \theta^{(0)}$. Então, $\hat{\phi}$ segue de

$$\log \hat{\phi} - \psi(\hat{\phi}) = g, \quad (5.59)$$

onde $g = \log(\bar{y}/\bar{y}_g)$, sendo $\bar{y}_g$ a média geométrica dos dados, e $\tilde{\phi}$ segue de

$$\log \tilde{\phi} - \psi(\tilde{\phi}) = g', \quad (5.60)$$

onde $g' = \log(\theta^{(0)}/\bar{y}_g) + \bar{y}/\theta^{(0)} - 1$. Observe-se que (5.60) reduz-se a (5.59) quando $\theta^{(0)}$ é substituído por $\bar{y}$. Soluções aproximadas para $\hat{\phi}$ (ou $\tilde{\phi}$) vêm das fórmulas de Greenwood e Durand (1960), válidas para os intervalos $0 \leq g(g') \leq 0,5772$ e $0,5772 \leq g(g') \leq 17,0$. Uma outra solução aproximada para $\hat{\phi}$ é dada por $\hat{\phi} = \{1 + (1 + 4g/3n)^{1/2}\}/4g$ (Cordeiro, 1982). A solução aproximada para $\tilde{\phi}$ segue desta mesma equação com g' no lugar de g . A estatística da razão de verossimilhança, neste caso, é expressa por (5.35) e a correspondente correção de Bartlett por (5.36), com ϕ estimado segundo a hipótese nula, isto é, por $\tilde{\phi}$.

Para a realização do estudo de simulação é desenvolvido um programa FORTRAN para calcular as taxas de rejeição de ambas estatísticas, ω e ω^* , usando a biblioteca de sub-rotinas do programa NAG (“FORTRAN Mark 11 Library”). Detalhes do programa são descritos a seguir.

1. Geração da distribuição gama: se $Y \sim G(\theta, \phi)$, então $Y \sim \frac{\theta}{2\phi} \chi_{2\phi}^2$, isto é, os dados $y_1, \dots, y_n$ da distribuição gama são gerados a partir dos valores especificados para θ e ϕ , via a sub-rotina G05DHF, de geração da distribuição qui-quadrado.
2. A solução de (5.59) e (5.60) é obtida através da sub-rotina C05ADF, qua calcula a raiz de uma equação transcendental qualquer, via o algoritmo de Bus e Dekker (vide NAG Mini-Manual ou NAG FORTRAN Library Manual). A NAG FORTRAN Mark 11 Library tem 524 sub-rotinas acessíveis.
3. Cálculo das estatísticas ω e $\omega^* = \tilde{c}^{-1}\omega$ é feito de forma trivial com a obtenção da função gama $\Gamma(\cdot)$, via a sub-rotina S14AAF.
4. Cálculo da correção de Bartlett $\tilde{c}$ é feito através de macros especiais do GLIM (versão 3.77), para computador VAX, que produzem as funções poligamas.
5. Cálculo das taxas de rejeição de ω e $\omega^* = \tilde{c}^{-1}\omega$, onde $\tilde{c}$ é a estimativa de máxima verossimilhança de c , em (5.36), segundo H , isto é, substituindo ϕ por $\tilde{\phi}$, é realizado através da sub-rotina M01ANF, de ordenação de componentes de um vetor em ordem ascendente.

Apresentamos, agora, os resultados das simulações das taxas de rejeição $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, com ϕ variando em 0.5, 1.0, 2.0 e 5.0 e $n = 5, 10, 15$ e 20, onde x_α é o $(1 - \alpha)$ -ésimo percentil da distribuição χ_1^2 ($\alpha = 5\%$ e 10%). A média verdadeira $\theta^{(0)}$ da hipótese nula foi tomada igual a 4 em todas as simulações. As Tabelas 5.2 e 5.3 mostram as taxas de rejeição de ω e ω^* para o teste da hipótese $H : \theta = 4$, baseadas em 10.000 simulações, para cada combinação de ϕ , n e $\alpha = 5\%$ ($x_\alpha = 3,841$) e $\alpha = 10\%$ ($x_\alpha = 2,706$), respectivamente.

Tabela 5.2: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste da hipótese $H : \theta = 4$ no modelo gama, considerando nível de significância $\alpha = 5\%$.

n	ϕ		0,5		1,0		2,0		5,0	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5	8,5	7,1	8,4	7,2	8,6	7,1	8,5	6,9		
10	6,6	5,7	6,5	5,7	6,6	5,7	6,5	5,4		
15	6,0	5,5	6,0	5,4	6,2	5,4	6,2	5,3		
20	5,7	5,1	5,6	5,1	5,8	5,2	5,7	5,1		

Nota: Os números (1) e (2) correspondem a $P(\omega \geq 3,841)$ e $P(\omega^* \geq 3,841)$, respectivamente.

Tabela 5.3: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste da hipótese $H : \theta = 4$ no modelo gama, considerando nível de significância $\alpha = 10\%$.

n	ϕ		0,5		1,0		2,0		5,0	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5	14,4	12,1	14,1	12,2	14,4	12,0	14,3	11,8		
10	12,3	11,3	12,1	11,3	12,4	11,0	12,2	10,8		
15	11,6	11,0	11,4	10,8	11,9	10,7	11,6	10,5		
20	11,1	10,7	10,8	10,4	10,8	10,4	10,7	10,2		

Nota: Os números (1) e (2) correspondem a $P(\omega \geq 2,706)$ e $P(\omega^* \geq 2,706)$, respectivamente.

Os resultados destas tabelas mostram que, para os valores considerados de ϕ no intervalo de 0,5 a 5,0, as taxas de rejeição de ω e ω^* , praticamente independem de ϕ . Observa-se, ainda, que a distribuição de ω^* é melhor aproximada por χ_1^2 do que aquela de ω , pois todas as taxas de rejeição de ω^* estão mais próximas dos correspondentes níveis nominais do que aquelas de ω . Para $\phi = 5$, em ambas tabelas, a distribuição de ω^* parece ser um pouco melhor aproximada por χ_1^2 do que quando $\phi = 0,5, 1,0$ ou $2,0$. Claramente, a correção de Bartlett, em qualquer caso, torna-se menos necessária quando n cresce, pois ω aproxima-se mais de χ_1^2 .

Para o teste do parâmetro de forma, $H : \phi = \phi^{(0)}$ versus $A : \phi \neq \phi^{(0)}$, a estimativa de máxima verossimilhança irrestrita $\hat{\phi}$ de ϕ segue de (5.59) e a estatística da razão de verossimilhança ω , para este teste, é dada por (5.37). A correspondente correção de Bartlett para melhorar ω é calculada por (5.39). Nosso estudo, como antes, visa a comparar as taxas de rejeição de ω e ω^* , $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$, respectivamente. Para isso, simulamos o modelo gama 10.000 vezes para cada combinação de $\phi^{(0)} = 0,5, 1,0, 2,0$ e $5,0$ e $n = 5, 10, 15$ e 20 , tomando o valor verdadeiro de θ igual a 5 . O procedimento da simulação é o mesmo descrito no teste anterior, com a diferença de que as correções dos passos 4 e 5 são calculadas de (5.39) e avaliadas em $\phi^{(0)}$. As taxas de rejeição são, então, computadas, como as proporções de vezes que os valores simulados de ω e ω^* superam os pontos críticos $x_{0,05}$ e $x_{0,10}$ da distribuição χ_1^2 . Os resultados estão apresentados nas Tabelas 5.4 e 5.5, para níveis nominais de 5% e 10%, respectivamente.

Tabela 5.4: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste da hipótese $H : \phi = \phi^{(0)}$ no modelo gama com $\theta = 5$, considerando nível de significância $\alpha = 5\%$.

$\phi^{(0)}$	0,5		1,0		2,0		5,0	
n	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5	8,7	7,3	9,0	7,6	9,3	7,7	9,4	7,8
10	6,9	6,0	7,0	6,0	7,1	6,2	7,4	6,5
15	6,3	5,7	6,4	5,9	6,5	5,9	6,8	6,4
20	5,9	5,3	5,9	5,3	6,1	5,6	6,2	5,8

Nota: Os números (1) e (2) correspondem a $P(\omega \geq 3,841)$ e $P(\omega^* \geq 3,841)$, respectivamente.

Tabela 5.5: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem), para o teste da hipótese $H : \phi = \phi^{(0)}$ no modelo gama com $\theta = 5$, considerando nível de significância $\alpha = 10\%$.

$\phi^{(0)}$	0,5		1,0		2,0		5,0	
n	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
5	14,9	12,4	15,4	12,8	15,9	12,9	15,8	12,8
10	12,7	11,4	12,9	11,7	12,9	11,8	13,0	11,6
15	11,7	10,6	11,9	10,7	11,9	10,5	12,0	10,5
20	10,8	10,2	10,9	10,4	11,2	10,4	11,2	10,2

Nota: Os números (1) e (2) correspondem a $P(\omega \geq 2,706)$ e $P(\omega^* \geq 2,706)$, respectivamente.

Analisando os valores destas tabelas, pode-se concluir que o desempenho da estatística corrigida ω^* é superior ao da estatística ω usual, pois em todos os casos considerados as taxas de rejeição de ω^* estão mais próximas do correspondentes níveis nominais do que aquelas de ω , particularmente, para tamanhos de amostra pequenos ($n = 5$ ou 10). Observa-se, ainda, tal como para o teste da média θ que, para ϕ no intervalo considerado, as taxas de rejeição de ω e ω^* apresentam uma certa estabilidade, isto é, praticamente independem do valor de ϕ . Como esperado, a correção de Bartlett, em qualquer caso, torna-se menos necessária quando n cresce, já que ω aproxima-se mais de χ_1^2 , embora a aproximação para ω^* , mesmo nestes casos, seja mais rápida.

CAPÍTULO 6

CORREÇÕES DE BARTLETT PARA MODELOS NORMAIS HETEROCEDÁSTICOS

6.1 INTRODUÇÃO

Definimos, no Capítulo 2, a classe dos modelos normais lineares, cuja estrutura de covariância dos erros é não-escalar e dependente de um conjunto de parâmetros desconhecidos. Encontramos expressões matriciais bastante simples para as correções de Bartlett, associadas a alguns testes de hipóteses sobre os parâmetros em modelos especiais.

Em muitas situações práticas, é bastante comum supormos que os erros aleatórios são não-correlacionados com variâncias quaisquer. Neste capítulo, nosso objetivo é estudar, então, os modelos normais lineares para os quais existe independência entre os erros, ou, equivalentemente, entre as observações, ou seja, trabalhamos com os modelos normais heterocedásticos,

sendo as variâncias dependentes, através de uma função arbitrária, de um conjunto de parâmetros desconhecidos e as covariâncias nulas. Esta classe de modelos é, portanto, uma generalização daquela estudada por Cordeiro (1993b) referente aos modelos normais heterocedásticos multiplicativos e um caso especial importante dos modelos descritos no Capítulo 2. É importante destacar que a razão principal de estudarmos este modelo a parte é devido à sua grande aplicabilidade em econometria e, principalmente, pelo desconhecimento dos econometristas nas correções de Bartlett. Obter estes resultados diretamente daqueles gerais do Capítulo 2 é muito difícil por causa da diferença de notação. Devemos, ainda, mencionar que no contexto dos modelos lineares generalizados com componentes sistemática para o parâmetro de dispersão, este estudo está desenvolvido na tese de doutorado de D.A. Botter (1993).

Seja o modelo normal heterocedástico para o vetor de observações aleatórias $Y(n \times 1)$, definido pela componente sistemática

$$\begin{aligned} E(Y) &= \mu = X\beta, \\ \sigma_l^2 &= \text{Var}(Y_l) = H(\omega_l^\top \gamma), \end{aligned} \quad (6.1)$$

para $l = 1, \dots, n$, sendo $X(n \times p)$, $\beta(p \times 1)$, $\gamma(q \times 1)$ e $\omega_l^\top = (\omega_{l1}, \dots, \omega_{lq})(1 \times q)$ definidos tal como na Seção 2.2, e $H(\cdot)$ uma função monótona e diferenciável até terceira ordem. Seja $W = (\omega_1, \dots, \omega_n)^\top$ uma matriz $n \times q$ de constantes conhecidas. Os vetores de parâmetros β e γ não têm, em geral, elementos em comum. Em alguns casos especiais, uma coluna de W pode ser uma função conhecida de algumas colunas da matriz modelo X . Supomos, então, que o vetor aleatório $Y(n \times 1)$ segue uma distribuição normal n -variada com vetor média $\mu = X\beta$ e matriz de covariância $\Lambda = \Lambda(\gamma) = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$, com $\sigma_l^2 = H(\omega_l^\top \gamma)$ para $l = 1, \dots, n$.

A função de variância em (6.1) tem grande aplicabilidade em vários campos de pesquisa, tais como, engenharia, economia, ciências físicas e biológicas, entre outros. Um caso especial de interesse corresponde a tomarmos $\omega_l^\top = (1, \log z_{l2}, \dots, \log z_{lq})$, onde z_{lj} são constantes conhecidas estritamente positivas, e $H(\cdot) = \exp(\cdot)$, que resulta $\sigma_l^2 = \sigma^2 \prod_{j=2}^q z_{lj}^{\gamma_j}$,

$l = 1, \dots, n$, com $\sigma^2 = \exp(\gamma_1)$. Assim, o teste relativo a homocedasticidade é equivalente a testar $\gamma_2 = \dots = \gamma_q = 0$. Testes similares para homocedasticidade foram propostos por Goldfeld e Quandt (1965), Glejser (1969), Harrison e McCabe (1979), Jarque (1981), Koenker e Bassett (1982) e Cook e Weisberg (1983).

Na Seção 6.2 apresentamos uma expressão matricial para a quantidade ε_{p+q} . Esta quantidade é examinada, na Seção 6.3, para casos especiais do modelo (6.1). Diversos testes da razão de verossimilhança são apresentados na Seção 6.4. Finalizando este capítulo, apresentamos na Seção 6.5 um pequeno estudo de simulação.

6.2 OBTENÇÃO DA QUANTIDADE ε_{p+q}

Para o modelo normal heterocedástico em (6.1), o logaritmo da função de verossimilhança total para $\theta = (\beta^\top, \gamma^\top)^\top$, dada as observações $y_1, \dots, y_n$, pode ser escrita como

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Lambda| - \frac{1}{2} \mathbf{1}_n^\top \Lambda^{-1} v, \quad (6.2)$$

onde $\mathbf{1}_n$ é um vetor $n \times 1$ de uns, $v = (v_1, \dots, v_n)^\top$ é um vetor $n \times 1$, com $v_l = (y_l - x_l^\top \beta)^2$, $l = 1, \dots, n$, $\log |\Lambda| = \sum_{l=1}^n \log H(\omega_l^\top \gamma)$ e $\mathbf{1}_n^\top \Lambda^{-1} v = \sum_{l=1}^n v_l / H(\omega_l^\top \gamma)$. Por simplicidade, denotamos $\rho_l = \omega_l^\top \gamma$, $\sigma_l^2 = H(\omega_l^\top \gamma) = H(\rho_l) = H_l$ e $H_l^{(i)} = H^{(i)}(\omega_l^\top \gamma) = \partial^i H(\rho_l) / \partial \rho_l^i$, $i = 1, 2, \dots$. Assim, (6.2) pode ser reescrita por

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{l=1}^n \log H_l - \frac{1}{2} \sum_{l=1}^n \frac{v_l}{H_l}. \quad (6.3)$$

A função $\ell(\theta)$ em (6.3) é regular com respeito às derivadas de β e γ até e incluindo as de quarta ordem. Os estimadores de máxima verossimilhança são obtidos resolvendo-se o sistema de equações formado igualando as derivadas de $\ell(\theta)$ com relação a β e γ a zero. Temos, então,

$$\frac{\partial \ell(\theta)}{\partial \beta} = -\frac{1}{2} \sum_{l=1}^n \frac{-2x_l(y_l - x_l^\top \beta)}{H_l} = X^\top \Lambda^{-1} y - (X^\top \Lambda^{-1} X) \beta$$

e

$$\frac{\partial \ell(\theta)}{\partial \gamma} = -\frac{1}{2} \sum_{l=1}^n \frac{H_l^{(1)}}{H_l} \omega_l + \frac{1}{2} \sum_{l=1}^n v_l \frac{H_l^{(1)}}{H_l^2} \omega_l = -\frac{1}{2} W^\top P 1_n + \frac{1}{2} W^\top P \Lambda^{-1} v ,$$

resultando nas equações de máxima verossimilhança,

$$\hat{\beta} = (X^\top \hat{\Lambda}^{-1} X)^{-1} X^\top \hat{\Lambda}^{-1} y \quad \text{e} \quad W^\top \hat{P} 1_n = W^\top \hat{P} \hat{\Lambda}^{-1} \hat{v} , \quad (6.4)$$

com $P = \text{diag}\{p_{ii}\} = \Lambda^{(1)} \Lambda^{-1}$, $\Lambda = \text{diag}\{H_1, \dots, H_n\}$ e $\Lambda^{(r)} = \text{diag}\{H_1^{(r)}, \dots, H_n^{(r)}\}$, para $r = 1, 2, \dots$. Os estimadores $\hat{\beta}$ e $\hat{\gamma}$ podem ser calculados, iterativamente, a partir das equações em (6.4).

Adotamos a mesma notação introduzida na Seção 2.3.1 para os cumulantes de derivadas do logaritmo da função de verossimilhança total e derivadas destes cumulantes, em relação aos componentes de β e γ .

Seja $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$ o estimador de máxima verossimilhança irrestrito de θ . Definimos o logaritmo da função de verossimilhança total maximizada em $\hat{\theta}$ por $\hat{\ell}_{p+q} = \ell(\hat{\theta})$ e o logaritmo da função de verossimilhança no valor verdadeiro do parâmetro θ por $\ell = \ell(\theta)$. Da expansão de Lawley (1956) temos, até ordem n^{-1} , que

$$2E(\hat{\ell}_{p+q} - \ell) = p + q + \varepsilon_{p+q} , \quad (6.5)$$

onde o termo ε_{p+q} , de ordem n^{-1} , é avaliado no valor verdadeiro do parâmetro e sob a hipótese nula, e expresso por

$$\varepsilon_{p+q} = \sum'_{\beta, \gamma} (\ell_{rstu} - \ell_{rstuvw}) . \quad (6.6)$$

Aqui, como no Capítulo 2, $\sum'_{\beta, \gamma}$ indica o somatório sobre todas as combinações dos p parâmetros em β e q parâmetros em γ , $\sum$ o somatório sobre os dados amostrais e os ℓ 's são obtidos de

$$\ell_{rstu} = \kappa^{rs} \kappa^{tu} \left(\frac{1}{4} \kappa_{rstu} - \kappa_{rt}^{(u)} + \kappa_{rt}^{(su)} \right) \quad (6.7)$$

e

$$\ell_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left[\kappa_{rtv} \left(\frac{1}{6} \kappa_{suw} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{1}{4} \kappa_{svw} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right] . \quad (6.8)$$

Vamos desenvolver o termo ε_{p+q} em (6.6) numa forma mais simplificada. Estamos denotando índices com letras minúsculas para representar componentes do vetor de parâmetros β e índices com letras maiúsculas para componentes do vetor de parâmetros γ . Diferenciando (6.3) e calculando valores esperados, encontramos os cumulantes conjuntos de derivadas de $\ell(\theta)$, em relação aos componentes de β e γ (vide Apêndice 3). A partição $\theta^\top = (\beta^\top, \gamma^\top)$ induz à correspondente partição na matriz de informação de Fisher bloco diagonal, pois

$$\begin{aligned} E_y \left[-\frac{\partial^2 \ell(\theta)}{\partial \gamma \partial \beta} \right] &= E_y \left[\frac{1}{2} \sum_{l=1}^n 2x_l (y_l - x_l^\top \beta) \omega_l^\top \frac{H_l^{(1)}}{H_l^2} \right] = 0, \\ E_y \left[-\frac{\partial^2 \ell(\theta)}{\partial \beta^2} \right] &= E_y [X^\top \Lambda^{-1} X] = X^\top \Lambda^{-1} X \end{aligned}$$

e

$$\begin{aligned} E_y \left[-\frac{\partial^2 \ell(\theta)}{\partial \gamma^2} \right] &= E_y \left[-\frac{1}{2} \sum_{l=1}^n \left(\frac{H_l^{(2)}}{H_l} - \frac{(H_l^{(1)})^2}{H_l^2} - \left\{ \frac{H_l^{(2)}}{H_l^2} - \frac{2(H_l^{(1)})^2}{H_l^3} \right\} v_l \right) \omega_l \omega_l^\top \right] \\ &= \frac{1}{2} \sum_{l=1}^n \left(\frac{H_l^{(1)}}{H_l} \right)^2 \omega_l \omega_l^\top = \frac{1}{2} W^\top P^2 W, \end{aligned}$$

onde $P^2 = P.P$ e $P = \Lambda^{(1)} \Lambda^{-1} = \text{diag} \left\{ \frac{H_1^{(1)}}{H_1}, \dots, \frac{H_n^{(1)}}{H_n} \right\}$. Portanto, a matriz de informação total de Fisher, dividida em β e γ , é

$$K(\theta) = \begin{array}{c} \beta \qquad \qquad \gamma \\ \begin{array}{cc} \beta & \gamma \end{array} \left[\begin{array}{cc} X^\top \Lambda^{-1} X & O_{p \times q} \\ O_{q \times p} & \frac{1}{2} W^\top P^2 W \end{array} \right], \end{array}$$

onde $K_{\beta,\beta} = X^\top \Lambda^{-1} X (p \times p)$ é a informação para β , $K_{\gamma,\gamma} = \frac{1}{2} W^\top P^2 W (q \times q)$ é a informação para γ e $O_{p \times q}$ e $O_{q \times p}$ são matrizes nulas, implicando que os vetores de parâmetros β e γ são globalmente ortogonais (Cox e Reid, 1987). Os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\gamma}$ são, então, assintoticamente independentes, devido à sua normalidade assintótica e à estrutura bloco diagonal da matriz de informação de Fisher.

A propriedade de diagonalidade em bloco da matriz de informação permite-nos obter considerável simplificação na fórmula final de ε_{p+q} . Em vista disto, as somas $\sum' \ell_{rstu}$ e $\sum' \ell_{rstuvw}$

podem ser obtidas em notação matricial somando-se sobre os dados depois de avaliarmos as somas sobre os parâmetros (vide Apêndice 3). Definimos as matrizes de dimensão $n \times n$ $A = W(W^\top P^2 W)^{-1} W^\top = 2W K_{\gamma, \gamma}^{-1} W^\top$ e $B = X(X^\top \Lambda^{-1} X)^{-1} X^\top = X K_{\beta, \beta}^{-1} X^\top$ e as matrizes diagonais, com elementos correspondentes, respectivamente, de A e B , $A_d = \text{diag}\{a_{11}, \dots, a_{nn}\}$ e $B_d = \text{diag}\{b_{11}, \dots, b_{nn}\}$. As matrizes $2A$ e B têm interpretações simples como estruturas de covariância assintótica de $W\hat{\gamma}$ e $\hat{\mu}$, respectivamente. Substituindo-se os cumulantes e derivadas de cumulantes nas expressões em (6.7) e (6.8) e efetuando-se os cálculos necessários, obtemos, após manipulação algébrica bastante trabalhosa (vide Apêndice 3),

$$\sum_{\beta, \gamma}' \ell_{rstu} = \text{tr} \left[A_d B_d \Lambda^{-1} (2P^2 - Q) - \frac{1}{2} A_d^2 (-Q^2 + 8P^2 Q - 6P^4) \right] \quad (6.9)$$

e

$$\begin{aligned} \sum_{\beta, \gamma}' \ell_{rstuvw} = & -1_n^\top P A_d A B_d \Lambda^{-1} P 1_n - \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n \\ & + 1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n - \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n \\ & - \frac{1}{3} 1_n^\top P [4P^2 A^{(3)} (3Q - 2P^2) - 3Q A^{(3)} Q] P 1_n, \end{aligned} \quad (6.10)$$

onde $B^{(2)} = B \otimes B$, $A^{(3)} = A^{(2)} \otimes A$, $Q = \Lambda^{(2)} \Lambda^{-1}$ e $\otimes$ simboliza produto direto, ou seja, produto de elementos correspondentes, por exemplo, $B^{(2)} = \{b_{ij}^2\}$, $A^{(3)} = \{a_{ij}^3\}$, e assim por diante.

Finalmente, o termo ε_{p+q} é dado pela diferença das expressões (6.9) e (6.10). Observando a expressão resultante notamos que ela depende da matriz modelo X , da matriz de constantes W e da matriz de covariâncias Λ . Ela não depende do vetor de parâmetros β . Assim, obtemos a seguinte decomposição para ε_{p+q} :

$$\varepsilon_{p+q} = h(X, W, \Lambda) = f(W, \Lambda) + g(X, W, \Lambda), \quad (6.11)$$

onde

$$\begin{aligned} f(W, \Lambda) = & -\frac{1}{2} \text{tr} \left[A_d^2 (-Q^2 + 8P^2 Q - 6P^4) \right] + \frac{1}{3} 1_n^\top P [4P^2 A^{(3)} (3Q - 2P^2) \\ & - 3Q A^{(3)} Q] P 1_n + \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n \end{aligned} \quad (6.12)$$

e

$$\begin{aligned}
g(X, W, \Lambda) = & \operatorname{tr} \left[A_d B_d \Lambda^{-1} (2P^2 - Q) \right] - 1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n \\
& + \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n + 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n .
\end{aligned} \tag{6.13}$$

Embora as expressões acima apresentem um aspecto bastante complicado, elas são obtidas com facilidade através do uso de pacotes computacionais algébricos, que realizam operações com vetores e matrizes, tais como o REDUCE (Rayna, 1987), o MAPLE (Abell e Braselton, 1994) ou o MATHEMATICA (Wolfram, 1991). Entretanto, a facilidade em calculá-las é acompanhada pela dificuldade em interpretá-las. Fundamentalmente, a dificuldade de interpretação está no fato que os termos envolvidos nestas expressões não são invariantes sob reparametrização, e suas interpretações dependem da escolha do sistema de coordenadas proposto para os parâmetros.

Para modelos lineares normais heterocedásticos com forma fechada para as matrizes de covariância assintótica $2A$ e B , podemos obter, facilmente, uma forma fechada para a quantidade $\varepsilon_{p+q} = h(X, W, \Lambda)$. Apresentamos, a seguir, o cálculo da quantidade ε_{p+q} , dado em (6.11) – (6.13), para alguns casos especiais, os quais produzem simplificações consideráveis.

6.3 ALGUNS CASOS ESPECIAIS

Examinamos, nesta seção, casos especiais do modelo (6.1), que originam simplificações nas expressões (6.11) – (6.13) e as verificam parcialmente.

1º caso: $\sigma_l^2 = \operatorname{Var}(y_l) = \exp(\omega_l^\top \gamma)$, $\ell = 1, 2, \dots, n$.

Nesta situação temos,

$$H_l = H(\omega_l^\top \gamma) = H(\rho_l) = \exp(\rho_l) ,$$

e, portanto, $H_l = H_l^{(1)} = H_l^{(2)}$, implicando que $p_{ll} = q_{ll} = 1$ e $P = Q = I_n$. Substituindo em

(6.11) – (6.13) obtemos

$$\begin{aligned} \varepsilon_{p+q} = & \operatorname{tr}(A_d B_d \Lambda^{-1} - \frac{1}{2} A_d^2) + 1_n^\top A_d A B_d \Lambda^{-1} 1_n + \frac{1}{2} 1_n^\top \Lambda^{-1} B_d A B_d \Lambda^{-1} 1_n \\ & - 1_n^\top \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} 1_n + \frac{1}{2} 1_n^\top A_d A A_d 1_n + \frac{1}{3} 1_n^\top A^{(3)} 1_n, \end{aligned} \quad (6.14)$$

sendo

$$f(W, \Lambda) = -\frac{1}{2} \operatorname{tr}(A_d^2) + \frac{1}{3} 1_n^\top A^{(3)} 1_n + \frac{1}{2} 1_n^\top A_d A A_d 1_n \quad (6.15)$$

e

$$\begin{aligned} g(X, W, \Lambda) = & \operatorname{tr}(A_d B_d \Lambda^{-1}) - 1_n^\top \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} 1_n \\ & + \frac{1}{2} 1_n^\top \Lambda^{-1} B_d A B_d \Lambda^{-1} 1_n + 1_n^\top A_d A B_d \Lambda^{-1} 1_n. \end{aligned} \quad (6.16)$$

As fórmulas (6.14) – (6.16) coincidem com as expressões (3.6) – (3.8) de Cordeiro (1993b), que considera esta situação especial relativa ao modelo normal heterocedástico multiplicativo. Isto verifica parcialmente a exatidão da fórmula ε_{p+q} , dada por (6.11) – (6.13).

2º caso: $q = 1$, $\Lambda = \sigma^2 I_n$ com $W = 1_n$.

Temos, $\Lambda = \sigma^2 I_n$, $P = \tau I_n$, $Q = \delta I_n$, onde $\sigma^2 = H(\gamma_1)$, $\tau = H^{(1)}(\gamma_1)/H(\gamma_1)$ e $\delta = H^{(2)}(\gamma_1)/H(\gamma_1)$. Portanto,

$$A = \frac{1}{n\tau^2} 1_n 1_n^\top \quad \text{e} \quad B = \sigma^2 X(X^\top X)^{-1} X^\top.$$

Apresentamos, a seguir, a obtenção dos sete termos envolvidos na expressão de $\varepsilon_{p+1} = h(X, 1_n, \sigma^2 I_n)$:

$$(1) \operatorname{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q)] = \frac{1}{n\tau^2 \sigma^2} (2\tau^2 - \delta) \operatorname{tr}(B_d) = \frac{2\tau^2 - \delta}{n\tau^2 \sigma^2} \sigma^2 p = \frac{2p}{n} - \frac{\delta p}{n\tau^2},$$

$$\text{pois } \operatorname{tr}(B_d) = \operatorname{tr}(B) = \sigma^2 \operatorname{tr}(X(X^\top X)^{-1} X^\top) = \sigma^2 p;$$

$$(2) -\frac{1}{2} \operatorname{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)] = -\frac{1}{2} \frac{(-\delta^2 + 8\tau^2 \delta - 6\tau^4)}{n^2 \tau^4} \operatorname{tr}(I_n) = \frac{\delta^2}{2n\tau^4} - \frac{4\delta}{n\tau^2} + \frac{3}{n};$$

$$(3) \quad 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n = \frac{\tau^2}{\sigma^2 n^2 \tau^4} 1_n^\top 1_n 1_n^\top B_d 1_n = \frac{n \sigma^2 p}{\sigma^2 n^2 \tau^2} = \frac{p}{n \tau^2},$$

$$\text{pois } 1_n^\top B_d 1_n = \text{tr}(B_d 1_n 1_n^\top) = \text{tr}(B_d) = \text{tr} B = \sigma^2 p;$$

$$(4) \quad \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n = \frac{1}{2} \frac{\tau^2}{\sigma^4 n \tau^2} 1_n^\top B_d 1_n 1_n^\top B_d 1_n = \frac{\sigma^4 p^2}{2 \sigma^4 n} = \frac{p^2}{2n};$$

$$(5) \quad -1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n = -\frac{\tau^2}{\sigma^4 n \tau^2} 1_n^\top B^{(2)} 1_n = -\frac{\sigma^4 p}{\sigma^4 n} = -\frac{p}{n},$$

$$\text{pois } 1_n^\top B^{(2)} 1_n = \sigma^4 p;$$

$$(6) \quad \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n = \frac{1}{2} \tau^2 \frac{\delta^2}{n^3 \tau^6} 1_n^\top 1_n 1_n^\top 1_n = \frac{\tau^2 \delta^2}{2 n^3 \tau^6} n^2 = \frac{\delta^2}{2 n \tau^4};$$

$$(7) \quad \frac{1}{3} 1_n^\top P (12 P^2 A^{(3)} Q - 3 Q A^{(3)} Q - 8 P^2 A^{(3)} P^2) P 1_n = \frac{1}{3} \frac{\tau^2 (12 \tau^2 \delta - 3 \delta^2 - 8 \tau^4)}{n^3 \tau^6} 1_n^\top 1_n 1_n^\top 1_n \\ = \left(\frac{4 \delta}{n^3 \tau^2} - \frac{\delta^2}{n^3 \tau^4} - \frac{8}{3 n^3} \right) n^2 = \frac{4 \delta}{n \tau^2} - \frac{\delta^2}{n \tau^4} - \frac{8}{3 n}.$$

Portanto, somando os termos (1) – (7) acima, e efetuando simplificações, encontramos

$$\varepsilon_{p+1} = \frac{3(p+1)^2 - 1}{6n} - \frac{(\delta - 1)p}{n \tau^2}. \quad (6.17)$$

Quando $H(\cdot) = \exp(\cdot)$, temos o modelo heterocedástico multiplicativo e vale $H(\gamma_1) = H^{(1)}(\gamma_1) = H^{(2)}(\gamma_1)$, que produz $\delta = H^{(2)}(\gamma_1)/H(\gamma_1) = 1$, resultando $\varepsilon_{p+1} = [3(p+1)^2 - 1]/6n$. Esta última forma corresponde ao último termo da fórmula (4.1) de Cordeiro (1993b).

3º caso: $p = 1$ e $X = 1_n$.

Neste caso, temos: $X = 1_n$, $A = W(W^\top P^2 W)^{-1} W^\top$, $B = E_n / \sum_{l=1}^n \sigma_l^{-2}$, onde $E_n = 1_n 1_n^\top$, resultando, portanto, os seguintes termos de ε_{1+q} , após simplificações:

$$(1) \quad \text{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q)] = \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \text{tr}[A_d \Lambda^{-1} (2P^2 - Q)];$$

$$(2) \quad -\frac{1}{2} \text{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)];$$

$$(3) \quad 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n = \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} 1_n^\top P A_d A \Lambda^{-1} P 1_n;$$

$$(4) \quad \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n = \frac{1}{2 \left(\sum_{l=1}^n \sigma_l^{-2} \right)^2} 1_n^\top P \Lambda^{-1} A \Lambda^{-1} P 1_n ;$$

$$(5) \quad -1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n = -\frac{1}{\left(\sum_{l=1}^n \sigma_l^{-2} \right)^2} 1_n^\top P \Lambda^{-1} A \Lambda^{-1} P 1_n ;$$

$$(6) \quad \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n ;$$

$$(7) \quad \frac{1}{3} 1_n^\top P (4P^2 A^{(3)} (3Q - 2P^2) - 3Q A^{(3)} Q) P 1_n .$$

A soma dos sete termos acima produz,

$$\begin{aligned} \varepsilon_{1+q} = & \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \text{tr}[A_d \Lambda^{-1} (2P^2 - Q)] - \frac{1}{2} \text{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)] \\ & + \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} 1_n^\top P A_d A \Lambda^{-1} P 1_n - \frac{1}{2 \left(\sum_{l=1}^n \sigma_l^{-2} \right)^2} 1_n^\top P \Lambda^{-1} A \Lambda^{-1} P 1_n \\ & + \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n + \frac{1}{3} 1_n^\top P [4P^2 A^{(3)} (3Q - 2P^2) - 3Q A^{(3)} Q] P 1_n. \quad (6.18) \end{aligned}$$

No caso do modelo heterocedástico multiplicativo, ou seja, $H(\cdot) = \exp(\cdot)$, que implica $P = Q = I_n$, a última expressão reduz-se a

$$\begin{aligned} \varepsilon_{1+q} = & \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \text{tr}(A_d \Lambda^{-1}) - \frac{1}{2} \text{tr}(A_d^2) + \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} 1_n^\top A_d A \Lambda^{-1} 1_n \\ & - \frac{1}{2 \left(\sum_{l=1}^n \sigma_l^{-2} \right)^2} 1_n^\top \Lambda^{-1} A \Lambda^{-1} 1_n + \frac{1}{2} 1_n^\top A_d A A_d 1_n + \frac{1}{n} 1_n^\top A^{(3)} 1_n . \end{aligned}$$

Não é difícil provarmos que esta expressão coincide com as expressões (3.6) - (3.8) de Cordeiro (1993b), bastando considerar naquelas $B = E_n / \sum_{l=1}^n \sigma_l^{-2}$, que implica $B_d = I_n / \sum_{l=1}^n \sigma_l^{-2}$ e $B^{(2)} \otimes A = A / \left(\sum_{l=1}^n \sigma_l^{-2} \right)^2$.

4º caso: Modelo para classificação a um fator

Vamos considerar que variáveis aleatórias independentes foram tomadas de distribuições $N(\mu_1, \sigma_1^2), \dots, N(\mu_p, \sigma_p^2)$, em conjuntos de tamanhos $n_1, \dots, n_p$, respectivamente, com $n_j \geq 1$, $j = 1, \dots, p$ e $\sum_{j=1}^p n_j = n$. Podemos escrever o vetor de observações $y = (y_{11}, \dots, y_{1n_1}, \dots,$

$y_{p1}, \dots, y_{pn_p})^\top$ e definir a componente sistemática do modelo por

$$\mu_j = \beta + \beta_j \quad \text{e} \quad \sigma_j^2 = H(\gamma + \gamma_j), \quad j = 1, \dots, p,$$

com $\sum_{j=1}^p \beta_j = \sum_{j=1}^p \gamma_j = 0$, onde β e γ são escalares que representam médias gerais e β_j e γ_j são os efeitos na média e na variância da variável resposta, provenientes da j -ésima população, respectivamente.

Notamos que, neste caso,

$$W^\top P^2 W (p \times p) = \text{diag}\{n_1 p_{11}^2, \dots, n_p p_{pp}^2\}$$

e

$$X^\top \Lambda^{-1} X (p \times p) = \text{diag}\{n_1 \sigma^{-2}, \dots, n_p \sigma_p^{-2}\}.$$

Então, as matrizes A e B de ordem n têm elementos típicos dados por

$$A = \{a_{lm}\}, \text{ com } a_{lm} = \frac{\delta_{lm}}{n_l p_{ll}^2} \quad \text{e} \quad B = \{b_{lm}\}, \text{ com } b_{lm} = \frac{\delta_{lm}}{n_l \sigma_l^2},$$

onde

$$\delta_{lm} = \begin{cases} 1 & , \text{ se os índices } l \text{ e } m \text{ variam dentro da mesma população} \\ 0 & , \text{ caso contrário.} \end{cases}$$

Portanto,

$$A = \begin{pmatrix} \begin{pmatrix} n_1^{-1} p_{11}^{-2} & \cdots & n_1^{-1} p_{11}^{-2} \\ \vdots & & \vdots \\ n_1^{-1} p_{11}^{-2} & \cdots & n_1^{-1} p_{11}^{-2} \end{pmatrix}_{n_1 \times n_1} & \cdots & 0 \\ & \ddots & \\ 0 & & \begin{pmatrix} n_p^{-1} p_{pp}^{-2} & \cdots & n_p^{-1} p_{pp}^{-2} \\ \vdots & & \vdots \\ n_p^{-1} p_{pp}^{-2} & \cdots & n_p^{-1} p_{pp}^{-2} \end{pmatrix}_{n_p \times n_p} \end{pmatrix},$$

$$B = \begin{pmatrix} \begin{pmatrix} n_1^{-1} \sigma_1^2 & \cdots & n_1^{-1} \sigma_1^2 \\ \vdots & & \vdots \\ n_1^{-1} \sigma_1^2 & \cdots & n_1^{-1} \sigma_1^2 \end{pmatrix}_{n_1 \times n_1} & \cdots & 0 \\ & \ddots & \\ 0 & & \begin{pmatrix} n_p^{-1} \sigma_p^2 & \cdots & n_p^{-1} \sigma_p^2 \\ \vdots & & \vdots \\ n_p^{-1} \sigma_p^2 & \cdots & n_p^{-1} \sigma_p^2 \end{pmatrix}_{n_p \times n_p} \end{pmatrix},$$

$$A_d = \text{diag}\{n_1^{-1}p_{11}^{-2}, \dots, n_1^{-1}p_{11}^{-2}, \dots, n_p^{-1}p_{pp}^{-2}, \dots, n_p^{-1}p_{pp}^{-2}\},$$

$$B_d = \text{diag}\{n_1^{-1}\sigma_1^2, \dots, n_1^{-1}\sigma_1^2, \dots, n_p^{-1}\sigma_p^2, \dots, n_p^{-1}\sigma_p^2\},$$

$$P = \text{diag}\{p_{11}, \dots, p_{11}, \dots, p_{pp}, \dots, p_{pp}\},$$

$$Q = \text{diag}\{q_{11}, \dots, q_{11}, \dots, q_{pp}, \dots, q_{pp}\} \quad \text{e}$$

$$\Lambda = \text{diag}\{\sigma_1^2, \dots, \sigma_1^2, \dots, \sigma_p^2, \dots, \sigma_p^2\}.$$

Assim, os termos de ε_{2p} resultam em

$$(1) \quad \text{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q)] = \sum_{l=1}^n n_j n_j^{-1} p_{jj}^{-2} n_j^{-1} \sigma_j^2 \sigma_j^{-2} (2p_{jj}^2 - q_{jj}) = \sum_{l=1}^p n_j^{-1} (2 - q_{jj} p_{jj}^{-2}) ;$$

$$(2) \quad -\frac{1}{2} \text{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)] = -\frac{1}{2} \sum_{j=1}^p n_j n_j^{-2} p_{jj}^{-4} (-q_{jj}^2 + 8p_{jj}^2 q_{jj} - 6p_{jj}^4) \\ = -\frac{1}{2} \sum_{j=1}^n n_j^{-1} (-q_{jj}^2 p_{jj}^{-4} + 8q_{jj} p_{jj}^{-2} - 6) ;$$

$$(3) \quad 1_n^T P A_d B_d \Lambda^{-1} P 1_n = \sum_{j=1}^p n_j^2 p_{jj} n_j^{-1} p_{jj}^{-2} n_j^{-1} p_{jj}^{-2} n_j^{-1} \sigma_j^2 \sigma_j^{-2} p_{jj} = \sum_{j=1}^p n_j^{-1} p_{jj}^{-2} ;$$

$$(4) \quad \frac{1}{2} 1_n^T P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n = \frac{1}{2} \sum_{j=1}^p n_j^2 p_{jj} \sigma_j^{-2} n_j^{-1} p_{jj}^{-2} n_j^{-1} \sigma_j^2 \sigma_j^{-2} p_{jj} = \frac{1}{2} \sum_{j=1}^p n_j^{-1} ;$$

$$(5) \quad -1_n^T P \Lambda^{-1} [B^{(2)} \otimes A] \Lambda^{-1} P 1_n = -\sum_{j=1}^p n_j^2 p_{jj} \sigma_j^{-2} n_j^{-2} \sigma_j^4 n_j^{-1} p_{jj}^{-2} \sigma_j^{-2} p_{jj} = -\sum_{j=1}^p n_j^{-1} ;$$

$$(6) \quad \frac{1}{2} 1_n^T P Q A_d A A_d Q P 1_n = \frac{1}{2} \sum_{j=1}^p n_j^2 p_{jj} q_{jj} n_j^{-1} p_{jj}^{-2} n_j^{-1} p_{jj}^{-2} n_j^{-1} p_{jj}^{-2} n_j^{-1} p_{jj}^{-2} q_{jj} p_{jj} \\ = \frac{1}{2} \sum_{j=1}^p n_j^{-1} q_{jj}^2 p_{jj}^{-4} ;$$

$$(7) \quad \frac{1}{3} 1_n^T P (12P^2 A^{(3)} Q - 3Q A^{(3)} Q - 8P^2 A^{(3)} P^2) P 1_n = \frac{1}{3} \sum_{j=1}^p n_j^2 p_{jj} (12p_{jj}^2 n_j^{-3} p_{jj}^{-6} q_{jj} \\ - 3q_{jj}^2 n_j^{-3} p_{jj}^{-6} - 8p_{jj}^4 n_j^{-3} p_{jj}^{-6}) p_{jj} = \frac{1}{3} \sum_{j=1}^p n_j^{-1} (12p_{jj}^{-2} q_{jj} - 3p_{jj}^{-4} q_{jj}^2 - 8) .$$

Portanto,

$$\varepsilon_{2p} = \sum_{j=1}^p n_j^{-1} \left(2 - q_{jj} p_{jj}^{-2} + \frac{1}{2} q_{jj}^2 p_{jj}^{-4} - 4q_{jj} p_{jj}^{-2} + 3 + p_{jj}^{-2} + \frac{1}{2} - 1 \right. \\ \left. + \frac{1}{2} q_{jj}^2 p_{jj}^{-4} + 4q_{jj} p_{jj}^{-2} - q_{jj}^2 p_{jj}^{-4} - \frac{8}{3} \right) ,$$

que implica

$$\varepsilon_{2p} = \sum_{j=1}^p n_j^{-1} \left(\frac{11}{6} + p_{jj}^{-2} (1 - q_{jj}) \right) .$$

6.4 APERFEIÇOAMENTO DE TESTES DA RAZÃO DE VEROSSIMILHANÇA

Nesta seção, apresentamos o aperfeiçoamento do teste da estatística da razão de verossimilhança para algumas hipóteses nulas importantes dos modelos lineares normais heterocedásticos. Nestes modelos é bastante natural estarmos interessados em verificar homocedasticidade e, também, homogeneidade de médias. Esta verificação pode ser feita através do teste das hipóteses nulas $H_1 : \gamma_2 = \dots = \gamma_q = 0$ e $H_2 : \beta_2 = \dots = \beta_p = 0$, respectivamente, ambas testadas contra a alternativa A : violação de pelo menos uma igualdade. Muitas outras situações podem ser de interesse, tais como, testes sobre o vetor β , testes sobre o vetor γ , testes simultâneos sobre os vetores β e γ , ou, até mesmo, testes sobre subconjuntos dos vetores β e γ , simultâneos ou isolados.

No decorrer das subseções seguintes existe alguma notação que é comum e, para evitar repetição, estamos introduzindo-a agora. Denotamos por $\hat{\beta}$ e $\hat{\gamma}$ os estimadores de máxima verossimilhança irrestritos de β e γ , que podem ser obtidos, iterativamente, de (6.4). Consequentemente, $\hat{v}_l = (y_l - x_l^\top \hat{\beta})^2$, $\hat{\sigma}_l^2 = H(\omega_l^\top \hat{\gamma})$ e $\hat{v} = (\hat{v}_1, \dots, \hat{v}_n)^\top$. Definimos, também, $\tau = H^{(1)}(\gamma_1)/H(\gamma_1)$ e $\delta = H^{(2)}(\gamma_1)/H(\gamma_1)$.

6.4.1 TESTE DE HOMOCEDASTICIDADE

Para o teste de H_1 contra A temos que $\beta = (\beta_1, \dots, \beta_p)^\top$ e γ_1 são os parâmetros de perturbação. Podemos, ainda, admitir sem perda de generalidade $\omega_{l1} = 1$, para $l = 1, \dots, n$. A

estatística da razão de verossimilhança para o teste de H_1 é dada por

$$\omega_1 = -2 \log \lambda = 2(\hat{\ell}_{p+1} - \tilde{\ell}_{p+1}) , \quad (6.19)$$

sendo $\hat{\ell}_{p+q} = \ell(\hat{\beta}, \hat{\gamma})$ o logaritmo da função de verossimilhança avaliado em $\hat{\beta}$ e $\hat{\gamma}$ e $\tilde{\ell}_{p+1} = \ell(\tilde{\beta}, \tilde{\gamma}_1)$ o logaritmo da função de verossimilhança avaliado nos estimadores de máxima verossimilhança sob H_1 , $\tilde{\beta}$ e $\tilde{\gamma}_1$, respectivamente. Para encontrarmos $\tilde{\beta}$ e $\tilde{\gamma}_1$ consideramos a função em (6.3), sob H_1 , ou seja, fazendo $\gamma_2 = \dots = \gamma_q = 0$ e $\sigma^2 = H(\gamma_1)$, resultando,

$$\ell(\beta, \gamma_1) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{l=1}^n v_l . \quad (6.20)$$

Tomando-se derivadas em relação a γ_1 e β_i , $i = 1, \dots, p$, e igualando-as a zero, resulta no sistema de equações

$$\sum_{l=1}^n (y_l x_{li} - x_l^\top \beta x_{li}) = 0 , \quad i = 1, \dots, p, \quad \text{e} \quad -\frac{n}{2} \tau + \frac{\tau}{2\sigma^2} \sum_{l=1}^n v_l = 0,$$

cujas soluções é dada, em notação matricial, por

$$\tilde{\beta} = (X^\top X)^{-1} X^\top y, \quad \tilde{\sigma}^2 = \frac{1}{n} 1_n^\top \tilde{v} , \quad (6.21)$$

onde $\tilde{v} = (\tilde{v}_1, \dots, \tilde{v}_n)^\top$ com $\tilde{v}_l = (y_l - x_l^\top \tilde{\beta})^2$, $l = 1, \dots, n$. Convém observar, aqui, que as equações em (6.21) poderiam ser obtidas diretamente de (6.4), acrescentando-se as restrições da hipótese nula. Assim,

$$\hat{\ell}_{p+1} = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \tilde{\sigma}^2 - \frac{n}{2} . \quad (6.22)$$

Então, a estatística da razão de verossimilhança em (6.19) é calculada por

$$\omega_1 = n(\log \tilde{\sigma}^2 + 1) - \sum_{l=1}^n \log \hat{\sigma}_l^2 - \sum_{l=1}^n \frac{\hat{v}_l}{\hat{\sigma}_l^2} . \quad (6.23)$$

Vimos que, até primeira ordem, ω_1 tem distribuição aproximada χ_{q-1}^2 . Esta aproximação pode ser melhorada, até ordem n^{-1} , através da correção de Bartlett usando a estatística modificada

$$\omega_1^* = c_1^{-1} \omega_1 ,$$

onde $c_1 = (q-1)^{-1}E(\omega_1)$ é a correção de Bartlett e $E(\omega_1)$ é avaliado até ordem n^{-1} e sob H_1 , ou seja, $E(\omega_1) = q-1 + \varepsilon_{p+q} - \varepsilon_{p+1}$, com ε_{p+1} dado por (6.17) e ε_{p+q} calculado de (6.11) - (6.13), sob a hipótese de homocedasticidade. Para a determinação de ε_{p+q} segundo H_1 , observamos que $\Lambda = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\} = \sigma^2 I_n$, sendo $\sigma^2 = H(\gamma_1)$, que resulta $P = \tau I_n$ e $Q = \delta I_n$. Definindo $M = W(W^\top W)^{-1}W^\top$ e $N = X(X^\top X)^{-1}X^\top$ podemos escrever $A = M/\tau^2$ e $B = \sigma^2 N$. Assim, obtemos, facilmente, os seguintes termos de ε_{p+q} sob H_1 , em (6.11) - (6.13), após simplificações:

$$\begin{aligned}
(1) \quad & \text{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q)] = \frac{(2\tau^2 - \delta)}{\tau^2} \text{tr}(M_d N_d); \\
(2) \quad & -\frac{1}{2} \text{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)] = \frac{(\delta^2 + 8\tau^2 \delta - 6\tau^4)}{2\tau^4} \text{tr}(M_d^2); \\
(3) \quad & 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n = \frac{1}{\tau^2} 1_n^\top M_d M N_d 1_n; \\
(4) \quad & \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n = \frac{1}{2} 1_n^\top N_d M N_d 1_n; \\
(5) \quad & -1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n = -1_n^\top (N^{(2)} \otimes M) 1_n; \\
(6) \quad & \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n = \frac{\delta^2}{2\tau^4} 1_n^\top M_d M M_d 1_n; \\
(7) \quad & \frac{1}{3} 1_n^\top P (12P^2 A^{(3)} Q - 3Q A^{(3)} Q - 8P^2 A^{(3)} P^2) P 1_n = \left(\frac{4\delta}{\tau^2} - \frac{\delta^2}{\tau^4} - \frac{8}{n} \right) 1_n^\top M^{(3)} 1_n,
\end{aligned}$$

onde M_d e N_d são matrizes diagonais com os elementos diagonais iguais aos das matrizes M e N , respectivamente. Portanto, denotando $\varepsilon_{p+q}(H_1)$ o valor da quantidade ε_{p+q} calculado sob a hipótese nula H_1 , vem

$$\begin{aligned}
\varepsilon_{p+q}(H_1) = & \frac{1}{2} \text{tr} \left[\frac{2(2\tau^2 - \delta)}{\tau^2} M_d N_d - \frac{(-\delta^2 + 8\tau^2 - 6\tau^4)}{\tau^4} M_d^2 \right] + \frac{1}{6} 1_n^\top \left[2 \left(\frac{12\delta}{\tau^2} - \frac{3\delta^2}{\tau^4} - 8 \right) M^{(3)} \right. \\
& \left. + \frac{3\delta^2}{\tau^4} M_d M M_d - 6(N^{(2)} \otimes M) + 3N_d M N_d + \frac{6}{\tau^2} M_d M N_d \right] 1_n. \quad (6.24)
\end{aligned}$$

Assim, $c_1 = (q-1)^{-1}E(\omega_1) = 1 + \frac{\varepsilon_{p+q}(H_1) - \varepsilon_{p+1}}{q-1}$ implicando

$$c_1 = 1 + \frac{\varepsilon_{p+q}(H_1)}{q-1} - \frac{1}{q-1} \left[\frac{3(p+1)^2 - 1}{6n} - \frac{(\delta-1)p}{n\tau^2} \right] \quad (6.25)$$

para a correção de Bartlett associada ao teste de homocedasticidade nos modelos normais heterocedásticos. Na situação em que $H(\cdot) = \exp(\cdot)$, temos $\tau = \delta = 1$ e $P = Q = I_n$, resultando

$$\varepsilon_{p+q}(H_1) = \frac{1}{2} \text{tr}[2M_d N_d - M_d^2] + \frac{1}{6} 1_n^\top [2M^{(3)} + 3M_d M M_d - 6(N^{(2)} \otimes M) + 3N_d M N_d + 6M_d M N_d] 1_n$$

e

$$\varepsilon_{p+1} = \frac{3(p+1)^2 - 1}{6n} .$$

Estas últimas expressões coincidem com aquelas em (4.1) de Cordeiro (1993b), que estuda os modelos normais heterocedásticos multiplicativos, observando que $\varepsilon_{p+q}(H_1)$ aqui é a função $h(X, W, \sigma^2 I_n)$ lá.

6.4.2 TESTE DE HOMOGENEIDADE DE MÉDIAS

Sem perda de generalidade, podemos admitir $x_{l1} = 1$, para $l = 1, \dots, n$. A hipótese H_2 de homogeneidade de médias é equivalente a $H_2 : \mu_1 = \dots = \mu_n = \beta_1$, sendo $\gamma = (\gamma_1, \dots, \gamma_q)^\top$ e β_1 os parâmetros de incômodo. A estatística da razão de verossimilhança, neste caso, é dada por

$$\omega_2 = -2 \log \lambda = 2(\hat{\ell}_{p+q} - \hat{\ell}_{1+q}) , \quad (6.26)$$

sendo $\hat{\ell}_{1+q} = \ell(\tilde{\beta}_1, \tilde{\gamma})$, com $\tilde{\beta}_1$ e $\tilde{\gamma}$ os estimadores de máxima verossimilhança de β_1 e γ sob H_2 .

Como no teste de homocedasticidade, temos

$$\hat{\ell}_{p+q} = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{l=1}^n \log \hat{\sigma}_l^2 - \frac{1}{2} \sum_{l=1}^n \frac{\hat{v}_l}{\hat{\sigma}_l^2} . \quad (6.27)$$

Sob a hipótese nula H_2 vem

$$\ell = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{l=1}^n \log \sigma_l^2 - \frac{1}{2} \sum_{l=1}^n \frac{v_l}{\sigma_l^2} ,$$

onde $v_l = (y_l - \beta_1)^2$, $l = 1, \dots, n$. Tomando-se derivadas de ℓ em relação a β_1 e a γ_j , para $j = 1, \dots, q$ e igualando-as a zero, resulta nas seguintes equações:

$$\sum_{l=1}^n \frac{y_l}{\tilde{\sigma}_l^2} = \tilde{\beta}_1 \sum_{l=1}^n \frac{1}{\tilde{\sigma}_l^2} \quad \text{e} \quad \sum_{l=1}^n \frac{\tilde{H}_l^{(1)}}{\tilde{H}_l} \omega_l = \sum_{l=1}^n \frac{\tilde{v}_l \tilde{H}_l^{(1)}}{\tilde{H}_l^2} \omega_l ,$$

sendo $\tilde{v}_l = (y_l - \tilde{\beta}_1)^2$, $\tilde{H}_l = H(\omega_l^\top \tilde{\gamma})$ e $\tilde{H}_l^{(1)} = H^{(1)}(\omega_l^\top \tilde{\gamma})$, que podem ser expressas, matricialmente, por

$$\tilde{\beta}_1 = \frac{1_n^\top \tilde{\Lambda}^{-1} y}{1_n^\top \tilde{\Lambda}^{-1} 1_n} \quad \text{e} \quad W^\top \tilde{P} 1_n = W^\top \tilde{P} \tilde{\Lambda}^{-1} \tilde{v} . \quad (6.28)$$

Devemos observar que as equações em (6.28) poderiam, também, ser obtidas diretamente de (6.4), acrescentando-se as restrições da hipótese nula. Então,

$$\hat{\ell}_{1+q} = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{l=1}^n \log \tilde{\sigma}_l^2 - \frac{1}{2} \sum_{l=1}^n \frac{\tilde{v}_l}{\tilde{\sigma}_l^2} ,$$

que resulta

$$\begin{aligned} \omega_2 &= 2 \left\{ -\frac{1}{2} \sum_{l=1}^n \log \hat{\sigma}_l^2 - \frac{1}{2} \sum_{l=1}^n \frac{\hat{v}_l}{\hat{\sigma}_l^2} + \frac{1}{2} \sum_{l=1}^n \log \tilde{\sigma}_l^2 + \frac{1}{2} \sum_{l=1}^n \frac{\tilde{v}_l}{\tilde{\sigma}_l^2} \right\} \\ &= \sum_{l=1}^n \log \left(\frac{\tilde{\sigma}_l^2}{\hat{\sigma}_l^2} \right) + 1_n^\top (\tilde{\Lambda}^{-1} \tilde{v} - \hat{\Lambda}^{-1} \hat{v}) . \end{aligned} \quad (6.29)$$

A estatística aperfeiçoada, neste caso, é

$$\omega_2^* = c_2^{-1} \omega_2 ,$$

onde $c_2 = (p-1)^{-1} E(\omega_2)$ é a correção de Bartlett e $E(\omega_2)$ é avaliado até ordem n^{-1} sob a hipótese nula H_2 , ou seja, $E(\omega_2) = p-1 + \varepsilon_{p+q} - \varepsilon_{1+q}$, com ε_{1+q} dado por (6.18) e ε_{p+q} dado por (6.11). Escrevemos, então,

$$E(\omega_2) = p-1 + h(X, W, \Lambda) - h(1_n, W, \Lambda) , \quad (6.30)$$

onde

$$h(X, W, \Lambda) - h(1_n, W, \Lambda) = \varepsilon_{p+q} - \varepsilon_{1+q}$$

$$\begin{aligned}
&= \text{tr}\{A_d B_d \Lambda^{-1}(2P^2 - Q)\} - \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \text{tr}\{A_d \Lambda^{-1}(2P^2 - Q)\} \\
&\quad + 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n - \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} 1_n^\top P A_d A \Lambda^{-1} P 1_n \\
&\quad + \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n + \frac{1}{2 \sum_{l=1}^n \sigma_l^{-2}} 1_n^\top P \Lambda^{-1} A \Lambda^{-1} P 1_n \\
&\quad - 1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n .
\end{aligned} \tag{6.31}$$

No caso particular em que $q = 1$ ($W = 1_n$), temos $\Lambda = \sigma^2 I_n$, $P = \tau I_n$ e $Q = \delta I_n$, resultando $A = \frac{1}{n\tau^2} 1_n 1_n^\top$ e $B = \sigma^2 X(X^\top X)^{-1} X^\top$. Então, $h(X, W, \Lambda) = h(X, \Lambda)$ iguala a ε_{p+1} em (6.17) e $h(1_n, W, \Lambda) = h(\Lambda)$ é calculado por (6.17) fazendo $p = 1$. Assim, obtemos

$$E(\omega_2) = p - 1 + \frac{(p+1)^2 - 4}{2n} - \frac{(p-1)(\delta-1)}{n\tau^2} . \tag{6.32}$$

Se nesta situação consideramos $H(\cdot) = \exp(\cdot)$ então, $\tau = \delta = 1$, resultando $E(\omega_2) = p - 1 + \frac{(p+1)^2 - 4}{2n}$. Este resultado é o mesmo obtido por Cordeiro (1993b), expressão (4.3), para os modelos heterocedásticos multiplicativos.

6.4.3 TESTE SOBRE OS VETORES DE PARÂMETROS β E γ

Podemos estar interessados em testar a hipótese nula simples $H_3 : \beta = \beta^{(0)}, \gamma = \gamma^{(0)}$, sendo $\beta^{(0)}$ e $\gamma^{(0)}$ vetores especificados de dimensão $p \times 1$ e $q \times 1$, respectivamente. Esta hipótese, em geral, não é frequente na prática. Nas seções 6.4.4 a 6.4.6, estudamos hipóteses de maior interesse prático. Temos, neste caso, zero parâmetros na hipótese nula e $p + q$ na alternativa. A estatística da razão de verossimilhança é dada por

$$\omega_3 = -2 \log \lambda = 2 \{ \ell(\hat{\beta}, \hat{\gamma}) - \ell(\beta^{(0)}, \gamma^{(0)}) \} .$$

Através de (6.3), calculada em $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$ e em $\theta^{(0)} = (\beta^{(0)\top}, \gamma^{(0)\top})^\top$, encontramos

$$\omega_3 = \sum_{l=1}^n \log \left(\frac{\sigma_l^{(0)2}}{\hat{\sigma}_l^2} \right) + \sum_{l=1}^n \frac{(y_l - x_l^\top \beta^{(0)})}{\sigma_l^{(0)2}} - \sum_{l=1}^n \frac{(y_l - x_l^\top \hat{\beta})^2}{\hat{\sigma}_l^2} , \tag{6.33}$$

que pode ser escrita em notação matricial, em analogia com (6.2), por

$$\omega_3 = \log \frac{|\Lambda^{(0)}|}{|\hat{\Lambda}|} + \mathbf{1}_n^\top [\Lambda^{(0)-1} v^{(0)} - \hat{\Lambda}^{-1} \hat{v}] . \quad (6.34)$$

A distribuição assintótica de primeira ordem ω_3 é χ_{p+q}^2 . A aproximação pode ser melhorada até ordem n^{-2} por $\omega_3^* = c_3^{-1} \omega_3$, onde c_3 é a correção de Bartlett calculada por

$$c_3 = (p + q)^{-1} E(\omega_3) ,$$

sendo $E(\omega_3) = p + q + \varepsilon_{p+q}$ até ordem n^{-1} , com ε_{p+q} obtido de (6.11) - (6.13) e avaliado sob H_3 .

Se $q = 1$, com $W = \mathbf{1}_n$, temos $\sigma_l^2 = \sigma^2 = H(\gamma_1)$, para $l = 1, \dots, n$, e, neste caso, a hipótese nula é $H_3 : \sigma^2 = \sigma^{(0)2}, \beta = \beta^{(0)}$. Então, até ordem n^{-1} , $E(\omega_3) = p + 1 + \varepsilon_{p+1}$, com ε_{p+1} de ordem n^{-1} , calculado por (6.17), ou seja, até esta ordem, $E(\omega_3) = p + 1 + \frac{3(p+1)^2 - 1}{6n} - \frac{(\delta-1)p}{n\tau^2}$. Quando $p = 1$, com $X = \mathbf{1}_n$, temos $\mu_l = \mu$, para $l = 1, \dots, n$ e a hipótese nula é da forma $H_3 : \mu = \mu^{(0)}, \gamma = \gamma^{(0)}$. O valor esperado da estatística da razão de verossimilhança até ordem n^{-1} é $E(\omega_3) = 1 + q + \varepsilon_{1+q}$, com ε_{1+q} obtido de (6.18). Se temos, agora, $q = 1$, então $A = \mathbf{1}_n \mathbf{1}_n^\top / n\tau^2$, $A_d = I_n / n\tau^2$ e $\Lambda = \sigma^2 I_n$, produzindo em (6.18) $\varepsilon_{1+1} = \frac{11}{6n} - \frac{(\delta-1)}{n\tau^2}$. Uma outra forma de obtermos este resultado é através de (6.17) fazendo $p = 1$.

6.4.4 TESTE SOBRE O VETOR DE PARÂMETROS β

Supomos, agora, que nosso interesse é o vetor de parâmetros β , sendo, portanto, γ o vetor de parâmetros de incômodo. Desejamos testar, então, a hipótese nula composta $H_4 : \beta = \beta^{(0)}$ contra a alternativa $A : \beta \neq \beta^{(0)}$, sendo $\beta^{(0)}$ um vetor $p \times 1$ de valores pré-fixados. A estatística da razão de verossimilhança, cuja distribuição assintótica de primeira ordem é χ_p^2 , sob H_4 , tem a forma

$$\omega_4 = 2\{\ell(\hat{\beta}, \hat{\gamma}) - \ell(\beta^{(0)}, \tilde{\gamma})\} , \quad (6.35)$$

$\tilde{\gamma}$ sendo a estimativa de máxima verossimilhança de γ sob H_4 . O logaritmo da função de verossimilhança, segundo H_4 , pode ser escrito por

$$\ell = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{l=1}^n \log H_l - \frac{1}{2} \sum_{l=1}^n \frac{v_l^{(0)}}{H_l}, \quad (6.36)$$

onde $v_l^{(0)} = (y_l - x_l^\top \beta^{(0)})^2$. A estimativa $\tilde{\gamma}$ é dada como solução da equação obtida igualando-se a zero a derivada de ℓ em relação a γ . Esta operação resulta em

$$W^\top \tilde{P} 1_n = W^\top \tilde{P} \tilde{\Lambda}^{-1} v^{(0)}, \quad (6.37)$$

com $v^{(0)} = (v_1^{(0)}, \dots, v_n^{(0)})^\top$ e $\tilde{P} = P(\tilde{\gamma})$ e $\tilde{\Lambda} = \Lambda(\tilde{\gamma})$. Assim, ω_4 em (6.35) é calculada por

$$\omega_4 = \sum_{l=1}^n \log \frac{\tilde{\sigma}_l^2}{\hat{\sigma}_l^2} + \sum_{l=1}^n \left(\frac{v_l^{(0)}}{\tilde{\sigma}_l^2} - \frac{\hat{v}_l}{\hat{\sigma}_l^2} \right), \quad (6.38)$$

sendo $\tilde{\sigma}_l^2 = H(\omega_l^\top \tilde{\gamma})$. Novamente, através de Lawley (1956), temos que, até ordem n^{-1} e sob H_4 , $E(\omega_4) = p + \varepsilon_{p+q} - \varepsilon_q$, com ε_{p+q} obtido de (6.11) - (6.13) e ε_q obtido a partir de (6.6) observando, agora, que o somatório $\sum'_{\beta, \gamma}$ daquela expressão estende-se apenas sobre os q componentes do vetor γ , ou seja, $\varepsilon_q = \sum'_\gamma \ell_{RSTU} - \sum'_\gamma \ell_{RSTUVW}$. Não é difícil provarmos, utilizando cumulantes apropriados e suas derivadas, apresentados no Apêndice 3, que

$$\begin{aligned} \varepsilon_q = & -\frac{1}{2} \text{tr}[A_d^2(-Q^2 + 8P^2Q - 6P^4)] + \frac{1}{3} 1_n^\top P[4P^2 A^{(3)}(3Q - 2P^2) - 3QA^{(3)}Q]P1_n \\ & + \frac{1}{2} 1_n^\top PQ A_d A A_d Q P 1_n, \end{aligned} \quad (6.39)$$

resultando,

$$\begin{aligned} E(\omega_4) = & p + \text{tr}[A_d B_d \Lambda^{-1}(2P^2 - Q)] - 1_n^\top P \Lambda^{-1}(B^{(2)} \otimes A) \Lambda^{-1} P 1_n \\ & + \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n + 1_n^\top P A_d A B_d \Lambda^{-1} P 1_n, \end{aligned} \quad (6.40)$$

onde Λ é estimado por $\tilde{\Lambda}$. O fator de correção de Bartlett c_4 associado à estatística ω_4 é determinado pela relação $c_4 = p^{-1} E(\omega_4)$, sendo que $E(\omega_4)$ é dada por (6.40). A estatística modificada $\omega_4^* = c_4^{-1} \omega_4$ tem distribuição χ_p^2 até ordem n^{-1} , sob H_4 .

Como um caso particular de (6.39), considerando $q = 1$, com $W = 1_n$, que resulta $\Lambda = \sigma^2 I_n$, onde $\sigma^2 = H(\gamma_1)$, temos $P = \tau I_n$, $Q = \delta I_n$, $A = 1_n 1_n^\top / n \tau^2$ e $B = \sigma^2 N$ com $N = X(X^\top X)^{-1} X^\top$. Assim, o valor esperado da estatística ω_4 em (6.40) reduz-se a

$$E(\omega_4) = p + \frac{p^2 + 4p}{2n} - \frac{\delta p}{\tau^2 n}. \quad (6.41)$$

Logo,

$$c_4 = 1 + \frac{p + 4}{2n} - \frac{\delta}{\tau^2 n}.$$

Temos, ainda, neste caso que $\tilde{\sigma}^2 = \sum_{l=1}^n \frac{(y_l - x_l^\top \beta^{(0)})^2}{n}$ e $\hat{\sigma}^2 = \sum_{l=1}^n \frac{(y_l - x_l^\top \hat{\beta})^2}{n}$, com $\hat{\beta} = (X^\top X)^{-1} X^\top y$, produzindo

$$\omega_4 = \sum_{l=1}^n \log \left(\frac{\tilde{\sigma}_l^2}{\hat{\sigma}_l^2} \right) = \sum_{l=1}^n \log \left\{ \frac{\sum_{l=1}^n (y_l - x_l^\top \beta^{(0)})^2}{\sum_{l=1}^n (y_l - x_l^\top \hat{\beta})^2} \right\}. \quad (6.42)$$

A expressão (6.41) pode ser verificada, também, expandindo (6.42) em série de Taylor e tomando o valor esperado.

6.4.5 TESTE SOBRE O VETOR DE PARÂMETROS γ

Suponhamos, agora, que nosso interesse é o vetor de parâmetros γ , sendo β o vetor de parâmetros de perturbação. Queremos, então, testar a hipótese nula $H_5 : \gamma = \gamma^{(0)}$ contra a alternativa $A : \gamma \neq \gamma^{(0)}$, onde $\gamma^{(0)}$ é um vetor especificado de dimensão $q \times 1$. A estatística da razão de verossimilhança para testar H_5 é expressa por

$$\omega_5 = 2[\ell(\hat{\beta}, \hat{\gamma}) - \ell(\tilde{\beta}, \gamma^{(0)})], \quad (6.43)$$

cuja distribuição assintótica de primeira ordem, sob H_5 , é χ_q^2 , onde $\tilde{\beta}$ é o estimador de máxima verossimilhança de β , segundo a hipótese nula H_5 . Podemos encontrar $\tilde{\beta}$ resolvendo a equação obtida derivando-se o logaritmo da função de verossimilhança (6.3) em relação a β , segundo H_5 , e igualando a zero. Temos, então,

$$\ell(\beta, \gamma^{(0)}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{l=1}^n \log \sigma_l^{(0)^2} - \frac{1}{2} \sum_{l=1}^n \frac{(y_l - x_l^\top \beta)^2}{\sigma_l^{(0)^2}}$$

e

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2} \sum_{l=1}^n \frac{-2x_l(y_l - x_l^\top \beta)}{\sigma_l^{(0)2}} = X^\top \Lambda^{(0)-1} y - (X^\top \Lambda^{(0)-1} X) \beta = 0 ,$$

que produz

$$\tilde{\beta} = (X^\top \Lambda^{(0)-1} X)^{-1} X^\top \Lambda^{(0)-1} y . \quad (6.44)$$

Assim, a estatística ω_5 em (6.43) é calculada por

$$\omega_5 = \sum_{l=1}^n \log \left(\frac{\sigma_l^{(0)2}}{\hat{\sigma}_l^2} \right) + \sum_{l=1}^n \left[\frac{(y_l - x_l^\top \tilde{\beta})^2}{\sigma_l^{(0)2}} - \frac{(y_l - x_l^\top \hat{\beta})^2}{\hat{\sigma}_l^2} \right] . \quad (6.45)$$

De maneira análoga aos outros testes estudados, temos que, segundo Lawley (1956), $E(\omega_5) = q + \varepsilon_{p+q} - \varepsilon_p + O(n^{-2})$, onde ε_{p+q} é obtido de (6.11) - (6.13) e ε_p de (6.6) com o somatório $\sum'_{\beta, \gamma}$ estendendo-se apenas sobre os p componentes de β . Dos resultados do Apêndice 3, facilmente, encontramos $\varepsilon_p = 0$. Assim, até ordem n^{-1} e sob H_5 , $E(\omega_5) = q + \varepsilon_{p+q}$, com $\varepsilon_{p+q} = h(X, W, \Lambda^{(0)})$ obtido por (6.11). O fator de correção de Bartlett c_5 associado a ω_5 é determinado, então, por

$$c_5 = q^{-1} E(\omega_5) = 1 + \frac{\varepsilon_{p+q}}{q} . \quad (6.46)$$

Se $q = 1$, com $W = 1_n$, $c_5 = 1 + \varepsilon_{p+1}$, com ε_{p+1} de ordem n^{-1} obtido de (6.17). Assim, $c_5 = 1 + \frac{3(p+1)^2 - 1}{6n} - \frac{(\delta-1)p}{n\tau^2}$. Para $p = 1$, com $X = 1_n$, resulta $c_5 = 1 + \varepsilon_{1+q}/q$, sendo $\varepsilon_{1+q} = h(1_n, W, \Lambda^{(0)})$, avaliado até ordem n^{-1} sob H_5 , cuja expressão é obtida de (6.18), com $\Lambda = \Lambda^{(0)}$. Finalmente, se temos $p = q = 1$, ou seja, β e γ são parâmetros escalares, então $E(\omega_5) = 1 + \varepsilon_{1+1}$ até ordem n^{-1} , sendo $\varepsilon_{1+1} = \frac{11}{6n} - \frac{(\delta-1)}{n\tau^2}$, conforme já apresentado no final da Seção 6.4.3. Neste caso, $c_5 = 1 + \frac{11}{6n} - \frac{(\delta-1)}{n\tau^2}$.

6.4.6 TESTE SOBRE UM SUBCONJUNTO DO VETOR $(\beta^\top, \gamma^\top)^\top$

Em muitos problemas as restrições sob um teste envolve um subconjunto dos vetores de parâmetros β e γ . Dividindo estes vetores paramétricos como $\beta = (\beta_1^\top, \beta_2^\top)^\top$ e $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$,

onde $\beta_1 = (\beta_1, \dots, \beta_{p_1})^\top$, $\beta_2 = (\beta_{p_1+1}, \dots, \beta_p)^\top$, $\gamma_1 = (\gamma_1, \dots, \gamma_{q_1})^\top$ e $\gamma_2 = (\gamma_{q_1+1}, \dots, \gamma_q)^\top$, com $p_1 \leq p$ e $q_1 \leq q$, estamos interessados em testar a hipótese nula $H_6 : \beta_1 = \beta_1^{(0)}$, $\gamma_1 = \gamma_1^{(0)}$ contra a alternativa A : violação de pelo menos uma igualdade, onde $\beta_1^{(0)}$ e $\gamma_1^{(0)}$ são vetores especificados de dimensões p_1 e q_1 , respectivamente, e β_2 e γ_2 são vetores de parâmetros de perturbação. As partições de β e γ induzidas pela hipótese H_6 correspondem às partições $X = (X_1, X_2)$ e $W = (W_1, W_2)$, onde X_1, X_2, W_1 e W_2 são matrizes conhecidas de postos completos e dimensões iguais a $n \times p_1$, $n \times (p - p_1)$, $n \times q_1$ e $n \times (q - q_1)$, respectivamente. A estatística da razão de verossimilhança para testar H_6 é calculada por

$$\omega_6 = \log \frac{|\tilde{\Lambda}|}{|\hat{\Lambda}|} + 1_n^\top (\tilde{\Lambda}^{-1} \tilde{v} - \hat{\Lambda}^{-1} \hat{v}) , \quad (6.47)$$

onde os estimadores de máxima verossimilhança sujeitos a H_6 são denotados com um til ($\sim$). Seguindo o raciocínio utilizado nos testes anteriores, a correção de Bartlett c_6 , que aperfeiçoa a estatística ω_6 , segue diretamente de (6.11), e é dada por

$$c_6 = 1 + \frac{h(X, W, \Lambda) - h(X_2, W_2, \Lambda)}{p_1 + q_1} . \quad (6.48)$$

Assim, o teste aperfeiçoado para a hipótese H_6 compara a estatística modificada $\omega_6^* = c_6^{-1} \omega_6$ com o $(1 - \alpha)$ -ésimo percentil da distribuição $\chi_{p_1+q_1}^2$.

6.4.7 TESTE SOBRE SUBCONJUNTOS DE β E DE γ

a) Teste sobre um subconjunto de β

Consideremos, agora, a mesma partição da Seção 6.4.6 para o vetor β , ou seja, $\beta^\top = (\beta_1^\top, \beta_2^\top)$, onde β_1 é $p_1 \times 1$ e β_2 é $(p - p_1) \times 1$, e a correspondente partição da matriz modelo $X = (X_1, X_2)$. O interesse é testar $H_7 : \beta_1 = \beta_1^{(0)}$ versus $A : \beta_1 \neq \beta_1^{(0)}$, onde $\beta_1^{(0)}$ é um vetor especificado de dimensão p_1 e β_2 e γ são vetores de parâmetros de incômodo. O critério da razão de verossimilhança para o teste de H_7 utiliza a estatística

$$\omega_7 = 2 \{ \ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma}) \} , \quad (6.49)$$

cuja distribuição assintótica de primeira ordem, sob H_7 é $\chi_{p_1}^2$. Reescrevendo ω_7 em (6.49) como

$$\omega_7 = 2 \left\{ \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] - \left[\ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma) \right] \right\}, \quad (6.50)$$

temos, segundo Lawley (1956), que, até ordem n^{-1} , $2E\{\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma)\} = p + q + \varepsilon_{p+q}$ e $2E\{\ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\gamma}) - \ell(\beta_1^{(0)}, \beta_2, \gamma)\} = (p - p_1) + q + \varepsilon_{(p-p_1)+q}$, com $\varepsilon_{(p-p_1)+q}$ obtido de (6.6) com o somatório $\sum'_{\beta, \gamma}$ estendendo-se sobre os componentes de β_2 e γ , isto é, sobre $(p - p_1) + q$ parâmetros. As quantidades em (6.50) assinaladas com til são as correspondentes estimativas de máxima verossimilhança, segundo a hipótese nula H_7 . Sob H_7 , obtemos, então, que até ordem n^{-1} ,

$$E(\omega_7) = p_1 + \varepsilon_{p+q} - \varepsilon_{(p-p_1)+q}. \quad (6.51)$$

A correspondente correção de Bartlett c_7 , associada à estatística ω_7 , é calculada por $c_7 = p_1^{-1} E(\omega_7)$. Utilizando a expressão (6.11) em (6.51), resulta

$$c_7 = 1 + \frac{h(X, W, \Lambda) - h(X_2, W, \Lambda)}{p_1},$$

e o teste compara a estatística aperfeiçoada $\omega_7^* = \tilde{c}_7^{-1} \omega_7$ com o $(1 - \alpha)$ -ésimo percentil da distribuição $\chi_{p_1}^2$, sendo que $\tilde{c}_7$ é o valor c_7 acima avaliado em $\tilde{\Lambda}$.

b) Teste sobre um subconjunto de γ

Consideramos, agora, a partição do vetor γ como $\gamma^\top = (\gamma_1^\top, \gamma_2^\top)$, com $\gamma_1 = (\gamma_1, \dots, \gamma_{q_1})^\top$ e $\gamma_2 = (\gamma_{q_1+1}, \dots, \gamma_q)^\top$, e a correspondente partição da matriz $W = (W_1, W_2)$, com W_1 e W_2 de dimensões $n \times q_1$ e $n \times (q - q_1)$, respectivamente. Para testar $H_8 : \gamma_1 = \gamma_1^{(0)}$ contra $A : \gamma_1 \neq \gamma_1^{(0)}$, com $\gamma_1^{(0)}$ um vetor especificado de dimensão q_1 e γ_2 e β os vetores de parâmetros de perturbação, a estatística da razão de verossimilhança é dada por

$$\omega_8 = 2[\ell(\hat{\beta}, \hat{\gamma}) - \ell(\hat{\beta}, \gamma_1^{(0)}, \tilde{\gamma}_2)] , \quad (6.52)$$

que, para amostras grandes e sob H_8 é, aproximadamente, $\chi_{q_1}^2$. O fator de correção de Bartlett c_8 associado a ω_8 é determinado por $c_8 = q_1^{-1} E(\omega_8)$, com $E(\omega_8)$ avaliado até ordem n^{-1} sob

H_8 . Análogo aos testes anteriores obtemos, segundo Lawley (1956), que até ordem n^{-1} e sob esta hipótese,

$$E(\omega_8) = q_1 + \varepsilon_{p+q} - \varepsilon_{p+(q-q_1)} , \quad (6.53)$$

com ε_{p+q} calculado por (6.11) - (6.13) e sob H_8 , e $\varepsilon_{p+(q-q_1)}$ de (6.6), considerando, agora, que o somatório $\sum'_{\beta, \gamma}$ estende-se somente sobre os componentes de β e γ_2 , isto é, sobre $p + (q - q_1)$ parâmetros. Assim, $E(\omega_8)$ em (6.53) pode ser calculada por

$$E(\omega_8) = q_1 + h(X, W, \Lambda) - h(X, W_2, \Lambda) , \quad (6.54)$$

onde $h(X, W, \Lambda)$ é obtida por (6.11) e sob H_8 . Como a correção de Bartlett associada à estatística ω_8 é $c_8 = q_1^{-1} E(\omega_8)$, então, vem

$$c_8 = 1 + \frac{h(X, W, \Lambda) - h(X, W_2, \Lambda)}{q_1} , \quad (6.55)$$

e a estatística aperfeiçoada $\omega_8^* = c_8^{-1} \omega_8$ tem, sob H_8 e até ordem n^{-1} , distribuição $\chi^2_{q_1}$.

Para finalizar, é importante destacarmos que nosso objetivo principal foi obter uma expressão matricial para o fator de correção de Bartlett, associado a testes de hipóteses sobre parâmetros dos modelos normais heterocedásticos. Esta situação generaliza aquela considerada por Cordeiro (1993b), que estuda os modelos normais heterocedásticos multiplicativos. Em várias situações de interesse em econometria, a hipótese de $H(\cdot)$ ser a função exponencial pode ser muito restritiva. Estudamos o aperfeiçoamento da estatística da razão de verossimilhança apropriado para testar oito hipóteses nulas. Entretanto, as hipóteses consideradas não esgotam todas as possibilidades, mas ajudam na dedução do aperfeiçoamento de estatísticas da razão de verossimilhança para os testes de outras hipóteses nulas. Podemos, ainda, futuramente comparar o desempenho da estatística da razão de verossimilhança corrigida com o da estatística escore corrigida, para os modelos normais heterocedásticos com erros independentes, para o teste de heterocedasticidade considerado por Cribari-Neto e Ferrari (1995). Pretendemos, também, encontrar uma interpretação estatística das correções obtidas.

6.5 UM ESTUDO DE SIMULAÇÃO

Finalizando, apresentamos, nesta seção, um pequeno estudo de simulação, cujo objetivo é verificar o desempenho da estatística da razão de verossimilhança modificada, comparado ao da estatística usual, em amostras pequenas e moderadas, para uma situação particular dos modelos normais heterocedásticos.

Consideramos o modelo normal heterocedástico definido pelas componentes sistemáticas

$$\begin{aligned}\mu_l &= \beta_0 + \beta_1 x_{l1} + \beta_2 x_{l2} + \beta_3 x_{l3} , \\ \sigma_l^2 &= \sigma^2(\gamma_0 + \gamma_1 \omega_{l1} + \gamma_2 \omega_{l2})^{2\lambda} ,\end{aligned}\tag{6.56}$$

para $l = 1, \dots, n$. Desejamos testar a hipótese nula $H : \beta_1 = \beta_2 = 2, \gamma_2 = 0$ contra a alternativa de violação de pelo menos uma igualdade. Para as simulações, são fixados os parâmetros de perturbação em $\beta_0 = 5, \beta_3 = 3, \gamma_0 = 3$ e $\gamma_1 = 1$ e as covariáveis x_{li} 's e ω_{li} 's, obtidas de simulações da distribuição uniforme nos intervalos $(0, 20)$ e $(0, 5)$, respectivamente. Amostras são geradas fixando $\sigma^2 = 1$ e $\sigma^2 = 5$, considerando tamanhos de amostra $n = 10, 20$ e 30 e os valores $\lambda = 0, 5, 1$ e 2 . Nossas simulações são, então, conduzidas através dos passos descritos a seguir, sendo que as sub-rotinas utilizadas são do programa NAG:

1. Utilizando a sub-rotina G05DDF, geramos, no computador VAX 8700, 20.000 amostras $y_1, \dots, y_n$, para cada combinação de n, λ e σ^2 . As covariáveis são geradas através da sub-rotina G05DAF e mantidas fixas em todas as simulações.
2. Resolve-se, em cada simulação, o sistema das equações (6.4), sem restrições e com as restrições da hipótese nula, obtendo-se as estimativas de máxima verossimilhança irrestritas e segundo H . Isto é conseguido através das sub-rotinas C05NBF e C05PBF.
3. Calcula-se, para cada amostra gerada, as estatísticas ω e $\omega^* = c^{-1}\omega$, com ω dada por (6.47) e c por (6.48), com $p = 4, q = 3, p_1 = 2$ e $q_1 = 1, h(\cdot)$ calculada por (6.11) e a matriz

de covariâncias Λ estimada sob a hipótese nula. Estes valores são obtidos via as rotinas de operação de matrizes F01AAF (inversão da matriz de informação), F01CRF (transposição de matriz), F01CKF (multiplicação de matriz), F01CLF (multiplicação da matriz por transposta), F01CSF (multiplicação de vetor por matriz simétrica) e F01CNF (transferência de vetor para linha de matriz).

4. Contabilizam-se as proporções de valores de ω e ω^* que superam o $(1 - \alpha)$ -ésimo percentil da distribuição χ_3^2 (com $\alpha = 5\%$ e 1%), em 20.000 ensaios, via a rotina M01AJF do módulo “sorting” do NAG.

Os resultados obtidos estão nas Tabelas 6.1 e 6.2, correspondentes a $\sigma^2 = 1$ e $\sigma^2 = 5$, respectivamente.

Tabela 6.1: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem) para o teste de $H : \beta_1 = \beta_2 = 2, \gamma_2 = 0$, no modelo normal heterocedástico, com $\sigma^2 = 1$.

nível		$\lambda = 0,5$		$\lambda = 0,1$		$\lambda = 2,0$	
n	nominal						
	α (%)	(1)	(2)	(1)	(2)	(1)	(2)
10	5	12,5	6,8	14,2	7,1	16,4	7,1
	1	7,2	2,5	8,5	3,2	9,2	3,5
20	5	8,3	5,4	9,8	6,2	10,1	6,4
	1	3,1	2,0	5,4	2,5	6,1	2,4
30	5	6,1	5,2	7,1	5,8	7,4	5,9
	1	1,9	1,1	2,8	1,3	3,5	1,6

Nota: Os números (1) e (2) referem-se às taxas de rejeição de ω e ω^* , respectivamente.

Tabela 6.2: Valores simulados de $P(\omega \geq x_\alpha)$ e $P(\omega^* \geq x_\alpha)$ (em porcentagem) para o teste de $H : \beta_1 = \beta_2 = 2, \gamma_2 = 0$, no modelo normal heterocedástico, com $\sigma^2 = 5$.

nível		$\lambda = 0,5$		$\lambda = 0,1$		$\lambda = 2,0$	
n	nominal						
	α (%)	(1)	(2)	(1)	(2)	(1)	(2)
10	5	12,6	6,9	14,9	7,4	16,5	8,1
	1	7,2	2,4	8,6	3,1	9,4	3,5
20	5	8,4	5,5	10,0	6,2	10,3	6,5
	1	3,2	2,0	6,0	2,5	6,4	2,3
30	5	6,5	5,5	7,2	5,9	7,5	5,8
	1	2,0	1,2	2,3	1,4	3,6	1,5

Nota: Os números (1) e (2) referem-se às taxas de rejeição de ω e ω^* , respectivamente.

Observando os valores destas tabelas, é bastante evidente que a distribuição da estatística corrigida ω^* é melhor aproximada por χ^2 do que a distribuição de ω , pois em todos os casos simulados as taxas de rejeição de ω^* estão muito mais próximas dos níveis nominais do que as correspondentes de ω , principalmente para tamanhos de amostras não muito grandes. Evidentemente, como esperado, a correção tem menor efeito quando o tamanho de amostra torna-se grande. Entretanto, as aproximações por qui-quadrado para ambas as estatísticas são piores quando λ aumenta, isto é, quando a variância torna-se muito grande. Os estudos de simulação apresentados nas Tabelas 6.1 e 6.2 ilustram que as correções de Bartlett desenvolvidas aqui são, realmente, de grande importância prática.

CAPÍTULO 7

UMA EXTENSÃO ÀS CORREÇÕES DE BARTLETT

7.1 INTRODUÇÃO

Cordeiro e Ferrari (1991) desenvolvem uma metodologia para melhorar estatísticas equivalentes, assintoticamente, à razão de máxima verossimilhança, ou seja, obtêm correções para uma estatística S , cuja distribuição assintótica é qui-quadrado. Esta extensão às correções de Bartlett é denominada de correção de Bartlett generalizada, conforme visto na Seção 1.5, ou, ainda, de correção tipo-Bartlett no caso específico da estatística original ser a estatística escore (vide Seção 1.3.4).

Neste capítulo, o objetivo é apresentar outros resultados sobre aperfeiçoamento de estatísticas, baseados nas idéias desenvolvidas por Cordeiro e Ferrari (1991). Primeiramente, consideramos o aperfeiçoamento de uma estatística que é resultante do quociente entre duas

variáveis aleatórias independentes, ambas convergindo para distribuições qui-quadrados centrais. Na Seção 7.2, mostramos que a distribuição deste quociente converge, até primeira ordem, para a distribuição do quociente entre os dois qui-quadrados de referência envolvidos no quociente original. A expansão assintótica, até ordem n^{-1} , da função distribuição do quociente de interesse é expressa pela soma da função distribuição da variável quociente entre os dois qui-quadrados de referência, mais uma parcela, de ordem n^{-1} , envolvendo combinações lineares dos coeficientes, de mesma ordem, das expansões das duas variáveis qui-quadrados assintóticas envolvidas. A correção que melhora a aproximação da distribuição da variável quociente considerada é apresentada na Seção 7.3. Mostramos, ainda, nesta seção, que é possível utilizarmos a distribuição F de Snedecor de referência, partindo do quociente padronizado entre os dois qui-quadrados assintóticos. Na Seção 7.4, apresentamos um resultado bem mais geral, pertinente a qualquer estatística-tese e distribuição de referência, que contém diversos tipos conhecidos de correção, como, por exemplo, as correções para as estatísticas V (Cordeiro e Perez-Abreu, 1995). Encerramos este capítulo com algumas aplicações dos resultados apresentados na Seção 7.4. Isto é feito na Seção 7.5.

7.2 EXPANSÃO ASSINTÓTICA DA DISTRIBUIÇÃO DO QUOCIENTE DE VARIÁVEIS ASSINTÓTICAS QUI-QUADRADOS

Sejam S_1 e S_2 duas variáveis aleatórias independentes, cujas distribuições assintóticas são qui-quadrados com r_1 e r_2 graus de liberdade, respectivamente. Seja Q a variável resultante do quociente entre estas variáveis, isto é, $Q = S_1/S_2$. Nosso objetivo, nesta seção, é obter uma expansão assintótica da função distribuição de Q , a partir das expansões assintóticas de S_1 e S_2 . Antes, porém, definimos $F(a, b)$ como uma variável aleatória com distribuição F de Snedecor com graus de liberdade a e b . A função distribuição acumulada de $F(a, b)$ é dada por

$$P(F(a, b) \leq x) = 1 - I_z\left(\frac{b}{2}, \frac{a}{2}\right), \quad (7.1)$$

onde $z = b/(b + ax)$ e $I_z(\cdot)$ é a função beta incompleta calculada por

$$I_z(r, s) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^z t^{r-1}(1-t)^{s-1} dt \quad (0 \leq z \leq 1) .$$

Como

$$I_z(r, s) = 1 - I_{1-z}(s, r) , \quad (7.2)$$

vem

$$P(F(a, b) \leq x) = I_{1-z} \left(\frac{a}{2}, \frac{b}{2} \right) , \quad (7.3)$$

com $1 - z = ax/(ax + b)$. Além disso, vale a relação

$$\begin{aligned} I_z(r+1, s) &= I_z(r, s) - \frac{\Gamma(r+s)}{\Gamma(r+1)\Gamma(s)} z^r (1-z)^s \\ &= I_z(r, s) - \frac{z^r (1-z)^s}{r B(r, s)} , \end{aligned} \quad (7.4)$$

onde $B(r, s) = \Gamma(r)\Gamma(s)/\Gamma(r+s)$ é a função beta completa. Através da relação de recorrência em (7.4) prova-se, facilmente, para l inteiro positivo, que

$$I_z(r+l, s) = I_z(r, s) - z^r (1-z)^s \sum_{k=0}^{l-1} \frac{z^k}{(r+k)B(r+k, s)} . \quad (7.5)$$

Como S_1 e S_2 têm distribuições assintóticas qui-quadrados centrais com r_1 e r_2 graus de liberdade, respectivamente, satisfeitas as condições de regularidade, podemos escrever suas funções de distribuição acumulada, até ordem n^{-1} , respectivamente, por

$$F_{S_1}(x) = G_{r_1}(x) + \sum_{l=0}^{\kappa_1} a_l G_{r_1+2l}(x) \quad (7.6)$$

e

$$F_{S_2}(x) = G_{r_2}(x) + \sum_{l=0}^{\kappa_2} b_l G_{r_2+2l}(x) , \quad (7.7)$$

onde $G_r(\cdot)$ representa a função distribuição da χ_r^2 , isto é, $G_r(x) = P(\chi_r^2 \leq x)$, e os coeficientes a_l 's e b_l 's são $O(n^{-1})$, tais que $\sum_{l=1}^{\kappa_1} a_l = \sum_{m=1}^{\kappa_2} b_m = 0$. Seja $Q = S_1/S_2$ e h_Q a função densidade de probabilidade de Q . Queremos obter a expansão assintótica de $F_Q(x)$, a função distribuição

de Q . Como $h_Q(x) = \int_0^\infty f_{S_1}(xy)f_{S_2}(y)ydy$, e utilizando as expansões em (7.6) e (7.7), temos até ordem n^{-1} ,

$$\begin{aligned} h_Q(x) &= \int_0^\infty \left[g_{r_1}(xy) + \sum_{l=0}^{\kappa_1} a_l g_{r_1+2l}(xy) \right] \left[g_{r_2}(y) + \sum_{m=0}^{\kappa_2} b_m g_{r_2+2m}(y) \right] y dy \\ &= \int_0^\infty \left[g_{r_1}(xy)g_{r_2}(y)ydy + \sum_{l=0}^{\kappa_1} a_l g_{r_1+2l}(xy)g_{r_2}(y)ydy \right. \\ &\quad \left. + \sum_{m=0}^{\kappa_2} b_m g_{r_1}(xy)g_{r_2+2m}(y)ydy \right]. \end{aligned}$$

Efetuada-se esta integral, obtemos até ordem n^{-1} ,

$$h_Q(x) = f_{r_1, r_2}(x) + \sum_{l=0}^{\kappa_1} a_l f_{r_1+2l, r_2}(x) + \sum_{m=0}^{\kappa_2} b_m f_{r_1, r_2+2m}(x), \quad (7.8)$$

onde $f_{a,b}(x)$ é a função densidade da variável $G = \chi_a^2/\chi_b^2$, com estas variáveis qui-quadrados supostas independentes. Integrando a densidade (7.8), obtemos para a função distribuição de Q , até ordem n^{-1} ,

$$F_Q(x) = F_{r_1, r_2}(x) + \sum_{l=0}^{\kappa_1} a_l F_{r_1+2l, r_2}(x) + \sum_{m=0}^{\kappa_2} b_m F_{r_1, r_2+2m}(x), \quad (7.9)$$

onde $F_{a,b}(x) = \int_0^x f_{a,b}(x)dx$ é a função distribuição da variável G . Por outro lado, através de (7.3), facilmente mostramos que

$$F_{r_1, r_2}(x) = I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right), \quad (7.10)$$

com $\omega = x/(1+x)$. Alternativamente,

$$F_{r_1, r_2}(x) = 1 - I_{1-\omega} \left(\frac{r_2}{2}, \frac{r_1}{2} \right), \quad (7.11)$$

com $1 - \omega = (1+x)^{-1}$. Logo, por (7.10), temos

$$F_{r_1+2l, r_2}(x) = I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right), \quad F_{r_1, r_2+2m}(x) = I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} + m \right). \quad (7.12)$$

Usando (7.10) e (7.12) em (7.9), vem

$$F_Q(x) = I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) + \sum_{l=0}^{\kappa_1} a_l I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right) + \sum_{m=0}^{\kappa_2} b_m I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} + m \right), \quad (7.13)$$

com $\omega = x/(1+x)$. Mas, por (7.2), podemos, ainda, expressar (7.13) por

$$\begin{aligned}
 F_Q(x) &= I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) + \sum_{l=0}^{\kappa_1} a_l I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right) + \sum_{m=0}^{\kappa_2} b_m \left\{ 1 - I_{1-\omega} \left(\frac{r_2}{2} + m, \frac{r_1}{2} \right) \right\} \\
 &= I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) + a_0 I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) + \sum_{l=1}^{\kappa_1} a_l I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right) \\
 &\quad + 0 - b_0 I_{1-\omega} \left(\frac{r_2}{2}, \frac{r_1}{2} \right) - \sum_{m=1}^{\kappa_2} b_m I_{1-\omega} \left(\frac{r_2}{2} + m, \frac{r_1}{2} \right) \\
 &= I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) (1 + a_0) - b_0 I_{1-\omega} \left(\frac{r_2}{2}, \frac{r_1}{2} \right) \\
 &\quad + \sum_{l=1}^{\kappa_1} a_l I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right) - \sum_{m=1}^{\kappa_2} b_m I_{1-\omega} \left(\frac{r_2}{2} + m, \frac{r_1}{2} \right) .
 \end{aligned}$$

Utilizando, novamente, a relação (7.2) na segunda parcela da última igualdade, e rearranjando termos, resulta

$$\begin{aligned}
 F_Q(x) &= (1 + a_0 + b_0) I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) - b_0 + \sum_{l=1}^{\kappa_1} a_l I_\omega \left(\frac{r_1}{2} + l, \frac{r_2}{2} \right) \\
 &\quad - \sum_{m=1}^{\kappa_2} b_m I_{1-\omega} \left(\frac{r_2}{2} + m, \frac{r_1}{2} \right) .
 \end{aligned} \tag{7.14}$$

Reescrevendo a expressão (7.5) por

$$I_z(r+l, s) = I_z(r, s) - z^r (1-z)^s P_{r,s}^{(l-1)}(z) , \tag{7.15}$$

onde $P_{r,s}^{(l-1)}(z)$ é um polinômio de grau $l-1$ em z , indexado por r e s , isto é, $P_{r,s}^{(l-1)}(x) = \sum_{k=0}^{l-1} \frac{z^k}{(r+k)B(r+k, s)}$, podemos escrever (7.14) por

$$\begin{aligned}
 F_Q(x) &= (1 + a_0 + b_0) I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) - b_0 + \sum_{l=1}^{\kappa_1} a_l \left\{ I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) - \omega^{\frac{r_1}{2}} (1-\omega)^{\frac{r_2}{2}} P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \right\} \\
 &\quad - \sum_{m=1}^{\kappa_2} b_m \left\{ I_{1-\omega} \left(\frac{r_2}{2}, \frac{r_1}{2} \right) - (1-\omega)^{\frac{r_2}{2}} \omega^{\frac{r_1}{2}} P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(m-1)}(1-\omega) \right\} \\
 &= (1 + a_0 + b_0) I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) - b_0 + (-a_0) I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) - \omega^{\frac{r_1}{2}} (1-\omega)^{\frac{r_2}{2}} \sum_{l=1}^{\kappa_1} a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \\
 &\quad + b_0 I_{1-\omega} \left(\frac{r_2}{2}, \frac{r_1}{2} \right) + \omega^{\frac{r_1}{2}} (1-\omega)^{\frac{r_2}{2}} \sum_{m=1}^{\kappa_2} b_m P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(m-1)}(1-\omega) .
 \end{aligned}$$

Finalmente, após simplificações na última igualdade, resulta que, até ordem n^{-1} , a função distribuição da variável $Q = S_1/S_2$ é dada por:

$$F_Q(x) = I_\omega \left(\frac{r_1}{2}, \frac{r_2}{2} \right) + \omega^{\frac{r_1}{2}} (1-\omega)^{\frac{r_2}{2}} \left\{ \sum_{m=1}^{\kappa_2} b_m P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(m-1)}(1-\omega) - \sum_{l=1}^{\kappa_1} a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \right\} , \tag{7.16}$$

onde $P_{a,b}^{(l-1)}(x) = \sum_{k=0}^{l-1} \frac{x^k}{(a+k)B(a+k,b)}$ e $\omega = \frac{x}{1+x}$.

Observando a expansão em (7.16) podemos concluir, omitindo-se termos de ordem inferior a n^{-1} , que a função distribuição da variável $Q = S_1/S_2$, onde S_1 e S_2 são variáveis assintóticas qui-quadrados, independentes, com r_1 e r_2 graus de liberdade, respectivamente, é calculada pela soma da função distribuição da variável $G = \chi_{r_1}^2/\chi_{r_2}^2$, com $\chi_{r_1}^2$ e $\chi_{r_2}^2$ independentes, mais um termo de ordem n^{-1} envolvendo os coeficientes das expansões assintóticas, até esta ordem, das funções distribuição das variáveis S_1 e S_2 .

Vamos supor, agora, que $\kappa_1 = \kappa_2 = t$. Esta suposição é fraca, pois na grande maioria das aplicações isto acontece com $t \leq 3$. Neste caso, a expressão (7.16) torna-se

$$F_Q(x) = I_\omega\left(\frac{r_1}{2}, \frac{r_2}{2}\right) + \omega^{\frac{r_1}{2}}(1-\omega)^{\frac{r_2}{2}} \sum_{l=1}^t \left\{ b_l P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(l-1)}(1-\omega) - a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \right\}, \quad (7.17)$$

onde $\omega = \frac{x}{1+x}$ e $1-\omega = \frac{1}{1+x}$.

7.3 APERFEIÇOAMENTO DO QUOCIENTE ENTRE QUI-QUADRADOS ASSINTÓTICOS

O desenvolvimento utilizado por Cordeiro e Ferrari (1991) sugere que a estatística Q pode ser aperfeiçoada pela estatística $Q^* = Q \left(1 - \sum_{i=1}^t c_i Q^{i-1}\right)$, cuja função distribuição é expandida, segundo Cox e Reid (1987), até ordem n^{-1} , como

$$F_{Q^*}(x) = F_Q(x) + f_Q(x) \sum_{i=1}^t c_i x^i, \quad (7.18)$$

onde os c_i 's são coeficientes de ordem n^{-1} a serem determinados. Uma vez que os c_i 's são $O(n^{-1})$ e, até primeira ordem, Q tem a mesma distribuição de G , chega-se a

$$F_{Q^*}(x) = F_Q(x) + f_G(x) \sum_{i=1}^t c_i x^i, \quad (7.19)$$

até ordem n^{-1} . Agora, combinando (7.17) e (7.19) vem, até esta ordem,

$$F_{Q^*}(x) = F_G(x) + \omega^{\frac{r_1}{2}}(1-\omega)^{\frac{r_2}{2}} \sum_{l=1}^t \left\{ b_l P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(l-1)}(1-\omega) - a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \right\} + f_G(x) \sum_{i=1}^t c_i x^i, \quad (7.20)$$

onde $G = \chi_{r_1}^2 / \chi_{r_2}^2$, com função distribuição $F_G(x)$ e função densidade $f_G(x)$.

Seguindo Cordeiro e Ferrari (1991), a idéia é determinar os c_i 's, de ordem n^{-1} , em função dos coeficientes a_l 's e b_l 's, de tal forma que Q^* tenha a mesma distribuição de G , até esta ordem e sob a hipótese nula. A função densidade de G pode ser expressa por

$$f_G(x) = \frac{x^{\frac{r_1}{2}-1}}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right) (1+x)^{\frac{r_1+r_2}{2}}},$$

onde $B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)$ é a função beta completa, isto é, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Como $\omega = x/(1+x)$, vem $x = \omega/(1-\omega)$ e $(1+x) = 1/(1-\omega)$ e, então,

$$f_G(x) = \frac{\omega^{\frac{r_1}{2}-1}(1-\omega)^{\frac{r_1}{2}}(1-\omega)^{\frac{r_2}{2}}}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right) (1-\omega)^{\frac{r_1}{2}-1}} = \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \frac{(1-\omega)}{\omega} \omega^{\frac{r_1}{2}} (1-\omega)^{\frac{r_2}{2}}. \quad (7.21)$$

Através de (7.20) e (7.21), $F_{Q^*}(x) = F_G(x)$, até ordem n^{-1} , se, e somente se,

$$\sum_{l=1}^t \left\{ b_l P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(l-1)}(1-\omega) - a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) \right\} + \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \frac{(1-\omega)}{\omega} \sum_{i=1}^t \frac{c_i \omega^i}{(1-\omega)^i} = 0.$$

A última igualdade pode, ainda, ser expressa por

$$\sum_{l=1}^t \left\{ b_l P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(l-1)}(1-\omega) - a_l P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) + \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \frac{\omega^{l-1}}{(1-\omega)^{l-1}} c_l \right\} = 0. \quad (7.22)$$

Mas,

$$P_{\frac{r_1}{2}, \frac{r_2}{2}}^{(l-1)}(\omega) = \sum_{k=0}^{l-1} \frac{\omega^k}{B\left(\frac{r_1}{2} + k, \frac{r_2}{2}\right) \left(\frac{r_1}{2} + k\right)} \quad \text{e} \quad P_{\frac{r_2}{2}, \frac{r_1}{2}}^{(l-1)}(1-\omega) = \sum_{k=0}^{l-1} \frac{(1-\omega)^k}{B\left(\frac{r_2}{2} + k, \frac{r_1}{2}\right) \left(\frac{r_2}{2} + k\right)}$$

e, então, a equação (7.22) resulta

$$\begin{aligned} & \sum_{l=1}^t \left\{ b_l \sum_{k=0}^{l-1} \frac{(1-\omega)^k}{B\left(\frac{r_2}{2} + k, \frac{r_1}{2}\right) \left(\frac{r_2}{2} + k\right)} - a_l \sum_{k=0}^{l-1} \frac{\omega^k}{B\left(\frac{r_1}{2} + k, \frac{r_2}{2}\right) \left(\frac{r_1}{2} + k\right)} \right\} \\ & + \sum_{l=1}^t \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \left(\frac{\omega}{1-\omega} \right)^{l-1} c_l = 0. \end{aligned}$$

Rearranjando somatórios, a equação anterior pode ser escrita como

$$\sum_{k=0}^{t-1} \left(\sum_{l=k+1}^t b_l \right) \frac{(1-\omega)^k}{B\left(\frac{r_2}{2} + k, \frac{r_1}{2}\right) \left(\frac{r_2}{2} + k\right)} - \left(\sum_{l=k+1}^t a_l \right) \frac{\omega^k}{B\left(\frac{r_1}{2} + k, \frac{r_2}{2}\right) \left(\frac{r_1}{2} + k\right)} + \sum_{l=1}^t \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \left(\frac{\omega}{1-\omega} \right)^{l-1} c_l = 0 .$$

Sejam $B_k = \sum_{l=k+1}^t b_l$ e $A_k = \sum_{l=k+1}^t a_l$. Logo, escrevemos

$$\sum_{k=0}^{t-1} \left\{ \frac{(1-\omega)^k B_k}{B\left(\frac{r_2}{2} + k, \frac{r_1}{2}\right) \left(\frac{r_2}{2} + k\right)} - \frac{A_k \omega^k}{B\left(\frac{r_1}{2} + k, \frac{r_2}{2}\right) \left(\frac{r_1}{2} + k\right)} \right\} + \sum_{k=0}^{t-1} \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \left(\frac{\omega}{1-\omega} \right)^k c_{k+1} = 0 .$$

Finalmente, obtém-se o seguinte sistema, que permite calcular os c_i 's em termos dos a_i 's e b_i 's:

$$\sum_{k=0}^{t-1} \left\{ B'_k (1-\omega)^k - A'_k \omega^k + c'_{k+1} \left(\frac{\omega}{1-\omega} \right)^k \right\} = 0 , \quad (7.23)$$

onde $B'_k = \frac{B_k}{B\left(\frac{r_2}{2} + k, \frac{r_1}{2}\right) \left(\frac{r_2}{2} + k\right)}$, $A'_k = \frac{A_k}{B\left(\frac{r_1}{2} + k, \frac{r_2}{2}\right) \left(\frac{r_1}{2} + k\right)}$ e $c'_{k+1} = \frac{c_{k+1}}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)}$. Quando $t = 1$ temos o quociente entre duas estatísticas da razão de máxima verossimilhança. O caso $t = 3$ corresponde ao quociente entre duas estatísticas escore. O caso $t = 1$ é o mais comum nos modelos de regressão não-linear com o uso nas tabelas de análise de variância, pois temos quociente de somas de quadrados, que são autênticas razões de máxima verossimilhança. A título de ilustração, vamos encontrar os coeficientes quando $t = 1$. Neste caso, temos somente um coeficiente a determinar, c , e o sistema de equações (7.23) produz $B'_0 - A'_0 + c' = 0$, implicando $c' = A'_0 - B'_0$, onde $A'_0 = \frac{A_0}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right) \frac{r_1}{2}} = \frac{a_1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right) \frac{r_1}{2}}$, $B'_0 = \frac{B_0}{B\left(\frac{r_2}{2}, \frac{r_1}{2}\right) \frac{r_2}{2}} = \frac{b_1}{B\left(\frac{r_2}{2}, \frac{r_1}{2}\right) \frac{r_2}{2}}$.

Assim, $c' = \frac{a_1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right) \frac{r_1}{2}} - \frac{b_1}{B\left(\frac{r_2}{2}, \frac{r_1}{2}\right) \frac{r_2}{2}}$, que resulta

$$c = 2 \left(\frac{a_1}{r_1} - \frac{b_1}{r_2} \right) , \quad (7.24)$$

ou seja, o fator de correção tipo-Bartlett, que aperfeiçoa o quociente Q entre duas estatísticas da razão de máxima verossimilhança, é calculado por

$$1 - c = 1 - 2 \left(\frac{a_1}{r_1} - \frac{b_1}{r_2} \right) . \quad (7.25)$$

Denominando por $1 - c_1$ e $1 - c_2$ as correções de Bartlett das estatísticas da razão de verossimilhança do numerador e denominador, respectivamente, temos, utilizando a expressão (16) de Cordeiro e Ferrari (1991), que $1 - c_1 = 1 - 2a_1/r_1$ e $1 - c_2 = 1 - 2b_1/r_2$, ou seja, $c_1 = 2a_1/r_1$ e $c_2 = 2b_1/r_2$. Assim, é fácil verificar que

$$1 - c = 1 - [1 - c_2 - (1 - c_1)] . \quad (7.26)$$

A equação (7.26) mostra que a correção tipo-Bartlett para aperfeiçoar o quociente de estatísticas da razão de verossimilhança iguala à diferença das correções de Bartlett associadas às estatísticas do denominador e do numerador.

Para o caso $t = 3$, temos o quociente entre duas estatísticas escore, e os coeficientes c_1, c_2, c_3 , da expansão (7.18) são determinados pelo sistema de equações dado por (7.23).

Vemos da expansão (7.20), para a função distribuição da variável modificada Q^* , que a variável de referência é $G = \chi_{r_1}^2 / \chi_{r_2}^2$, ou seja, $P(Q^* \leq x) = P(G \leq x) + O(n^{-2})$. Definimos $R = \frac{S_1/r_1}{S_2/r_2} = \frac{r_2}{r_1}Q$ e $F(r_1, r_2) = \frac{r_2}{r_1}G$. Então, $F_R(x) = P(R \leq x) = P(Q \leq r_1x/r_2) = F_Q(r_1x/r_2)$ e $F_G(x) = P(G \leq x) = P(F(r_1, r_2) \leq r_2x/r_1) = F_F(r_2x/r_1)$. Logo, a função distribuição de G , $F_G(x)$, pode ser facilmente calculada da distribuição acumulada da F de Snedecor. O teste baseado em Q^* deve ser referenciado à distribuição de G . Podemos fazer o teste em termos da distribuição $F(r_1, r_2)$ de referência, modificando a estatística aperfeiçoada Q^* . Temos,

$$Q^* = \frac{r_1}{r_2}R \left\{ 1 - \sum_{i=1}^t c_i \left(\frac{r_1}{r_2} \right)^{i-1} R^{i-1} \right\}$$

e, então,

$$R^* = \frac{r_2}{r_1}Q^* = R \left\{ 1 - \sum_{i=1}^t c_i \left(\frac{r_1}{r_2} \right)^{i-1} R^{i-1} \right\} . \quad (7.27)$$

De $P(Q^* \leq x) = P(G \leq x) + O(n^{-2})$ vem

$$P(R^* \leq r_2x/r_1) = P(F(r_1, r_2) \leq r_2x/r_1) + O(n^{-2}) ,$$

que implica

$$P(R^* \leq y) = P(F(r_1, r_2) \leq y) + O(n^{-2}) .$$

Assim, o teste compara a estatística R^* , dada por (7.27), com a distribuição F de Snedecor com r_1 e r_2 graus de liberdade.

7.4 UMA EXTENSÃO DAS CORREÇÕES

Na seção anterior apresentamos a correção tipo-Bartlett para uma variável expressa pelo quociente entre duas variáveis aleatórias independentes, que convergem para distribuições qui-quadrados. Agora, vamos estender as correções de Bartlett generalizadas a estatísticas-teste quaisquer, cuja distribuição nula assintótica é arbitrária. Esta extensão baseia-se no resultado de Cox e Reid (1987), apresentado a seguir. Sejam U e W variáveis aleatórias com densidades marginais contínuas tendo suporte em $\mathbb{R}$, tais que $U = O_p(1)$ e $W = O_p(n^{-1})$. Considere-se a seqüência de variáveis aleatórias $\{S_n\}_{n \geq 1}$, tais que $S_n = U + W + O_p(n^{-3/2})$. Então, segundo condições gerais de regularidade, a função distribuição $F_n(x)$ de S_n tem expansão assintótica

$$F_n(x) = F_U(x) - f_U(x)E(W|U = x) + O(n^{-3/2}), \quad (7.28)$$

onde $F_U(x)$ e $f_U(x) = \frac{d}{dx}F_U(x)$ são as funções distribuição e densidade de U , respectivamente. Mostramos, agora, como obter uma estatística aperfeiçoada em generalidade, cuja distribuição é idêntica àquela de referência, até termos de ordem n^{-1} .

Seja S uma estatística qualquer, cuja distribuição assintótica, segundo alguma hipótese nula, é $F_R(x)$. Supomos $F_S(x) = F_R(x) + O(n^{-1})$. Objetivamente, queremos encontrar uma estatística aperfeiçoada S^* tal que, segundo esta hipótese, $F_{S^*}(x) = F_R(x) + O(n^{-2})$, ou seja, a variável modificada S^* coincide com a variável de referência até ordem n^{-1} . Considere que o termo de ordem n^{-1} , digamos $A_R(x)$, da expansão assintótica da distribuição nula da estatística S seja conhecido, ou seja,

$$F_S(x) = F_R(x) + A_R(x) + O(n^{-2}), \quad (7.29)$$

onde $F_R(x)$ e $A_R(x)$ são termos $O(1)$ e $O(n^{-1})$, respectivamente. Observe-se que, no caso

de estatísticas S que convergem, segundo a hipótese nula, para a distribuição qui-quadrado central com q graus de liberdade, a parcela $A_R(x)$ coincide com uma combinação linear de $(k+1)$ distribuições qui-quadrados com graus de liberdade $q, q+2, \dots, q+2k$, dado por (1.26).

Demonstramos, agora, que conhecendo-se $A_R(x)$, podemos definir uma estatística modificada S^* cuja distribuição nula, até ordem $O(n^{-1})$, coincide com $F_R(x)$. Seja a estatística

$$S^* = S\{1 - P(S)\} , \quad (7.30)$$

onde $P(S) = O_p(n^{-1})$. O termos entre chaves em (7.30) é uma *forma estendida da correção de Bartlett*, como função da própria estatística não-modificada S . O objetivo aqui é, então, determinar a forma da função $P(\cdot)$. Aplicando a expansão assintótica (7.28) para S^* até $O(n^{-1})$, encontramos

$$F_{S^*}(x) = F_S(x) + f_S(x)xP(x)$$

e, portanto, combinando esta equação com a equação (7.29), concluimos que S^* tem distribuição nula, até $O(n^{-1})$, igual a $F_R(x)$, se (supondo $f_R(x) > 0$)

$$P(x) = -\frac{A_R(x)}{xf_R(x)} . \quad (7.31)$$

Logo, admitindo $P(S=0) = 0$, o teste estatístico melhorado é baseado em

$$S^* = S \left\{ 1 + \frac{A_R(S)}{Sf_R(S)} \right\} . \quad (7.32)$$

A expressão entre chaves em (7.32) estende as correções de Bartlett generalizadas da equação (1.28) da Seção 1.5, pois, naquele caso, a distribuição nula assintótica é χ_q^2 , ou seja, $f_R(x)$ iguala a densidade $f_q(x)$ da χ_q^2 e pode-se demonstrar, facilmente, de (1.26), usando a fórmula de recorrência entre distribuições acumuladas de qui-quadrados defasadas por dois graus de liberdade, dada em (1.27), que

$$A_R(x) = -f_q(x) \sum_{i=1}^k c_i x^i .$$

Assim, (7.32) implica uma correção da forma expressa em (1.23), ou em (1.28). No caso das estatísticas V , onde a distribuição nula assintótica é uma combinação linear de qui-quadrados independentes, cada uma com um grau de liberdade, o resultado (7.32) generaliza o Teorema 1 de Cordeiro e Pérez-Abreu (1995).

É importante destacarmos que a fórmula (7.32) pode ser usada para corrigir estatísticas até ordem $n^{-1/2}$, quando na expansão (7.29) o termo $A_R(x)$ é de ordem $n^{-1/2}$ e não de ordem n^{-1} . Neste caso, o resto desta expansão é de ordem n^{-1} . Em outras palavras, a equação (7.32) permite corrigir estatísticas em duas situações: (i) até ordem $n^{-1/2}$ e, (ii) até ordem n^{-1} . No caso (ii), como vimos, quando não há termos de ordem $n^{-1/2}$ na expansão assintótica (7.29).

O resultado desenvolvido aqui é bastante poderoso para o aperfeiçoamento de classes amplas de estatísticas pois, para que correções sejam obtidas através da fórmula (7.32), é necessário apenas conhecer a expansão assintótica da distribuição da estatística original, sob a hipótese nula até ordem $n^{-1/2}$, ou até ordem n^{-1} quando o termo de ordem $n^{-1/2}$ é zero. Uma generalização da equação (7.32) pode ser obtida para corrigir qualquer estatística-teste S , função de variáveis aleatórias contínuas $X_1, X_2, \dots, X_n$, não necessariamente independentes e identicamente distribuídas, tendo aproximação de primeira ordem aperfeiçoada através da expansão da função de distribuição desta estatística, até ordem $n^{-3/2}$, incluindo nesta expansão um termo de ordem $n^{-1/2}$. Esta generalização tem grande interesse quando a distribuição de referência é a normal reduzida. Esta pesquisa está sendo desenvolvida por G. M. Cordeiro e S. L. P. Ferrari.

7.5 ALGUMAS APLICAÇÕES

Finalizando, mostramos a seguir algumas aplicações da técnica de obtenção de estatísticas aperfeiçoadas desenvolvida na seção anterior. O primeiro exemplo já foi considerado anteriormente, sendo revisto aqui a título de ilustração.

Exemplo 7.1

Como um primeiro exemplo, consideramos $A_R(x) = \sum_{l=0}^k a_l F_{q+2l}(x)$, que corresponde ao caso de estatísticas que convergem para a distribuição qui-quadrado, estudada na Seção 1.5, cuja expansão da função distribuição, até ordem n^{-1} , é escrita por (1.26). Usando as fórmulas de recorrência em (1.27) e a equação (7.32) chega-se à fórmula $S^* = S\{1 - \sum_{i=1}^k c_i S^{i-1}\}$ (Cordeiro e Ferrari, 1991), onde $c_i = 2\mu_i'^{-1} \sum_{l=i}^k a_l$ e μ_i' é o i -ésimo momento central da variável χ_q^2 , isto é, $\mu_i' = E\{(\chi_q^2)^i\}$. Conforme já mencionado na Seção 1.5, o caso $k = 1$ é a clássica correção de Bartlett.

Exemplo 7.2

Uma segunda situação, que é de grande interesse refere-se às estatísticas cuja aproximação de primeira ordem é a normal reduzida. Considere a estatística

$$S = \left(\sum_{i=1}^n X_i - n\delta_1 \right) / (\sqrt{n}\delta_2^{1/2}),$$

onde os X_i 's são variáveis aleatórias contínuas independentes e identicamente distribuídas, com cumulantes padronizados $\rho_r = \delta_r / \delta_2^{r/2}$ ($r \geq 1$), sendo δ_r o r -ésimo cumulante (não padronizado) de X_i . Observe-se que, usando a propriedade de que o cumulante de uma soma de variáveis aleatórias independentes iguala à soma dos cumulantes dessa variável, os cumulantes da variável S são $\kappa_{S_r} = \rho_r n^{1-r/2}$, para $r > 1$ e $\kappa_{S_1} = 0$. Tomando-se a distribuição de referência R como $N(0, 1)$, já que pelo teorema do limite central a variável S tem distribuição assintótica normal reduzida, temos a seguinte expansão bastante conhecida na literatura (vide, por exemplo, entre outros, o livro de McCullagh, 1987, fórmula (5.12))

$$F_S(x) = \Phi(x) - \phi(x) \left\{ \frac{\rho_3 h_2(x)}{6\sqrt{n}} + O(n^{-1}) \right\}, \quad (7.33)$$

onde $\Phi(\cdot)$ e $\phi(\cdot)$ são, respectivamente, a função distribuição acumulada e a função densidade da normal padrão e $h_2(x) = x^2 - 1$. Combinando (7.33) com (7.29) segue $A_R(x) =$

$-\phi(x)\rho_3h_2(x)/(6\sqrt{n})$ de ordem $n^{-1/2}$, e levando em (7.32) vem a estatística corrigida

$$S^* = S - \frac{\rho_3(S^2 - 1)}{6n^{1/2}}. \quad (7.34)$$

Note que $S^* \sim N(0, 1) + O_p(n^{-1})$ e, portanto, (7.34) corresponde à clássica transformação de Cornish-Fisher (vide Kendall e Stuart, 1969 e Hill e Davis, 1968) que normaliza a distribuição de S^* até $O_p(n^{-1/2})$.

Exemplo 7.3

Este exemplo refere-se à correção da distribuição da função escore padronizada até ordem $n^{-1/2}$, no contexto uniparamétrico. Considere um problema regular uniparamétrico, cujo logaritmo da função de verossimilhança total para θ é $\ell(\theta)$. A função escore total é $U(\theta) = d\ell(\theta)/d\theta$, cuja esperança é nula e cuja variância iguala à informação total de Fisher $\kappa_{\theta,\theta}$. A função escore padronizada $S = U(\theta)/\kappa_{\theta,\theta}^{1/2}$ é, freqüentemente, usada para construir intervalos de confiança para θ com erro de ordem $n^{-1/2}$. Nosso objetivo é construir intervalos de confiança para θ corretos até $O(n^{-1/2})$, isto é, com erro $O(n^{-1})$. Como antes, sejam κ_{S_r} e κ_{U_r} os cumulantes de S e U , respectivamente. Como S é uma soma de n contribuições individuais (vide Seção 4.2) dividida pela raiz quadrada da informação total, então S satisfaz a soma padronizada definida no Exemplo 7.2, com seus cumulantes dados por $\kappa_{S_r} = \kappa_{U_r} \kappa_{\theta,\theta}^{-r/2}$. Em particular, das condições de regularidade, $\kappa_{S_1} = 0$ e $\kappa_{S_2} = 1$. Assim, este caso é uma aplicação direta da expressão (7.34) para $S = U(\theta)\kappa_{\theta,\theta}^{-1/2}$, com o cumulante $\rho_3/\sqrt{n}$ desta expressão substituído por $\kappa_{S_3} = \kappa_{U_3}/\kappa_{\theta,\theta}^{3/2}$. O terceiro cumulante de U pode ser calculado através da identidade de Bartlett (vide Seção 1.3.1) $\kappa_{U_3} = 2\kappa_{\theta\theta\theta} + 3\kappa_{\theta\theta}^{(\theta)}$. Logo, da equação (7.34), chegamos à fórmula da função escore padronizada melhorada até $O(n^{-1/2})$

$$S^* = \kappa_{\theta,\theta}^{-1/2} \left\{ U - \frac{\kappa_{U_3}}{6\kappa_{\theta,\theta}^2} (U^2 - \kappa_{\theta,\theta}) \right\}. \quad (7.35)$$

O erro da fórmula (7.35) é $O_p(n^{-1})$.

Exemplo 7.4

Passamos, agora, a corrigir a distribuição da estimativa de máxima verossimilhança $\hat{\theta}$ do parâmetro único θ das distribuições uniparamétricas num problema regular. Consideramos a estatística padronizada $S = (\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$. Observe-se que S está padronizada para a distribuição de referência ser a normal reduzida, pois a estimativa de máxima verossimilhança $\hat{\theta}$ converge em distribuição para a distribuição $N(\theta, \kappa_{\theta,\theta}^{-1/2})$, em amostras de tamanho grande. Isto posto, escrevemos a expansão da distribuição acumulada de S até ordem $n^{-1/2}$ (vide livro de McCullagh, 1987, equação (5.6)) por

$$F_S(x) = \Phi(x) - \phi(x) \left\{ \frac{6\eta_1 + \eta_3 h_2(x)}{6} \right\}, \quad (7.36)$$

onde $\Phi(\cdot)$, $\phi(\cdot)$, $h_2(\cdot)$ estão definidos no Exemplo 7.2, $\eta_1 = \kappa_{\theta,\theta}^{1/2} b_1(\theta) + O(n^{-3/2})$ e $\eta_3 = \rho_{3_{\hat{\theta}}} + O(n^{-3/2})$, sendo $b_1(\theta)$ o viés de ordem n^{-1} da estimativa de máxima verossimilhança $\hat{\theta}$, isto é, $E(\hat{\theta}) = \theta + b_1(\theta) + O(n^{-2})$ e $\rho_{3_{\hat{\theta}}}$ o terceiro cumulante padronizado de $\hat{\theta}$, de ordem $n^{-1/2}$. Assim, η_1 e η_3 são quantidades calculadas até ordem $n^{-1/2}$. Identificando a expressão (7.36) com (7.29), obtemos o termo $A_R(x) = -\phi(x)(6\eta_1 + \eta_3 h_2(x))/6$, de ordem $n^{-1/2}$. Levando-se esta equação em (7.32) resulta, depois de alguma álgebra, que até ordem $n^{-1/2}$

$$S^* = S + \frac{\eta_3 - 6\eta_1}{6} - \frac{\eta_3}{6} S^2, \quad (7.37)$$

onde $S = (\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$. Assim, a estimativa de máxima verossimilhança padronizada corrigida S^* é um polinômio de segundo grau na estimativa de máxima verossimilhança padronizada S , até ordem $n^{-1/2}$. Observe-se que $S^* \sim N(0, 1) + O_p(n^{-1})$ enquanto $S \sim N(0, 1) + O_p(n^{-1/2})$. Note que podemos expressar (7.37) por $S^* = S + \sum_{i=0}^2 a_i S^i$, onde $a_0 = (\eta_3 - 6\eta_1)/6$, $a_1 = 0$ e $a_2 = -\eta_3/6$. Para calcular η_1 e η_3 é necessário conhecer as expressões de $b_1(\theta)$ e $\rho_{3_{\hat{\theta}}}$. A fórmula de $b_1(\theta)$ é obtida de Cordeiro e McCullagh (1991, equação (4.1)) fazendo todos os índices iguais a θ e, o terceiro cumulante padronizado de $\hat{\theta}$ é dado por Shenton e Bowman (1977, equação (2.31a)). Assim, depois de alguma álgebra, vem

$$b_1(\theta) = \frac{\kappa_{\theta\theta\theta}^{(\theta)} - \frac{1}{2}\kappa_{\theta\theta\theta}}{\kappa_{\theta,\theta}^2} \quad \text{e} \quad \rho_{3_{\hat{\theta}}} = \frac{3\kappa_{\theta\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}}{\kappa_{\theta,\theta}^{3/2}}.$$

Substituindo as fórmulas acima nas expressões para η_1 e η_3 , até ordem $n^{-1/2}$, resulta

$$\eta_1 = \frac{2\kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}}{2\kappa_{\theta,\theta}^{3/2}} \ , \quad \eta_3 = \frac{3\kappa_{\theta\theta}^{(\theta)} - \kappa_{\theta\theta\theta}}{\kappa_{\theta,\theta}^{3/2}} \ .$$

Considera-se aqui notação idêntica àquela usada para encontrar a correção de Bartlett nos modelos exponenciais uniparamétricos no Capítulo 4. Intervalos de confiança aperfeiçoados para θ , com erro $O(n^{-1})$ são facilmente deduzidos de S^* em (7.37), baseados na normal reduzida, ao invés de serem construídos da aproximação normal para a estimativa de máxima verossimilhança de θ , como é feito rotineiramente, e cujo erro é de ordem $O(n^{-1/2})$.

CONCLUSÃO

O enfoque principal deste trabalho foi obter, para alguns modelos especiais, o aperfeiçoamento da estatística do teste da razão de verossimilhança até ordem n^{-1} , através do cálculo do fator de correção de Bartlett. Outros pontos foram também desenvolvidos. Resumidamente, os principais resultados encontrados foram:

- (i) Fórmulas matriciais para fatores de correção de Bartlett na classe dos modelos de regressão linear heterocedásticos, em que a estrutura de variância-covariância dos erros é expressa por uma função de um vetor de parâmetros desconhecidos, e aplicação das fórmulas obtidas para alguns testes de hipóteses particulares. Ainda para esta classe de modelos, encontramos o viés até ordem n^{-1} do estimador de máxima verossimilhança para o vetor de parâmetros na estrutura de variância-covariância. Nossas fórmulas apresentam vantagens sob o ponto de vista numérico e simbólico já que requerem apenas operações elementares com matrizes e vetores, podendo ser implementadas facilmente em algum pacote computacional algébrico;
- (ii) Fórmulas para fatores de correção de Bartlett para modelos da família exponencial uniparamétrica e biparamétrica com parâmetros ortogonais. As expressões resultantes são bastante simples, envolvendo somente operações triviais com certas funções e suas derivadas. Na família

exponencial uniparamétrica, obtemos correções de Bartlett em 30 distribuições de grande interesse na Estatística. Através do método de reparametrização de Cox e Reid (1987) obtemos os parâmetros ortogonais na família exponencial biparamétrica, para algumas distribuições em que esta condição não se verifica e, então, aplicamos nossos resultados;

(iii) Extensão das correções tipo-Bartlett a outras estatísticas. Obtemos o aperfeiçoamento do quociente de variáveis aleatórias independentes que convergem para distribuições qui-quadrados e de variáveis aleatórias quaisquer com distribuição assintótica de primeira ordem arbitrária.

É claro que, dada a abrangência desta área de pesquisa, inúmeros tópicos poderão ser desenvolvidos a partir dos resultados desta tese. Alguns deles são listados a seguir:

- (i) Comparação, via simulação, do desempenho da estatística da razão de verossimilhança corrigida com a usual, para os modelos estudados no Capítulo 2;
- (ii) Obtenção do aperfeiçoamento da estatística da razão de verossimilhança para o teste de restrições lineares $H : A\beta = a$, onde A é uma matriz $m \times p$ de constantes conhecidas e a é um vetor $n \times 1$ de constantes conhecidas nos modelos estudados no Capítulo 6. Este teste é rotineiramente usado na área de econometria;
- (iii) Reconciliação da fórmula (2.9) com as expressões (3.6) - (3.8) obtidas por Cordeiro (1993b), que estuda o caso particular do modelo de regressão heterocedástico multiplicativo;
- (iv) Aperfeiçoamento de testes score em modelos de dispersão;
- (v) Explorar outras aplicações na área de econometria dentro do contexto do Capítulo 6 e dar uma interpretação estatística às correções que encontramos;
- (vi) Aperfeiçoamento da estatística do teste da razão de verossimilhança em modelos normais multivariados não-lineares;

(vii) Encontrar a expressão do viés de ordem superior a n^{-1} do estimador de máxima verossimilhança para o vetor de parâmetros β , do modelo considerado no Capítulo 3.

APÊNDICE 1

Apresentamos neste apêndice a obtenção direta, sem usar a fórmula de Lawley (1956), do valor esperado da estatística da razão de verossimilhança, até ordem n^{-1} , para o teste do vetor de parâmetros β , nos modelos normais lineares, sendo γ o parâmetro de incômodo, unidimensional e ortogonal ao vetor de parâmetros β .

Consideremos $Y_1, \dots, Y_n$ n variáveis aleatórias independentes, sendo que Y_l tem função densidade de probabilidade dada por:

$$\pi(y_l; \beta, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{(y_l - x_l^\top \beta)^2}{2\sigma^2} \right\},$$

onde y_l é o valor observado de Y_l , e $x_l^\top = (x_{l1}, \dots, x_{lp})$ é um vetor de valores fixados de p variáveis auxiliares para a l -ésima observação.

A função log-verossimilhança total para o vetor de parâmetros desconhecidos $\theta = (\beta^\top, \sigma^2)^\top$, dado o vetor de observações $(y_1, \dots, y_n)$, é

$$\ell = \ell(\theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{l=1}^n (y_l - x_l^\top \beta)^2,$$

que é maximizada em $\hat{\theta} = (\hat{\beta}^\top, \hat{\sigma}^2)^\top$, o estimador de máxima verossimilhança irrestrito de θ , obtido resolvendo o sistema de equações (2.3). É fácil mostrar que $\hat{\beta} = (X^\top X)^{-1} X^\top y$ e $\hat{\sigma}^2 = \sum_{l=1}^n (y_l - x_l^\top \hat{\beta})^2 / n$, onde X é a matriz modelo.

A hipótese nula de interesse é $H_3 : \beta = \beta^{(0)}$ contra $A_3 : \beta \neq \beta^{(0)}$, sendo $\beta^{(0)}$ um vetor de valores especificados e σ^2 o parâmetro de perturbação. Seja $\tilde{\sigma}^2$ o estimador de máxima

verossimilhança sob a hipótese nula H_3 de σ^2 . A estatística da razão de verossimilhança do teste é

$$\omega_3 = -2 \log \lambda = 2 \{ \ell(\hat{\beta}^\top, \hat{\sigma}^2) - \ell(\beta^{(0)\top}, \tilde{\sigma}^2) \} ,$$

onde $\tilde{\sigma}^2 = \frac{\sum_{l=1}^n (y_l - x_l^\top \beta^{(0)})^2}{n}$. Mas,

$$\begin{aligned} \ell(\hat{\beta}, \hat{\sigma}^2) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2} \sum_{l=1}^n (y_l - x_l^\top \hat{\beta})^2 \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2} \end{aligned}$$

e

$$\ell(\beta^{(0)}, \tilde{\sigma}^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \tilde{\sigma}^2 - \frac{n}{2} .$$

Assim,

$$\omega_3 = n \log \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right) . \quad (\text{A1.1})$$

Podemos, ainda, escrever

$$\tilde{\sigma}^2 = \frac{\sigma^2}{n} \sum_{l=1}^n \frac{(y_l - x_l^\top \beta^{(0)})^2}{\sigma^2} \stackrel{\text{d}}{=} \frac{\sigma^2}{n} \chi_n^2 \quad (\text{A1.2})$$

e

$$\hat{\sigma}^2 = \frac{\sigma^2}{n} \sum_{l=1}^n \frac{(y_l - x_l^\top \hat{\beta})^2}{\sigma^2} \stackrel{\text{d}}{=} \frac{\sigma^2}{n} \chi_{n-p}^2 , \quad (\text{A1.3})$$

onde $\stackrel{\text{d}}{=}$ indica que as variáveis aleatórias à esquerda e à direita das igualdades têm a mesma distribuição. Substituindo (A1.2) e (A1.3) em (A1.1) vem

$$\omega_3 \stackrel{\text{d}}{=} n \log \left(\frac{\chi_n^2}{\chi_{n-p}^2} \right) ,$$

resultando,

$$\begin{aligned} E(\omega_3) &= n \{ E(\log \chi_n^2) - E(\log \chi_{n-p}^2) \} \\ &= n \left\{ \psi \left(\frac{n}{2} \right) - \psi \left(\frac{n-p}{2} \right) \right\} , \end{aligned}$$

uma vez que, até ordem n^{-1} , $E(\log \chi_n^2) = \psi(n/2) + \log 2$, onde $\psi(x) = d \log \Gamma(x)/dx$ é a função digama. Usando a expansão da função digama (Abramowitz e Stegun, 1970; fórmula 6.3.18),

$$\psi(x) = \log x - \frac{1}{2x} - \frac{1}{12x^2} + O(x^{-4})$$

vem,

$$\psi\left(\frac{n}{2}\right) = \log\left(\frac{n}{2}\right) - \frac{1}{n} - \frac{1}{3n^2} + O(n^{-4})$$

e

$$\psi\left(\frac{n-p}{2}\right) = \log\left(\frac{n-p}{2}\right) - \frac{1}{n-p} - \frac{1}{3(n-p)^2} + O(n^{-4}).$$

Assim, até ordem n^{-2} ,

$$\begin{aligned} \psi\left(\frac{n}{2}\right) - \psi\left(\frac{n-p}{2}\right) &= \log \frac{n}{n-p} + \left(\frac{1}{n-p} - \frac{1}{n}\right) + \frac{1}{3} \left(\frac{1}{(n-p)^2} - \frac{1}{n^2}\right) \\ &= -\log\left(1 - \frac{p}{n}\right) + \frac{1}{n} \left(\frac{1}{1 - \frac{p}{n}} - 1\right) + \frac{1}{3n^2} \left(\frac{1}{\left(1 - \frac{p}{n}\right)^2} - 1\right). \end{aligned}$$

Mas,

$$\log(1+x) = x - \frac{x^2}{2} + O(x^{-3}), \quad \frac{1}{\left(1 - \frac{p}{n}\right)} = 1 + \frac{p}{n} + O(n^{-2})$$

e

$$\frac{1}{\left(1 - \frac{p}{n}\right)^2} = 1 + \frac{2p}{n} + O(n^{-2}).$$

Logo, até ordem n^{-1} , temos

$$\begin{aligned} E(\omega_3) &= -n \left(-\frac{p}{n} - \frac{p^2}{2n^2} \right) + \frac{p}{n} = p + \frac{p^2}{2n} + \frac{p}{n} \\ &= p + \frac{p(p+2)}{2n}, \end{aligned} \tag{A1.4}$$

que coincide com a fórmula (2.33) do Capítulo 2, obtida usando a expansão de Lawley. Para o caso $p = 1$, ou seja, β unidimensional e $X = 1_n$, temos o teste da hipótese H_1 , apresentada no item (i) da Seção 2.4.1, que por (A1.4), resulta

$$E(\omega_1) = 1 + \frac{3}{2n} + O(n^{-2}),$$

coincidindo com a expressão encontrada na Seção 2.4.1, usando a expansão de Lawley.

APÊNDICE 2

Neste apêndice apresentamos, com mais detalhes, o cálculo dos somatórios envolvidos na expressão de ε_{p+q} em (2.4), resultando na fórmula matricial de ε_{p+q} em (2.9), para os modelos normais lineares. Para tanto, encontramos, primeiramente, cumulantes de derivadas da função de log-verossimilhança que envolvem apenas componentes do vetor de parâmetros β , apenas componentes de γ e cumulantes mistos, envolvendo componentes de β e γ . Em seguida, mostramos a construção das parcelas $\Sigma'_{\beta,\gamma}\ell_{rstu}$ e $\Sigma'_{\beta,\gamma}\ell_{rstuvw}$ em notação matricial, que compõem a fórmula matricial final de ε_{p+q} dada por (2.9).

A2.1 OBTENÇÃO DE ALGUNS CUMULANTES DE DERIVADAS DO LOGARITMO DA FUNÇÃO DE VEROSSIMILHANÇA EM MODELOS NORMAIS LINEARES COM ESTRUTURA DE COVARIÂNCIA NÃO-ESCALAR

A função densidade de probabilidade para o modelo normal linear descrito na Seção 2.1 pode ser escrita por

$$\pi(y; \beta, \gamma) = \frac{1}{(2\pi)^{n/2} |V^{-1}|^{1/2}} \exp \left\{ -\frac{1}{2} (y - X\beta)^T V (y - X\beta) \right\} .$$

Então, o logaritmo da função de verossimilhança total para o vetor de parâmetros $\theta =$

$(\beta^\top, \gamma^\top)^\top$, dado o vetor de observações $y = (y_1, \dots, y_n)^\top$, tem a forma

$$\ell(\theta) = \ell(\beta, \gamma) = -\frac{n}{2} \log(2n) + \frac{1}{2} \log |V| - \frac{1}{2} (y - X\beta)^\top V (y - X\beta) . \quad (\text{A2.1})$$

Por simples diferenciação de $\ell(\theta)$ acima, em relação aos componentes de β , chega-se a

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \beta_r} &= x_r^\top V (y - X\beta) , \\ \frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \beta_s} &= -x_r^\top V x_s , \\ \text{e} \quad \frac{\partial^k \ell(\theta)}{\partial \beta_{i_1} \dots \partial \beta_{i_k}} &= 0 , \text{ para } k \geq 3, \end{aligned}$$

onde $\beta_{i_1}, \dots, \beta_{i_k} \in \{\beta_1, \dots, \beta_p\}$.

Como no Capítulo 2, convencionamos que índices com letras minúsculas referem-se a componentes do vetor de parâmetros β e índices com letras maiúsculas as componentes do vetor de parâmetros γ .

Tomando o valor esperado das derivadas acima, com respeito ao vetor aleatório Y , obtém-se os seguintes cumulantes relativos aos componentes de β ,

$$\begin{aligned} \kappa_{r,s} &= -\kappa_{rs} = x_r^\top V x_s \\ \kappa_{i_1, \dots, i_k} &= 0 , \text{ para qualquer } k \geq 3, \end{aligned}$$

com $i_1, \dots, i_k \in \{\beta_1, \dots, \beta_k\}$.

As derivadas do logaritmo da função de verossimilhança em relação aos componentes do vetor γ são obtidas de maneira análoga. Com a finalidade de simplificar a apresentação final de suas expressões, definimos $\bar{V}_R = V^{-1}V_R$, $\bar{V}^R = V^R V = -\bar{V}_R$, $\bar{V}_{RS} = V^{-1}V_{RS}$, $\bar{V}^{RS} = V^{RS} V$, $\bar{V}_{RST} = V^{-1}V_{RST}$, $\bar{V}^{RST} = V^{RST} V$, etc., $m_R = \text{tr}(\bar{V}_R)$, $m^R = \text{tr}(\bar{V}^R) = -m_R$, $m_{RS} = \text{tr}(\bar{V}_{RS})$, $m^{RS} = \text{tr}(\bar{V}^{RS})$, $m_{R,S} = \text{tr}(\bar{V}_R \bar{V}_S)$, $m^{R,S} = \text{tr}(\bar{V}^R \bar{V}^S)$, $m_{RS,T} = \text{tr}(\bar{V}_{RS} \bar{V}_T)$, $m^{RS,T} = \text{tr}(\bar{V}^{RS} \bar{V}^T)$, $m_{R,S,T} = \text{tr}(\bar{V}_R \bar{V}_S \bar{V}_T)$, $m^{R,S,T} = \text{tr}(\bar{V}^R \bar{V}^S \bar{V}^T)$, $m_{R,S,T,U} = \text{tr}(\bar{V}_R \bar{V}_S \bar{V}_T \bar{V}_U)$, $m^{R,S,T,U} = \text{tr}(\bar{V}^R \bar{V}^S \bar{V}^T \bar{V}^U)$, e assim por diante, onde $V_R = \partial V / \partial \gamma_R$, $V_{RS} = \partial^2 V / \partial \gamma_R \partial \gamma_S$ e $V_{RST} = \partial^3 V / \partial \gamma_R \partial \gamma_S \partial \gamma_T$ e V^R , V^{RS} e V^{RST} são as três últimas derivadas tomadas da matriz V^{-1} .

Os $\bar{V}$'s definidos acima satisfazem algumas relações, tais como

$$(a) \quad \bar{V}_{RS} = \bar{V}^R \bar{V}^S - \bar{V}^{RS} + \bar{V}^S \bar{V}^R,$$

$$(b) \quad \bar{V}^{RS} = \bar{V}_R \bar{V}_S - \bar{V}_{RS} + \bar{V}_S \bar{V}_R, \text{ etc.}$$

Para provar (a), escrevemos $\bar{V}_{RS} = V^{-1}V_{RS}$, sendo $V_{RS} = \partial V_R / \partial \gamma_S$. Mas da relação (i) apresentada na Seção 2.3.1 segue que $V_R = -VV^R V$, para $R \in \{\gamma_1, \dots, \gamma_q\}$. Então, $V_{RS} = \partial(-VV^R V) / \partial \gamma_S = -V_S V^R V - VV^{RS} V - VV^R V_S$. Usando, novamente a relação (i) da Seção 2.3.1 para V_S e substituindo na última relação, temos $V_{RS} = VV^S VV^R V - VV^{RS} V + VV^R VV^S V = V\bar{V}^S \bar{V}^R - V\bar{V}^{RS} + V\bar{V}^R \bar{V}^S$, que multiplicada por V^{-1} produz o resultado desejado. A demonstração de (b) é obtida de forma análoga.

Das relações existentes entre os V 's, tais como (a) e (b) acima, seguem algumas relações, que, em determinadas situações especiais, podem facilitar seus cálculos, como, por exemplo,

$$(c) \quad m_{RS} = 2m^{R,S} - m^{RS},$$

$$(d) \quad m^{RS} = 2m_{R,S} - m_{RS},$$

$$(e) \quad m_{RS,T} = m^{RS,T} - m^{R,S,T} - m^{S,R,T},$$

$$(f) \quad m_{R,S,T} = -m^{R,S,T} \text{ e assim por diante.}$$

A verificação das relações (c) – (f) é bastante simples. Basta usar a definição dos m 's, as relações (a) e (b) entre os V 's, a relação (i) da Seção 2.3.1 e propriedades do operador traço.

As expressões para as derivadas do logaritmo da função de verossimilhança em relação aos componentes de γ são apresentadas a seguir, usando as definições acima, onde omitimos as passagens intermediárias, uma vez que algumas destas derivadas envolvem um trabalho algébrico bastante tedioso. Temos,

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \gamma_R} &= \frac{1}{2} \left\{ m_R - \frac{1}{2} (y - X\beta)^\top V_R (y - X\beta) \right\}, \\ \frac{\partial^2 \ell(\theta)}{\partial \gamma_R \partial \gamma_S} &= -\frac{1}{2} \left\{ m_{R,S} - m_{RS} + (y - X\beta)^\top V_{RS} (y - X\beta) \right\}, \end{aligned}$$

$$\frac{\partial^3 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \gamma_T} = \frac{1}{2} \left\{ m_{RST} - m_{RS,T} - m_{RT,S} - m_{ST,R} + m_{R,S,T} \right. \\ \left. + m_{R,T,S} - (y - X\beta)^\top V_{RST} (y - X\beta) \right\}$$

e

$$\frac{\partial^4 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \gamma_T \partial \gamma_U} = \frac{1}{2} \left\{ m_{RSTU} - m_{RST,U} - m_{RSU,T} - m_{RTU,S} - m_{STU,R} \right. \\ - m_{RS,TU} - m_{RT,SU} - m_{RU,ST} + m_{RS,U,T} + m_{RS,T,U} \\ + m_{RT,U,S} + m_{RT,S,U} + m_{ST,R,U} + m_{ST,U,R} + m_{TU,S,R} \\ + m_{TU,R,S} + m_{SU,R,T} + m_{SU,T,R} + m_{RU,T,S} + m_{RU,S,T} \\ - m_{S,R,T,U} - m_{S,R,U,T} - m_{R,T,S,U} - m_{R,U,S,T} - m_{R,S,U,T} \\ \left. - m_{R,S,T,U} + (y - X\beta)^\top V_{RSTU} (y - X\beta) \right\} .$$

Os respectivos cumulantes são obtidos tomando-se o valor esperado das expressões acima e utilizando a relação (iii) apresentada na Seção 2.3.1. Assim, facilmente, encontramos

$$\kappa_{RS} = -\frac{1}{2} m_{R,S} , \\ \kappa_{RST} = -\frac{1}{2} (m_{RS,T} + m_{RT,S} + m_{ST,R} - m_{R,S,T} - m_{R,T,S})$$

e

$$\kappa_{RSTU} = -\frac{1}{2} \left(m_{RST,U} + m_{RSU,T} + m_{RTU,S} + m_{STU,R} + m_{RS,TU} + m_{RT,SU} + m_{RU,ST} \right. \\ - m_{RS,T,U} - m_{RS,U,T} - m_{RT,S,U} - m_{RT,U,S} - m_{ST,R,U} - m_{ST,U,R} - m_{SU,R,T} \\ - m_{SU,T,R} - m_{RU,T,S} - m_{RU,S,T} - m_{TU,S,R} - m_{TU,R,S} + m_{R,S,T,U} + m_{R,S,U,T} \\ \left. + m_{R,U,S,T} + m_{R,U,T,S} + m_{R,T,S,U} + m_{R,T,U,S} \right) .$$

Para obtermos cumulantes mistos, basta derivarmos $\ell(\theta)$, em (A2.1), com respeito ao(s) componente(s) de β e ao(s) componente(s) de γ e calcularmos o valor esperado destas derivadas. Assim, temos

$$\frac{\partial^2 \ell}{\partial \beta_r \partial \gamma_S} = x_r^\top V_S (y - X\beta) ,$$

$$\begin{aligned}
\frac{\partial^3 \ell}{\partial \beta_r \partial \beta_s \partial \gamma_T} &= -x_r^\top V_T x_s, \\
\frac{\partial^3 \ell}{\partial \beta_r \partial \gamma_s \partial \gamma_T} &= x_r^\top V_{ST} (y - X\beta), \\
\frac{\partial^4 \ell}{\partial \beta_r \partial \beta_s \partial \beta_t \partial \gamma_U} &= 0, \\
\frac{\partial^4 \ell}{\partial \beta_r \partial \beta_s \partial \gamma_T \partial \gamma_U} &= -x_r^\top V_{TU} x_s, \\
\frac{\partial^4 \ell}{\partial \beta_r \partial \gamma_s \partial \gamma_T \partial \gamma_U} &= x_r^\top V_{STU} (y - X\beta).
\end{aligned}$$

Tomando-se o valor esperado das derivadas acima, obtemos os seguintes cumulantes:

$$\begin{aligned}
\kappa_{rS} &= 0 = \kappa_{Rs} = \kappa_{rST} = \kappa_{RSt} = \kappa_{rstR} = \kappa_{rSTU}, \\
\kappa_{rST} &= -x_r^\top V_T x_s \\
\text{e} \quad \kappa_{rSTU} &= -x_r^\top V_{TU} x_s.
\end{aligned}$$

As derivadas de cumulantes são obtidas derivando-se qualquer um dos cumulantes acima, em relação a componentes de β ou a componentes de γ ou em relação a ambos. Assim, não é difícil mostrarmos que as seguintes derivadas de cumulantes são nulas:

$\kappa_{rs}^{(tu)}, \kappa_{rs}^{(tU)}, \kappa_{rs}^{(tu)}, \kappa_{rs}^{(tU)}, \kappa_{rs}^{(TU)}, \kappa_{rs}^{(tU)}, \kappa_{rs}^{(u)}, \kappa_{rs}^{(U)}, \kappa_{rsT}^{(u)}, \kappa_{rsT}^{(u)}, \kappa_{rsT}^{(U)}, \kappa_{rsT}^{(u)}$

1 2 5 3

4 6 9 15

8 10 12 13 14 16 20 17 18 22 19

Para os outros casos, obtemos

$$\begin{aligned}
4 \quad \kappa_{rs}^{(T)} &= -x_r^\top V_T x_s, \\
6 \quad \kappa_{RS}^{(T)} &= -\frac{1}{2} (m_{ST,R} + m_{RT,S} - m_{R,S,T} - m_{R,T,S}), \\
9 \quad \kappa_{rs}^{(TU)} &= -x_r^\top V_{TU} x_s, \\
15 \quad \kappa_{RS}^{(TU)} &= -\frac{1}{2} (m_{STU,R} + m_{RTU,S} + m_{RU,ST} + m_{RT,SU} - m_{ST,R,U} - m_{ST,U,R} - m_{RT,U,S} \\
&\quad - m_{RT,S,U} - m_{SU,T,R} - m_{SU,R,T} - m_{RU,S,T} - m_{RU,T,S} - m_{TU,R,S} - m_{TU,S,R} \\
&\quad + m_{R,S,T,U} + m_{R,S,U,T} + m_{R,U,S,T} + m_{R,U,T,S} + m_{R,T,S,U} + m_{R,T,U,S})
\end{aligned}$$

e

$$23 \quad \kappa_{RST}^{(U)} = -\frac{1}{2} (m_{RSU,T} + m_{RTU,S} + m_{STU,R} + m_{RS,TU} + m_{RT,SU} + m_{RU,ST} - m_{RS,T,U}$$

$$\begin{aligned}
& -m_{RS,U,T} - m_{RT,S,U} - m_{RT,U,S} - m_{ST,R,U} - m_{ST,U,R} - m_{SU,T,R} - m_{SU,R,T} \\
& -m_{TU,R,S} - m_{TU,S,R} - m_{RU,S,T} - m_{RU,T,S} + m_{R,S,T,U} + m_{R,S,U,T} + m_{R,T,S,U} \\
& + m_{R,T,U,S} + m_{R,U,S,T} + m_{R,U,T,S} \Big) .
\end{aligned}$$

Do apresentado acima temos que a matriz de informação de Fisher para β é escrita por $K_{\beta,\beta} = X^T V X$, cujo (r, s) -ésimo elemento é $\kappa_{r,s} = -\kappa_{rs} = x_r^T V x_s$ e a matriz de informação de Fisher para γ é $K_{\gamma,\gamma} = \{\kappa_{T,U}\}$, onde $\kappa_{T,U} = -\kappa_{TU} = \frac{1}{2}m_{T,U}$ é o (T, U) -ésimo elemento típico.

A2.2 OBTENÇÃO DE ε_{p+q}

Para obter a fórmula matricial de ε_{p+q} , seguindo a expansão de Lawley (1956), necessitamos avaliar os seguintes somatórios:

$$\sum_{\beta,\gamma}' \ell_{rstu} = \sum_{\beta,\gamma}' \kappa^{rs} \kappa^{tu} \left(\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} - \kappa_{rt}^{(su)} \right)$$

e

$$\begin{aligned}
\sum_{\beta,\gamma}' \ell_{rstuvw} &= \sum_{\beta,\gamma}' \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \kappa_{rtv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)} \right) \right. \\
&\quad \left. + \kappa_{rtu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right\} ,
\end{aligned}$$

onde cada um dos índices r, s, t, u, v e w representa um dos componentes do vetor de parâmetros β ou do vetor de parâmetros γ e $\sum_{\beta,\gamma}'$ denota o somatório sobre todos os componentes dos vetores β e γ .

Como visto na seção anterior, muitos cumulantes e derivadas de cumulantes se anulam, devido à ortogonalidade global entre os vetores β e γ . Assim, para o cálculo de $\sum_{\beta,\gamma}' \ell_{rstu}$ apenas quatro termos necessitam ser avaliados, a saber:

1. $\sum_{\beta}' \kappa^{rs} \kappa^{tu} \left(\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right),$
2. $\sum_{\beta,\gamma}' \kappa^{rs} \kappa^{TU} \left(\frac{1}{4} \kappa_{rsTU} - \kappa_{rsT}^{(U)} + \kappa_{rT}^{(sU)} \right),$

3. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} \left(\frac{1}{4} \kappa_{RS tu} - \kappa_{RS t}^{(u)} + \kappa_{R t}^{(Su)} \right),$
4. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{TU} \left(\frac{1}{4} \kappa_{RSTU} - \kappa_{RST}^{(U)} + \kappa_{RT}^{(SU)} \right),$

onde Σ'_{β} e Σ'_{γ} representam, respectivamente, os somatórios somente sobre os componentes do vetor β e somente sobre os componentes do vetor γ . A obtenção dos quatro termos acima é conseguida substituindo-se os cumulantes encontrados na seção anterior nestas expressões. O primeiro termo é nulo, já que as parcelas dentro do parênteses são todas nulas. No segundo termo a terceira parcela dentro dos parênteses é nula, e por substituição das parcelas restantes não-nulas, segue para o termo 2

$$\begin{aligned} & \frac{3}{4} \sum'_{\beta, \gamma} \kappa^{rs} \kappa^{TU} x_r^{\top} V_{TU} x_s = \frac{3}{4} \sum'_{T, U} \kappa^{T, U} \sum'_{r, s} \kappa^{r, s} x_r^{\top} V_{TU} x_s \\ &= \frac{3}{4} \sum'_{T, U} \kappa^{T, U} \sum_{r, s} \kappa^{r, s} G_{rs}^{(TU)} = \frac{3}{4} \sum'_{T, U} \kappa^{T, U} \text{tr}(K_{\beta, \beta}^{-1} G_{\gamma}^{(TU)}) , \end{aligned}$$

onde $G_{rs}^{(TU)}$ é o (r, s) -ésimo elemento da matriz de dimensão $p \times p$ $G_{\gamma}^{(TU)} = X^{\top} V_{TU} X$. Sejam ρ_T um q -vetor coluna com 1 no componente T e zero nos demais e $\tau_{1, \gamma, \beta}^{(T)}$ um q -vetor coluna, cujo elemento típico U é $\text{tr}(K_{\beta, \beta}^{-1} G_{\gamma}^{(TU)})$. Então, podemos escrever 2 em notação matricial, na forma

$$\frac{3}{4} \sum_T \rho_T^{\top} K_{\gamma, \gamma}^{-1} \tau_{1, \gamma, \beta}^{(T)} ,$$

onde T varre o conjunto $\{\gamma_1, \dots, \gamma_q\}$.

No terceiro termo apenas a primeira parcela dentro dos parênteses não é identicamente nula, resultando, para este termo

$$-\frac{1}{4} \sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} x_t^{\top} V_{RS} x_u .$$

Este resultado é idêntico àquele obtido para o termo 2, a menos da constante multiplicando o somatório, pela troca simultânea de (R, S) por (T, U) e (t, u) por (r, s) .

O quarto termo é obtido substituindo-se as parcelas dentro dos parênteses pelas determinadas na Seção A2.1. Uma considerável redução ocorre observando-se que vários m 's dentro do somatório são iguais quando: (i) permutamos os índices R com S e T com U e (ii) quando

trocamos simultaneamente os índices (R, S) pelos índices (T, U) . Depois de um longo trabalho algébrico resulta que o termo 4 é dado por

$$\begin{aligned} & \frac{1}{8} \sum_{\gamma} ' \kappa^{RS} \kappa^{TU} (2m_{RT,SU} - m_{RS,TU} + 4m_{RS,T,U} - 4m_{R,S,T,U} - 2m_{R,T,S,U}) \\ &= \frac{1}{8} \sum_{R,S} ' \kappa^{R,S} \sum_{T,U} ' \kappa^{T,U} d_{TU}^{(RS)} = \frac{1}{8} \sum_{R,S} ' \kappa^{R,S} \text{tr}(K_{\gamma,\gamma}^{-1} D_{\gamma}^{(RS)}) , \end{aligned}$$

onde $d_{TU}^{(RS)} = 2m_{RT,SU} - m_{RS,TU} + 4m_{RS,T,U} - 4m_{R,S,T,U} - 2m_{R,T,S,U}$ é o (T, U) -ésimo elemento da matriz quadrada de ordem q $D_{\gamma}^{(RS)}$, para $R, S \in \{\gamma_1, \dots, \gamma_q\}$. Sejam $\tau_{2\gamma}^{(R)}$ q vetores coluna de dimensão q , cujo S -ésimo elemento típico é $\text{tr}(K_{\gamma,\gamma}^{-1} D_{\gamma}^{(RS)})$ e ρ_R o q -vetor coluna com 1 no componente R e zero nos demais, para $R \in \{\gamma_1, \dots, \gamma_q\}$. Então, o termo 4 pode ser escrito, em notação matricial na forma

$$\frac{1}{8} \sum_R ' \rho_R^{\top} K_{\gamma,\gamma}^{-1} \tau_{2\gamma}^{(R)} .$$

Assim, temos,

$$\sum_{\beta,\gamma} ' \ell_{rstu} = \frac{1}{2} \sum_R ' \rho_R^{\top} K_{\gamma,\gamma}^{-1} \tau_{1\gamma,\beta}^{(R)} + \frac{1}{8} \sum_R ' \rho_R^{\top} K_{\gamma,\gamma}^{-1} \tau_{2\gamma}^{(R)} .$$

Definimos, ainda, as matrizes quadradas de ordem q , $M_{\gamma,\beta}$ e N_{γ} , formadas, respectivamente, pelos q -vetores coluna $\tau_{1\gamma,\beta}^{(R)}$ e $\tau_{2\gamma}^{(R)}$, para $R \in \{\gamma_1, \dots, \gamma_q\}$, ou seja,

$$M_{\gamma,\beta} = (\tau_{1\gamma,\beta}^{(\gamma_1)}, \dots, \tau_{1\gamma,\beta}^{(\gamma_q)}) \quad \text{e} \quad N_{\gamma} = (\tau_{2\gamma}^{(\gamma_1)}, \dots, \tau_{2\gamma}^{(\gamma_q)}) .$$

Também, podemos escrever para a inversa da informação relativa a γ ,

$$K_{\gamma,\gamma}^{-1} = (\kappa_1^{\gamma,\gamma}, \dots, \kappa_q^{\gamma,\gamma}) ,$$

com $\kappa_R^{\gamma,\gamma}$ denotando o vetor coluna $(\kappa^{\gamma_1,\gamma_R}, \dots, \kappa^{\gamma_q,\gamma_R})$ para $R = 1, \dots, q$. Sejam

$$m_{\gamma,\beta} = \text{vec}(M_{\gamma,\beta}) , \quad n_{\gamma} = \text{vec}(N_{\gamma}) \quad \text{e} \quad \kappa^{\gamma} = \text{vec}(K_{\gamma,\gamma}^{-1}) .$$

Finalmente, em notação matricial, temos

$$\sum_{\beta,\gamma} ' \ell_{rstu} = \frac{1}{2} \kappa^{\gamma\top} m_{\gamma,\beta} + \frac{1}{8} \kappa^{\gamma\top} n_{\gamma} . \quad (\text{A2.2})$$

Para o cálculo de $\sum_{\beta,\gamma} ' \ell_{rstuvw}$ devemos desenvolver apenas os oito termos abaixo, designados por a, b, c, d, e, f, g e h, devido à ortogonalidade entre os vetores de parâmetros β e γ :

- a. $\sum'_{\beta} \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \kappa_{rtv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right\},$
- b. $\sum'_{\beta, \gamma} \kappa^{rs} \kappa^{tu} \kappa^{VW} \left\{ \kappa_{rtV} \left(\frac{\kappa_{suW}}{6} - \kappa_{sW}^{(u)} \right) + \kappa_{rtu} \left(\frac{\kappa_{sVW}}{4} - \kappa_{sW}^{(V)} \right) + \kappa_{rt}^{(V)} \kappa_{sW}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sW}^{(V)} \right\},$
- c. $\sum'_{\beta, \gamma} \kappa^{rs} \kappa^{TU} \kappa^{vw} \left\{ \kappa_{rTv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(U)} \right) + \kappa_{rTu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{rT}^{(v)} \kappa_{sw}^{(U)} + \kappa_{rT}^{(U)} \kappa_{sw}^{(v)} \right\},$
- d. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} \kappa^{vw} \left\{ \kappa_{Rtv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)} \right) + \kappa_{Rtu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{Rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{Rt}^{(u)} \kappa_{sw}^{(v)} \right\},$
- e. $\sum'_{\beta, \gamma} \kappa^{rs} \kappa^{TU} \kappa^{VW} \left\{ \kappa_{rTV} \left(\frac{\kappa_{suW}}{6} - \kappa_{sW}^{(U)} \right) + \kappa_{rTu} \left(\frac{\kappa_{sVW}}{4} - \kappa_{sW}^{(V)} \right) + \kappa_{rT}^{(V)} \kappa_{sW}^{(U)} + \kappa_{rT}^{(U)} \kappa_{sW}^{(V)} \right\},$
- f. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} \kappa^{VW} \left\{ \kappa_{RtV} \left(\frac{\kappa_{suW}}{6} - \kappa_{sW}^{(u)} \right) + \kappa_{Rtu} \left(\frac{\kappa_{sVW}}{4} - \kappa_{sW}^{(V)} \right) + \kappa_{Rt}^{(V)} \kappa_{sW}^{(u)} + \kappa_{Rt}^{(u)} \kappa_{sW}^{(V)} \right\},$
- g. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{TU} \kappa^{vw} \left\{ \kappa_{RTv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(U)} \right) + \kappa_{RTu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{RT}^{(v)} \kappa_{sw}^{(U)} + \kappa_{RT}^{(U)} \kappa_{sw}^{(v)} \right\},$
- h. $\sum'_{\beta, \gamma} \kappa^{RS} \kappa^{TU} \kappa^{VW} \left\{ \kappa_{RTV} \left(\frac{\kappa_{suW}}{6} - \kappa_{sW}^{(U)} \right) + \kappa_{RTu} \left(\frac{\kappa_{sVW}}{4} - \kappa_{sW}^{(V)} \right) + \kappa_{RT}^{(V)} \kappa_{sW}^{(U)} + \kappa_{RT}^{(U)} \kappa_{sW}^{(V)} \right\}.$

Dos oito termos acima, facilmente verificamos que os termos a e e se anulam e que o termo h envolve apenas os componentes do vetor de parâmetros γ , cujo desenvolvimento deixamos por último. O desenvolvimentos matriciais dos termos b, c, d, f e g são semelhantes e assim, apresentamos, com detalhes, apenas a derivação matricial do termo b.

Como muitos dos cumulantes e suas derivadas são nulos, resulta para o termo b,

$$\begin{aligned} & - \sum'_{\beta, \gamma} \kappa^{r,s} \kappa^{t,u} \kappa^{V,W} \kappa_{rtV} \frac{\kappa_{suW}}{6} = - \frac{1}{6} \sum'_{\beta, \gamma} \kappa^{r,s} \kappa^{t,u} \kappa^{V,W} (x_r^\top V_V x_t x_s^\top V_W x_u) \\ & = - \frac{1}{6} \sum'_{r,s,t,u,V,W} \kappa^{r,s} \kappa^{t,u} \kappa^{V,W} G_{rt}^{(V)} G_{su}^{(W)}, \end{aligned}$$

onde $G_{rt}^{(V)}$ é o (r, t) -ésimo elemento da matriz de dimensão $p \times p$ $G_\gamma^{(V)} = X^\top V_V X$, para todo $V \in \{\gamma_1, \dots, \gamma_q\}$. O último resultado pode, ainda, ser escrito por

$$\begin{aligned} & - \frac{1}{6} \sum'_{r,s,V,W} \kappa^{V,W} \left(\sum_t' \kappa^{t,u} G_{rt}^{(V)} \right) \left(\sum_s' \kappa^{r,s} G_{su}^{(W)} \right) \\ & = - \frac{1}{6} \sum'_{r,u,V,W} \kappa^{V,W} \left(\rho_r^\top G_\gamma^{(V)} K_{\beta,\beta}^{-1} \rho_u \right) \left(\rho_r^\top K_{\beta,\beta}^{-1} G_\gamma^{(W)} \rho_u \right) \\ & = - \frac{1}{6} \sum'_{r,u,V,W} \kappa^{V,W} \rho_r^\top G_\gamma^{(V)} K_{\beta,\beta}^{-1} \rho_u \rho_u^\top G_\gamma^{(W)} K_{\beta,\beta}^{-1} \rho_r \\ & = - \frac{1}{6} \sum'_{r,V,W} \kappa^{V,W} \rho_r^\top G_\gamma^{(V)} K_{\beta,\beta}^{-1} \left(\sum_u' \rho_u \rho_u^\top \right) G_\gamma^{(W)} K_{\beta,\beta}^{-1} \rho_r. \end{aligned}$$

Como $\sum'_u \rho_u \rho_u^\top = I_q$, onde ρ_u é o q -vetor coluna com 1 no componente u e zero nos demais e I_q é a matriz identidade de ordem q , o termo **b** reduz-se a

$$\begin{aligned} & -\frac{1}{6} \sum'_{V,W} \kappa^{V,W} \sum_r \rho_r^\top G_\gamma^{(V)} K_{\beta,\beta}^{-1} G_\gamma^{(W)} K_{\beta,\beta}^{-1} \rho_r \\ &= -\frac{1}{6} \sum'_{V,W} \kappa^{V,W} \text{tr}(G_\gamma^{(V)} K_{\beta,\beta}^{-1} G_\gamma^{(W)} K_{\beta,\beta}^{-1}) = -\frac{1}{6} \sum'_{V,W} \kappa^{V,W} p_{VW} \\ &= -\frac{1}{6} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}), \end{aligned}$$

onde $p_{VW} = \text{tr}(G_\gamma^{(V)} K_{\beta,\beta}^{-1} G_\gamma^{(W)} K_{\beta,\beta}^{-1})$ é o (V, W) -ésimo elemento da matriz quadrada de ordem q $P_{\gamma,\beta}$.

Substituindo os cumulantes e derivadas de cumulantes envolvidos nos termos **c**, **d**, **f** e **g** e procedendo de maneira análoga ao desenvolvimento do termo **b**, obtemos para estes termos

$$\begin{aligned} \mathbf{c} &= \frac{5}{6} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}), \\ \mathbf{d} &= -\frac{1}{6} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) - \frac{1}{4} \tau_{1,\beta}^\top K_{\gamma,\gamma}^{-1} \tau_{1,\beta}, \\ \mathbf{f} &= \frac{1}{8} \tau_{1,\beta}^\top K_{\gamma,\gamma}^{-1} \tau_{3,\gamma}, \\ \mathbf{g} &= \frac{1}{8} \tau_{1,\beta}^\top K_{\gamma,\gamma}^{-1} \tau_{4,\gamma}, \end{aligned}$$

sendo $\tau_{1,\beta}$ o vetor coluna $q \times 1$ com elemento típico R dado por $\text{tr}(K_{\beta,\beta}^{-1} G_\gamma^{(R)})$ e $G_\gamma^{(R)} = X^\top V_R V$ de dimensão $p \times p$, $\tau_{3,\gamma}$ o vetor coluna $q \times 1$ com elemento típico R $\text{tr}(K_{\gamma,\gamma}^{-1} E_\gamma^{(R)})$ sendo $E_\gamma^{(R)}$ uma matriz de dimensão $q \times q$ cujo (I, J) -ésimo elemento é $e_{IJ}^{(R)} = 2m_{RI,J} + 3m_{R,IJ} - 6m_{R,I,J}$ e $\tau_{4,\gamma}$ o vetor coluna $q \times 1$ com elemento típico R igual a $\text{tr}(K_{\beta,\beta}^{-1} F_\gamma^{(R)})$ com $F_\gamma^{(R)}$ uma matriz $q \times q$, cujo (I, J) -ésimo elemento típico é $f_{IJ}^{(R)} = -2m_{RI,J} - m_{R,IJ} + 2m_{R,I,J}$.

O desenvolvimento matricial do oitavo termo de $\sum'_{\beta,\gamma} \ell_{rstuvw}$ é obtido substituindo-se, primeiramente, os cumulantes de terceira ordem e derivadas dos cumulantes de segunda ordem envolvidos, os quais são encontrados na Seção A2.1. Após simplificações obtemos para o termo **h**

$$\begin{aligned} & \sum_{R,S,T,U,V,W} \kappa^{R,S} \kappa^{T,U} \kappa^{V,W} \left\{ \frac{1}{8} m_{RT,V} m_{SU,W} - \frac{1}{16} m_{TU,R} (m_{VW,S} - m_{S,V,W}) \right. \\ & \left. - \frac{1}{12} m_{R,T,V} (m_{S,U,W} + m_{S,W,U}) - \frac{1}{4} m_{R,T,U} m_{S,V,W} \right\}. \end{aligned}$$

As quatro parcelas dentro do último somatório podem ser escritas, em notação matricial, por

$$\begin{aligned} \frac{1}{8} \sum_{R,S,T,U,V,W} ' \kappa^{R,S} \kappa^{T,U} \kappa^{V,W} m_{RT,V} m_{SU,W} &= \frac{1}{8} \sum_{V,W} ' \kappa^{V,W} \sum_{S,T} ' \left(\sum_R ' \kappa^{S,R} m_{V,R} \right) \\ \left(\sum_U ' \kappa^{T,U} m_{W,US} \right) &= \frac{1}{8} \sum_{V,W} ' \kappa^{V,W} \text{tr}(K_{\gamma,\gamma}^{-1} T^{(V)} K_{\gamma,\gamma}^{-1} T^{(W)}) = \frac{1}{8} \sum_{V,W} ' \kappa^{V,W} t_{VW} \\ &= \frac{1}{8} \text{tr}(K_{\gamma,\gamma}^{-1} T_\gamma) , \end{aligned}$$

$$\begin{aligned} & -\frac{1}{16} \sum_{R,S,T,U,V,W} ' \kappa^{R,S} \kappa^{T,U} \kappa^{V,W} m_{TU,R} (m_{VW,S} - 4m_{V,W,S}) \\ &= -\frac{1}{16} \sum_{R,S,T,U} ' \kappa^{R,S} \kappa^{T,U} m_{TU,R} \sum_{V,W} ' \kappa^{V,W} (m_{VW,S} - 4m_{V,W,S}) \\ &= -\frac{1}{16} \sum_{R,S,T,U} ' \kappa^{R,S} \kappa^{T,U} m_{TU,R} \text{tr}[K_{\gamma,\gamma}^{-1} (T^{(S)} - 4T_1^{(S)})] \\ &= -\frac{1}{16} \sum_{R,S} ' \kappa^{R,S} \text{tr}[K_{\gamma,\gamma}^{-1} (T^{(S)} - 4T_1^{(S)})] \text{tr}(K_{\gamma,\gamma}^{-1} T^{(R)}) = -\frac{1}{16} \sum_{R,S} ' \kappa^{R,S} u_{R,S} \\ &= -\frac{1}{16} \text{tr}(K_{\gamma,\gamma}^{-1} U_\gamma) , \end{aligned}$$

$$\begin{aligned} & -\frac{1}{12} \sum_{R,S,T,U,V,W} ' \kappa^{R,S} \kappa^{T,U} \kappa^{V,W} m_{R,T,V} (m_{S,U,W} + m_{S,W,U}) \\ &= -\frac{1}{12} \left\{ \sum_{V,W} ' \kappa^{V,W} \sum_{R,U} ' \left(\sum_T ' \kappa^{U,T} m_{R,T,V} \right) \left(\sum_S ' \kappa^{R,S} m_{S,U,W} \right) \right. \\ & \quad \left. + \sum_{V,W} ' \kappa^{V,W} \sum_{R,U} ' \left(\sum_T ' \kappa^{U,T} m_{R,T,V} \right) \left(\sum_S ' \kappa^{R,S} m_{S,W,U} \right) \right\} \\ &= -\frac{1}{12} \left\{ \sum_{V,W} ' \kappa^{V,W} \text{tr} \left[K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)} \right] + \sum_{V,W} ' \kappa^{W,V} \text{tr} \left[K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)\top} \right] \right\} \\ &= -\frac{1}{12} \left\{ \sum_{V,W} ' \kappa^{V,W} q_{VW} + \sum_{V,W} ' \kappa^{W,V} r_{W,V} \right\} = -\frac{1}{12} \left\{ \text{tr}(K_{\gamma,\gamma}^{-1} Q_\gamma) + \text{tr}(K_{\gamma,\gamma}^{-1} R_\gamma) \right\} \\ &= -\frac{1}{12} \text{tr} \left[K_{\gamma,\gamma}^{-1} (Q_\gamma + R_\gamma) \right] \end{aligned}$$

e

$$-\frac{1}{4} \sum_{R,S,T,U,V,W} ' \kappa^{R,S} \kappa^{T,U} \kappa^{V,W} m_{R,T,U} m_{S,V,W}$$

$$\begin{aligned}
&= -\frac{1}{4} \sum'_{R,S} \kappa^{R,S} \left(\sum'_{T,U} \kappa^{T,U} m_{R,T,U} \right) \left(\sum'_{V,W} \kappa^{V,W} m_{S,V,W} \right) \\
&= -\frac{1}{4} \sum'_{R,S} \kappa^{R,S} \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(R)}) \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(S)}) = -\frac{1}{4} \sum'_{R,S} \kappa^{R,S} s_{RS} \\
&= -\frac{1}{4} \text{tr}(K_{\gamma,\gamma}^{-1} S_\gamma) ,
\end{aligned}$$

onde $T_\gamma = \{t_{VW}\}$ é uma matriz $q \times q$ cujo (V, W) -ésimo elemento é $t_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T^{(V)} K_{\gamma,\gamma}^{-1} T^{(W)})$, sendo $T^{(V)}$ uma matriz $q \times q$, cujo (I, J) -ésimo elemento é $t_{IJ}^{(V)} = m_{V,IJ}$, $U_\gamma = \{u_{VW}\}$ é uma matriz $q \times q$ cujo (V, W) -ésimo elemento é $u_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T^{(V)}) \text{tr}[K_{\gamma,\gamma}^{-1} (T^{(W)} - 4T_1^{(W)})]$, sendo $T^{(V)}$ como acima e $T_1^{(W)}$ uma matriz $q \times q$, cujo (I, J) -ésimo elemento é $t_{IJ}^{(W)} = m_{W,IJ}$. Além disso, $Q_\gamma = \{q_{VW}\}$ é uma matriz $q \times q$, cujo (V, W) -ésimo elemento é $q_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)})$, $R_\gamma = \{r_{VW}\}$ é uma matriz $q \times q$, cujo elemento típico (V, W) é $r_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)\top} K_{\gamma,\gamma}^{-1} T_1^{(W)\top})$ e, $S_\gamma = \{s_{VW}\}$ é uma matriz $q \times q$, cujo elemento típico (V, W) é $s_{VW} = \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(V)}) \text{tr}(K_{\gamma,\gamma}^{-1} T_1^{(W)})$. Assim, em notação matricial, resulta para o termo h

$$\frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right] .$$

Logo, agrupando os oito termos desenvolvidos acima, segue, em notação matricial, que

$$\begin{aligned}
\sum'_{\beta,\gamma} \ell_{rstuvw} &= \frac{1}{2} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta} - \frac{1}{8} \tau_{1\gamma,\beta}^\top K_{\gamma,\gamma}^{-1} (2\tau_{1\gamma,\beta} - \tau_{3\gamma} - \tau_{4\gamma})) \\
&\quad + \frac{1}{8} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right] . \quad (\text{A2.3})
\end{aligned}$$

Calculando a diferença entre (A2.2) e (A2.3), encontramos

$$\begin{aligned}
\varepsilon_{p+q} &= \frac{1}{8} \kappa^{\gamma^\top} n_\gamma - \frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right] \\
&\quad + \frac{1}{2} \kappa^{\gamma^\top} m_{\gamma,\beta} - \frac{1}{2} \text{tr}(K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) + \frac{1}{8} \tau_{1\gamma,\beta}^\top K_{\gamma,\gamma}^{-1} (2\tau_{1\gamma,\beta} - \tau_{3\gamma} - \tau_{4\gamma}) .
\end{aligned}$$

Denotando $A_\beta = \sum_\beta '(\ell_{rstu} - \ell_{rstuvw})$, $B_\gamma = \sum_\gamma '(\ell_{rstu} - \ell_{rstuvw})$ e $C_{\beta,\gamma} = \sum_{\beta,\gamma} (\ell_{rstu} - \ell_{rstuvw})$, podemos decompor ε_{p+q} da seguinte forma:

$$\varepsilon_{p+q} = A_\beta + B_\gamma + C_{\beta,\gamma} ,$$

sendo

$$A_\beta = 0 ,$$

$$B_\gamma = \frac{1}{8} \kappa^{\gamma^\top} n_\gamma - \frac{1}{48} \text{tr} \left[K_{\gamma,\gamma}^{-1} (6T_\gamma - 3U_\gamma - 4(Q_\gamma + R_\gamma) - 12S_\gamma) \right]$$

e

$$C_{\beta,\gamma} = \frac{1}{2} \kappa^{\gamma^\top} m_{\gamma,\beta} - \frac{1}{2} \text{tr} (K_{\gamma,\gamma}^{-1} P_{\gamma,\beta}) + \frac{1}{8} \tau_{1,\gamma,\beta} K_{\gamma,\gamma}^{-1} (2\tau_{1,\gamma,\beta} - \tau_{3_\gamma} - \tau_{4_\gamma}) .$$

APÊNDICE 3

Nosso objetivo neste apêndice é apresentar o cálculo das derivadas da função de log-verossimilhança, os respectivos cumulantes e derivadas destes para o modelo normal heterocedástico estudado no Capítulo 6. Obtemos em seguida, as parcelas $\Sigma' \ell_{rstu}$ e $\Sigma' \ell_{rstuvw}$, utilizando álgebra matricial. A diferença entre estas duas somas fornece a quantidade ε_{p+q} , exibida em (6.11) – (6.13).

Considere o modelo normal heterocedástico descrito por (6.1). O logaritmo da função de verossimilhança associada a este modelo tem a forma

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{l=1}^n \log H_l - \frac{1}{2} \sum_{l=1}^n \frac{v_l}{H_l} .$$

Apresentamos, a seguir, as derivadas da função acima, os cumulantes em relação a estas derivadas e derivadas destes. Temos,

$$\frac{\partial \ell(\theta)}{\partial \beta_r} = \sum_{l=1}^n \frac{x_{lr}(y_l - x_l^\top \beta)}{H_l} = \sum_{l=1}^n \sigma_l^{-2} (y_l - x_l^\top \beta) x_{lr} ,$$

$$\frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \beta_s} = - \sum_{l=1}^n \frac{x_{lr} x_{ls}}{H_l} = - \sum_{l=1}^n \sigma_l^{-2} x_{lr} x_{ls} ,$$

$$\frac{\partial^{(i)} \ell(\theta)}{\partial \beta_{r_1} \cdots \partial \beta_{r_i}} = 0, \text{ para } i = 3, 4, \dots \text{ e } r_1, \dots, r_i \in \{1, \dots, p\} ,$$

$$\frac{\partial^3 \ell(\theta)}{\partial \beta_r \partial \beta_s \partial \gamma_T} = \sum_{l=1}^n \frac{H_l^{(1)}}{H_l^2} x_{lr} x_{ls} w_{lT} = \sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lr} x_{ls} w_{lT} ,$$

$$\begin{aligned}\frac{\partial^4 \ell(\theta)}{\partial \beta_r \partial \beta_s \partial \gamma_T \partial \gamma_U} &= \sum_{l=1}^n \left(\frac{H_l^{(2)}}{H_l^2} - \frac{2(H_l^{(1)})^2}{H_l^3} \right) x_{lr} x_{ls} w_{lT} w_{lU} \\ &= \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lr} x_{ls} w_{lT} w_{lU} ,\end{aligned}$$

onde $P(n \times n) = \text{diag}\{p_{ii}\} = \Lambda^{(1)} \Lambda^{-1}$ e $Q(n \times n) = \text{diag}\{q_{jj}\} = \Lambda^{(2)} \Lambda^{-1}$, com $\Lambda = \text{diag}\{H_1, \dots, H_n\}$.

Outras derivadas seguem abaixo.

$$\frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \gamma_T} = - \sum_{l=1}^n \sigma_l^{-2} p_{ll} (y_l - x_l^\top \beta) x_{lr} w_{lT} ,$$

$$\frac{\partial \ell(\theta)}{\partial \gamma_R} = -\frac{1}{2} \left[\sum_{l=1}^n \left(\frac{H_l^{(1)}}{H_l} - \frac{H_l^{(1)}}{H_l^2} v_l \right) w_{lR} \right] = -\frac{1}{2} \sum_{l=1}^n (1 - \sigma_l^{-2} v_l) p_{ll} w_{lR} ,$$

$$\begin{aligned}\frac{\partial^2 \ell(\theta)}{\partial \gamma_R \partial \gamma_S} &= -\frac{1}{2} \left\{ \sum_{l=1}^n \left[\frac{H_l^{(2)}}{H_l} - \frac{(H_l^{(1)})^2}{H_l^2} - \left(\frac{H_l^{(2)}}{H_l^2} - \frac{2(H_l^{(1)})^2}{H_l^3} \right) v_l \right] w_{lR} w_{lS} \right\} \\ &= -\frac{1}{2} \sum_{l=1}^n [q_{ll} - p_{ll}^2 - (q_{ll} - 2p_{ll}^2) \sigma_l^{-2} v_l] w_{lR} w_{lS} ,\end{aligned}$$

$$\frac{\partial^3 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \beta_t} = - \sum_{l=1}^n (q_{ll} - 2p_{ll}^2) \sigma_l^{-2} (y_l - x_l^\top \beta) x_{lt} w_{lR} w_{lS} ,$$

$$\frac{\partial^4 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \beta_t \partial \beta_u} = \sum_{l=1}^n (q_{ll} - 2p_{ll}^2) \sigma_l^{-2} x_{lu} x_{lt} w_{lR} w_{lS} ,$$

$$\begin{aligned}\frac{\partial^3 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \gamma_T} &= -\frac{1}{2} \sum_{l=1}^n \left[\left(\frac{H_l^{(3)}}{H_l} - \frac{H_l^{(2)} H_l^{(1)}}{H_l^2} - \frac{2H_l^{(1)} H_l^{(2)}}{H_l^2} + \frac{2(H_l^{(1)})^3}{H_l^3} \right) \right. \\ &\quad \left. - \left(\frac{H_l^{(3)}}{H_l^2} + \frac{2H_l^{(2)} H_l^{(1)}}{H_l^3} - \frac{4H_l^{(1)} H_l^{(2)}}{H_l^3} + \frac{6(H_l^{(1)})^3}{H_l^4} \right) v_l \right] w_{lR} w_{lS} w_{lT} \\ &= -\frac{1}{2} \sum_{l=1}^n \left[\left(\frac{H_l^{(3)}}{H_l} - 3p_{ll} q_{ll} + 2p_{ll}^3 \right) - \left(\frac{H_l^{(3)}}{H_l} - 6p_{ll} q_{ll} + 6p_{ll}^3 \right) \sigma_l^{-2} v_l \right] \\ &\quad w_{lR} w_{lS} w_{lT} ,\end{aligned}$$

$$\begin{aligned}\frac{\partial^4 \ell(\theta)}{\partial \gamma_R \partial \gamma_S \partial \gamma_T \partial \gamma_U} &= -\frac{1}{2} \sum_{l=1}^n \left[\left(\frac{H_l^{(4)}}{H_l} + \frac{12(H_l^{(1)})^2 H_l^{(2)}}{H_l^3} - \frac{3(H_l^{(2)})^2}{H_l^2} - \frac{4H_l^{(3)} H_l^{(1)}}{H_l^2} - \frac{6(H_l^{(1)})^4}{H_l^4} \right) \right. \\ &\quad \left. - \left(\frac{H_l^{(4)}}{H_l} - \frac{8H_l^{(3)} H_l^{(1)}}{H_l^2} - \frac{6(H_l^{(2)})^2}{H_l^2} + \frac{36(H_l^{(1)})^2 H_l^{(2)}}{H_l^3} - \frac{24(H_l^{(1)})^4}{H_l^4} \right) \sigma_l^{-2} v_l \right] \\ &\quad w_{lR} w_{lS} w_{lT} w_{lU} ,\end{aligned}$$

$$\begin{aligned}
& \omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU} , \\
& = -\frac{1}{2} \sum_{l=1}^n \left[\left(\frac{H_l^{(4)}}{H_l} + 12p_{ll}^2 q_{ll} - 3q_{ll}^2 - 4r_{ll}p_{ll} - 6p_{ll}^4 \right) - \left(\frac{H_l^{(4)}}{H_l} + 36p_{ll}^2 q_{ll} \right. \right. \\
& \quad \left. \left. - 6q_{ll}^2 - 8r_{ll}p_{ll} - 24p_{ll}^4 \right) \sigma_l^{-2} v_l \right] \omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU} ,
\end{aligned}$$

onde $R(n \times n) = \Lambda^{(3)}\Lambda^{-1}$.

Tomando esperanças em relação a y das derivadas acima, e observando que $E(v_l) = E(y_l - x_l^\top \beta)^2 = H_l = \sigma_l^2$, obtemos os seguintes cumulantes:

$$\begin{aligned}
\kappa_{rs} &= -\sum_{l=1}^n \sigma_l^{-2} x_{lr} x_{ls}, \\
\kappa_{rS} &= 0 = \kappa_{rst} = \kappa_{rstu} = \kappa_{rST} = \kappa_{rstU}, \\
\kappa_{rsT} &= \sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lr} x_{ls} \omega_{lT}, \\
\kappa_{rsTU} &= \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lr} x_{ls} \omega_{lT} \omega_{lU}, \\
\kappa_{RS} &= -\frac{1}{2} \sum_{l=1}^n p_{ll}^2 \omega_{lR} \omega_{lS}, \\
\kappa_{RST} &= \frac{1}{2} \sum_{l=1}^n (4p_{ll}^3 - 3p_{ll} q_{ll}) \omega_{lR} \omega_{lS} \omega_{lT}
\end{aligned}$$

e

$$\kappa_{RSTU} = -\frac{1}{2} \sum_{l=1}^n (-24p_{ll}^2 q_{ll} + 3q_{ll}^2 + 4r_{ll}p_{ll} + 18p_{ll}^4) \omega_{lR} \omega_{lS} \omega_{lT} \omega_{lU}.$$

Derivando os cumulantes acima, em relação a componentes do vetor β e do vetor γ , resulta, após simplificações algébricas,

$$\begin{aligned}
\kappa_{rs}^{(t)} &= 0 = \kappa_{rst}^{(u)} = \kappa_{rs}^{(tu)} = \kappa_{rs}^{(tU)} = \kappa_{rS}^{(t)} = \kappa_{rS}^{(U)} = \kappa_{rS}^{(tU)} = \kappa_{rs}^{(TU)} \\
&= \kappa_{rsT}^{(u)} = \kappa_{RS}^{(t)} = \kappa_{RS}^{(tu)} = \kappa_{RST}^{(u)} = \kappa_{rST}^{(u)} = \kappa_{rST}^{(U)}, \\
\kappa_{rs}^{(T)} &= \sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lr} x_{ls} \omega_{lT} = \kappa_{rsT}, \\
\kappa_{rsT}^{(U)} &= \kappa_{rsTU} = \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lr} x_{ls} \omega_{lT} \omega_{lU}, \\
\kappa_{RS}^{(T)} &= \sum_{l=1}^n (p_{ll}^3 - p_{ll} q_{ll}) \omega_{lR} \omega_{lS} \omega_{lT},
\end{aligned}$$

$$\kappa_{RST}^{(U)} = -\frac{1}{2} \sum_{l=1}^n (3q_{ll}^2 + 3p_{ll}r_{ll} - 18p_{ll}^2q_{ll} + 12p_{ll}^4)\omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU}$$

e

$$\kappa_{RS}^{(TU)} = -\sum_{l=1}^n (q_{ll}^2 + p_{ll}r_{ll} - 5p_{ll}^2q_{ll} + 3p_{ll}^4)\omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU}.$$

Mostramos, a seguir, com detalhes a obtenção dos dois somatórios, $\sum'_{\beta,\gamma} \ell_{rstu}$ e $\sum'_{\beta,\gamma} \ell_{rstuvw}$, envolvidos no cálculo de ε_{p+q} .

Para o cálculo de $\sum'_{\beta,\gamma} \ell_{rstu}$, onde $\ell_{rstu} = \kappa^{rs}\kappa^{tu} \left(\frac{1}{4}\kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right)$, desprezando-se os cumulantes e derivadas de cumulantes que são nulos, devemos avaliar a seguinte expressão:

$$\begin{aligned} \sum'_{\beta,\gamma} \ell_{rstu} &= \sum'_{\beta,\gamma} \kappa^{rs}\kappa^{TU} \left(\frac{1}{4}\kappa_{rsTU} - \kappa_{rsT}^{(U)} \right) + \sum'_{\beta,\gamma} \kappa^{RS}\kappa^{tu} \frac{1}{4}\kappa_{RStu} \\ &\quad + \sum'_{\gamma} \kappa^{RS}\kappa^{TU} \left(\frac{1}{4}\kappa_{RSTU} - \kappa_{RST}^{(U)} + \kappa_{RT}^{(SU)} \right). \end{aligned}$$

Substituindo as quantidades na expressão acima, temos

$$\begin{aligned} \sum'_{\beta,\gamma} \ell_{rstu} &= \sum'_{\beta,\gamma} \kappa^{rs}\kappa^{TU} \left(\frac{1}{4} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lr} x_{ls} \omega_{lT}\omega_{lU} - \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lr} x_{ls} \omega_{lT}\omega_{lU} \right) \\ &\quad + \frac{1}{4} \sum'_{\beta,\gamma} \kappa^{RS}\kappa^{tu} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) x_{lt} x_{lu} \omega_{lR}\omega_{lS} - \frac{1}{8} \sum'_{\gamma} \kappa^{RS}\kappa^{TU} \left[\sum_{l=1}^n (-24p_{ll}^2 q_{ll} \right. \\ &\quad \left. + 3q_{ll}^2 + 4r_{ll}p_{ll} + 18p_{ll}^4) \omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU} + \frac{1}{2} \sum_{l=1}^n (3q_{ll}^2 + 3p_{ll}r_{ll} - 18p_{ll}^2q_{ll} \right. \\ &\quad \left. + 12p_{ll}^4) \omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU} - \sum_{l=1}^n (q_{ll}^2 + p_{ll}r_{ll} - 5p_{ll}^2q_{ll} + 3p_{ll}^4) \omega_{lR}\omega_{lS}\omega_{lT}\omega_{lU} \right]. \end{aligned}$$

Agrupando-se termos iguais e efetuando-se a soma primeiro sobre os parâmetros e depois sobre a amostra, obtemos

$$\begin{aligned} \sum'_{\beta,\gamma} \ell_{rstu} &= -\frac{3}{4} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) \left(\sum'_{r,s} \kappa^{rs} x_{lr} x_{ls} \right) \left(\sum'_{T,U} \kappa^{TU} \omega_{lT}\omega_{lU} \right) \\ &\quad + \frac{1}{8} \sum_{l=1}^n (q_{ll}^2 - 8p_{ll}^2q_{ll} + 6p_{ll}^4) \left(\sum'_{R,S} \kappa^{RS} \omega_{lR}\omega_{lS} \right) \left(\sum'_{T,U} \kappa^{TU} \omega_{lT}\omega_{lU} \right) \\ &\quad + \frac{1}{4} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) \left(\sum'_{t,u} \kappa^{tu} x_{lt} x_{lu} \right) \left(\sum'_{R,S} \kappa^{RS} \omega_{lR}\omega_{lS} \right) \\ &= -\frac{3}{4} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) \left[x_l^T (-\kappa_{\beta,\beta}^{-1}) x_l \right] \left[\omega_l^T (-\kappa_{\gamma,\gamma}^{-1}) \omega_l \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{8} \sum_{l=1}^n (q_{ll}^2 - 8p_{ll}^2 q_{ll} + 6p_{ll}^4) [\omega_l^\top (-\kappa_{\gamma,\gamma}^{-1}) \omega_l] [\omega_l^\top (-\kappa_{\gamma,\gamma}^{-1}) \omega_l] \\
& + \frac{1}{4} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) [x_l^\top (-\kappa_{\beta,\beta}^{-1}) x_l] [\omega_l^\top (-\kappa_{\gamma,\gamma}^{-1}) \omega_l].
\end{aligned}$$

Usando as definições das matrizes A , B , A_d e B_d , introduzidas no Capítulo 6, podemos escrever

$$\begin{aligned}
\sum' \ell_{rstu} &= -\frac{3}{2} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) b_{ll} a_{ll} + \frac{1}{2} \sum_{l=1}^n (q_{ll} - 8p_{ll}^2 q_{ll} + 6p_{ll}^4) a_{ll}^2 \\
&+ \frac{1}{2} \sum_{l=1}^n \sigma_l^{-2} (q_{ll} - 2p_{ll}^2) b_{ll} a_{ll},
\end{aligned}$$

resultando, em notação matricial,

$$\begin{aligned}
\sum' \ell_{rstu} &= \text{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q)] - \frac{1}{2} \text{tr}[A_d^2 (-Q^2 + 8P^2 Q - 6P^4)] \\
&= \text{tr}[A_d B_d \Lambda^{-1} (2P^2 - Q) - \frac{1}{2} A_d^2 (-Q^2 + 8P^2 Q - 6P^4)]. \quad (\text{A3.1})
\end{aligned}$$

Para o cálculo de $\sum'_{\beta,\gamma} \ell_{rstuvw}$ devemos avaliar, depois de eliminadas as parcelas nulas, a expressão,

$$\begin{aligned}
\sum'_{\beta,\gamma} \ell_{rstuvw} &= \sum'_{\beta,\gamma} \kappa^{rs} \kappa^{tu} \kappa^{VW} \frac{\kappa_{rtV} \kappa_{suW}}{6} + \sum'_{\beta,\gamma} \kappa^{rs} \kappa^{TU} \kappa^{vw} \kappa_{rTv} \left(\frac{\kappa_{sUw}}{6} - \kappa_{sw}^{(U)} \right) \\
&+ \sum'_{\beta,\gamma} \kappa^{RS} \kappa^{tu} \kappa^{vw} \left(\frac{\kappa_{Rtv} \kappa_{Suw}}{6} + \frac{\kappa_{Rtu} \kappa_{Svw}}{4} \right) \\
&+ \sum'_{\beta,\gamma} \kappa^{RS} \kappa^{tu} \kappa^{VW} \kappa_{Rtu} \left(\frac{\kappa_{SVW}}{4} - \kappa_{SW}^{(V)} \right) + \sum'_{\beta,\gamma} \kappa^{RS} \kappa^{TU} \kappa^{vw} \frac{\kappa_{RTU} \kappa_{Svw}}{4} \\
&+ \sum'_{\gamma} \kappa^{RS} \kappa^{TU} \kappa^{VW} \left\{ \kappa_{RTV} \left(\frac{\kappa_{SUW}}{6} - \kappa_{SW}^{(U)} \right) + \kappa_{RTU} \left(\frac{\kappa_{SVW}}{4} - \kappa_{SW}^{(V)} \right) \right. \\
&\left. + \kappa_{RT}^{(V)} \kappa_{SW}^{(U)} + \kappa_{RT}^{(U)} \kappa_{SW}^{(V)} \right\}.
\end{aligned}$$

Substituindo cumulantes e derivadas de cumulantes na última expressão, vem

$$\begin{aligned}
\sum' \ell_{rstuvw} &= \\
&\sum'_{\beta,\gamma} \kappa^{rs} \kappa^{tu} \kappa^{VW} \frac{1}{6} \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lr} x_{lt} \omega_{lV} \right) \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{ls} x_{lu} \omega_{lW} \right) \\
&- \frac{5}{6} \sum'_{\beta,\gamma} \kappa^{rs} \kappa^{TU} \kappa^{vw} \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lr} x_{lv} \omega_{lT} \right) \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{ls} x_{lw} \omega_{lU} \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} \kappa^{vw} \left[\frac{1}{6} \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lt} x_{lv} \omega_{lR} \right) \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lu} x_{lw} \omega_{lS} \right) \right. \\
& + \frac{1}{4} \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lt} x_{lu} \omega_{lR} \right) \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lv} x_{lw} \omega_{lS} \right) \Big] \\
& + \sum'_{\beta, \gamma} \kappa^{RS} \kappa^{tu} \kappa^{VW} \frac{1}{8} \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lt} x_{lu} \omega_{lR} \right) \left(\sum_{l=1}^n (5p_{ll} q_{ll} - 4p_{ll}^3) \omega_{lS} \omega_{lV} \omega_{lW} \right) \\
& + \frac{1}{8} \sum'_{\beta, \gamma} \kappa^{RS} \kappa^{TU} \kappa^{vw} \left(\sum_{l=1}^n (4p_{ll}^3 - 3p_{ll} q_{ll}) \omega_{lR} \omega_{lT} \omega_{lU} \right) \left(\sum_{l=1}^n \sigma_l^{-2} p_{ll} x_{lv} x_{lw} \omega_{lS} \right) \\
& + \sum'_{\gamma} \kappa^{RS} \kappa^{TU} \kappa^{VW} \left[-\frac{1}{24} \left(\sum_{l=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) \omega_{lR} \omega_{lT} \omega_{lV} \right) \left(\sum_{l=1}^n (9p_{ll} q_{ll} - 8p_{ll}^3) \omega_{lS} \omega_{lU} \omega_{lW} \right) \right. \\
& - \frac{1}{16} \left(\sum_{l=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) \omega_{lR} \omega_{lT} \omega_{lU} \right) \left(\sum_{l=1}^n (5p_{ll} q_{ll} - 4p_{ll}^3) \omega_{lS} \omega_{lV} \omega_{lW} \right) \\
& + \left(\sum_{l=1}^n p_{ll} (q_{ll} - p_{ll}^2) \omega_{lR} \omega_{lT} \omega_{lV} \right) \left(\sum_{l=1}^n p_{ll} (q_{ll} - p_{ll}^2) \omega_{lS} \omega_{lW} \omega_{lU} \right) \\
& \left. + \left(\sum_{l=1}^n p_{ll} (q_{ll} - p_{ll}^2) \omega_{lR} \omega_{lT} \omega_{lU} \right) \left(\sum_{l=1}^n p_{ll} (q_{ll} - p_{ll}^2) \omega_{lS} \omega_{lW} \omega_{lV} \right) \right] \\
= & \frac{1}{6} \sum_{r,s,t,u,v,w} ' \kappa^{rs} \kappa^{tu} \kappa^{VW} \left(\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} x_{lr} x_{ms} x_{lt} x_{mu} \omega_{lV} \omega_{lW} \right) \\
& - \frac{5}{6} \sum_{r,s,T,U,v,w} ' \kappa^{rs} \kappa^{TU} \kappa^{vw} \left(\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} x_{lr} x_{ms} \omega_{lT} \omega_{lU} x_{lv} x_{lw} \right) \\
& - \frac{1}{24} \sum_{R,S,T,U,V,W} ' \kappa^{RS} \kappa^{TU} \kappa^{VW} \left[\sum_{l=1}^n \sum_{m=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) (9p_{mm} q_{mm} - 8p_{mm}^3) \omega_{lR} \omega_{mS} \omega_{lT} \omega_{mU} \omega_{lV} \omega_{mW} \right] \\
& - \frac{1}{16} \sum_{R,S,T,U,V,W} ' \kappa^{RS} \kappa^{TU} \kappa^{VW} \left[\sum_{l=1}^n \sum_{m=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) (5p_{mm} q_{mm} - 4p_{mm}^3) \omega_{lR} \omega_{mS} \omega_{lT} \omega_{lU} \omega_{mV} \omega_{mW} \right] \\
& + \sum_{R,S,T,U,V,W} ' \kappa^{RS} \kappa^{TU} \kappa^{VW} \left[\sum_{l=1}^n \sum_{m=1}^n p_{ll} (q_{ll} - p_{ll}^2) p_{mm} (q_{mm} - p_{mm}^2) \omega_{lR} \omega_{mS} \omega_{lT} \omega_{mU} \omega_{lV} \omega_{mW} \right] \\
& + \sum_{R,S,T,U,V,W} ' \kappa^{RS} \kappa^{TU} \kappa^{VW} \left[\sum_{l=1}^n \sum_{l=1}^m p_{ll} (q_{ll} - p_{ll}^2) p_{mm} (q_{mm} - p_{mm}^2) \omega_{lR} \omega_{mS} \omega_{lT} \omega_{lU} \omega_{mV} \omega_{mW} \right] \\
& - \frac{1}{8} \sum_{R,S,T,U,v,w} ' \kappa^{RS} \kappa^{TU} \kappa^{vw} \left[\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} p_{ll} (3p_{mm} q_{mm} - 4p_{mm}^3) \omega_{mR} x_{lS} \omega_{mT} \omega_{mU} x_{lv} x_{lw} \right] \\
& + \frac{1}{8} \sum_{R,S,t,u,v,w} ' \kappa^{RS} \kappa^{tu} \kappa^{VW} \left[\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} p_{ll} (5p_{mm} q_{mm} - 4p_{mm}^3) \omega_{lR} \omega_{mS} x_{lt} x_{lu} \omega_{mV} \omega_{mW} \right] \\
& + \frac{1}{6} \sum_{R,S,t,u,v,w} ' \kappa^{RS} \kappa^{tu} \kappa^{vw} \left[\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \omega_{lR} \omega_{mS} x_{lt} x_{mu} x_{lv} x_{mw} \right]
\end{aligned}$$

$$+\frac{1}{4} \sum_{R,S,t,u,v,w} ' \kappa^{RS} \kappa^{tu} \kappa^{vw} \left[\sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \omega_{lR} \omega_{mS} x_{lt} x_{lu} x_{mv} x_{mw} \right].$$

Podemos, ainda, trocar a ordem dos somatórios $\sum'$ e $\sum$, e rearranjar os termos, para obtermos

$$\begin{aligned} \sum_{\beta,\gamma} ' \ell_{rstuvw} = & \frac{1}{6} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \left(\sum_{r,s} ' \kappa^{rs} x_{lr} x_{ms} \right) \left(\sum_{t,u} ' \kappa^{tu} x_{lt} x_{mu} \right) \left(\sum_{V,W} ' \kappa^{VW} \omega_{lV} \omega_{lW} \right) \\ & - \frac{5}{6} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \left(\sum_{r,s} ' \kappa^{rs} x_{lr} x_{ms} \right) \left(\sum_{T,U} ' \kappa^{TU} \omega_{lT} \omega_{mU} \right) \left(\sum_{v,w} ' \kappa^{vw} x_{lv} x_{mw} \right) \\ & - \frac{1}{24} \sum_{l=1}^n \sum_{m=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) (9p_{mm} q_{mm} - 8p_{mm}^3) \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{mS} \right) \left(\sum_{T,U} ' \kappa^{TU} \omega_{lT} \omega_{mU} \right) \\ & \left(\sum_{V,W} ' \kappa^{VW} \omega_{lV} \omega_{lW} \right) - \frac{1}{16} \sum_{l=1}^n \sum_{m=1}^n (3p_{ll} q_{ll} - 4p_{ll}^3) (5p_{mm} q_{mm} - 4p_{mm}^3) \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{mS} \right) \\ & \left(\sum_{T,U} ' \kappa^{TU} \omega_{lT} \omega_{lU} \right) \left(\sum_{V,W} ' \kappa^{VW} \omega_{mV} \omega_{mW} \right) + \sum_{l=1}^n \sum_{m=1}^n p_{ll} (q_{ll} - p_{ll}^2) p_{mm} (q_{mm} - p_{mm}^2) \\ & \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{mS} \right) \left(\sum_{T,U} ' \kappa^{TU} \omega_{lT} \omega_{mU} \right) \left(\sum_{V,W} ' \kappa^{VW} \omega_{lV} \omega_{mW} \right) \\ & + \sum_{l=1}^n \sum_{m=1}^n p_{ll} (q_{ll} - p_{ll}^2) p_{mm} (q_{mm} - p_{mm}^2) \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{mS} \right) \left(\sum_{T,U} ' \kappa^{TU} \omega_{lT} \omega_{lU} \right) \\ & \left(\sum_{V,W} ' \kappa^{VW} \omega_{mV} \omega_{mW} \right) - \frac{1}{8} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} (3p_{mm} q_{mm} - 4p_{mm}^3) \left(\sum_{R,S} ' \kappa^{RS} \omega_{mR} \omega_{lS} \right) \\ & \left(\sum_{T,U} ' \kappa^{TU} \omega_{mT} \omega_{mU} \right) \left(\sum_{V,W} ' \kappa^{VW} x_{lV} x_{lW} \right) + \frac{1}{8} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} p_{ll} (5p_{mm} q_{mm} - 4p_{mm}^3) \\ & \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{lS} \right) \left(\sum_{t,u} ' \kappa^{tu} x_{lt} x_{lu} \right) \left(\sum_{V,W} ' \kappa^{VW} \omega_{mV} \omega_{mW} \right) \\ & + \frac{1}{6} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \left(\sum_{R,S} ' \kappa^{RS} \omega_{lR} \omega_{mS} \right) \left(\sum_{t,u} ' \kappa^{tu} x_{lt} x_{mu} \right) \left(\sum_{v,w} ' \kappa^{vw} x_{lv} x_{mw} \right) \\ & + \frac{1}{4} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} \left(\sum_{R,S} ' \kappa^{RS} x_{lR} x_{mS} \right) \left(\sum_{t,u} ' \kappa^{tu} x_{lt} x_{lu} \right) \left(\sum_{v,w} ' \kappa^{vw} x_{mv} x_{mw} \right). \end{aligned}$$

Efetando-se as operações na última expressão, e utilizando-se as definições das matrizes A , B ,

A_d e B_d introduzidas na Seção 6.2, chegamos a

$$\begin{aligned} \sum_{\beta, \gamma}' \ell_{rstuvw} = & \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} a_{lm} b_{lm}^2 - \frac{1}{3} \sum_{l=1}^n \sum_{m=1}^n (12 p_{ll}^3 p_{mm} q_{mm} - 3 p_{ll} q_{ll} p_{mm} q_{mm} - 8 p_{ll}^3 p_{mm}^3) a_{lm}^3 \\ & - \frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n (p_{ll} q_{ll} p_{mm} q_{mm} - 4 p_{ll} q_{ll} p_{mm}^3 + 4 p_{ll}^3 p_{mm} q_{mm}) q_{lm} a_{ll} a_{mm} \\ & - \frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} \sigma_m^{-2} p_{ll} p_{mm} a_{lm} b_{ll} b_{mm} - \sum_{l=1}^n \sum_{m=1}^n \sigma_l^{-2} p_{ll} p_{mm} a_{lm} a_{mm} b_{ll} . \end{aligned}$$

A observação cuidadosa da fórmula acima permite-nos expressá-la em notação matricial por

$$\begin{aligned} \sum_{\beta, \gamma}' \ell_{rstuvw} = & -1_n^\top P A_d A B_d \Lambda^{-1} P 1_n - \frac{1}{2} 1_n^\top P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1_n \\ & + 1_n^\top P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1_n - \frac{1}{2} 1_n^\top P Q A_d A A_d Q P 1_n \\ & - \frac{1}{3} 1_n^\top P (4 P^2 A^{(3)} (3Q - 2P^2) - 3Q A^{(2)} Q) P 1_n , \end{aligned} \quad (\text{A3.2})$$

onde, $B^{(2)} = B \otimes B$, $A^{(3)} = A^{(2)} \otimes A$, $P = \Lambda^{(1)} \Lambda^{-1}$, $Q = \Lambda^{(2)} \Lambda^{-1}$, 1_n n -vetor coluna de uns e $\otimes$ simboliza produto direto, ou seja, produto de elementos correspondentes.

Finalmente, o termo ε_{p+q} é dado pela diferença das expressões (A3.1) e (A3.2).

APÊNDICE 4

Neste apêndice apresentamos o cálculo dos cumulantes e suas derivadas utilizados na obtenção da correção de Bartlett para a estatística da razão de verossimilhança em testes de hipóteses sobre os parâmetros dos modelos biparamétricos pertencentes à família exponencial.

A4.1 CÁLCULO DE CUMULANTES DOS MODELOS BIPARAMÉTRICOS DA FAMÍLIA EXPONENCIAL

O logaritmo da função de verossimilhança total do vetor de parâmetros $(\theta, \phi)^T$ para o modelo biparamétrico da família exponencial em (5.1), dado o conjunto de observações $y_1, \dots, y_n$, pode ser escrita por

$$\begin{aligned}\ell = \ell(\theta, \phi) &= \sum t(y_l; \theta, \phi) \\ &= \alpha_1(\theta, \phi) \sum_{l=1}^n d_1(y_l) + \alpha_2(\theta, \phi) \sum_{l=1}^n d_2(y_l) - n\rho(\theta, \phi) + \sum_{l=1}^n v(y_l) .\end{aligned}\quad (\text{A4.1})$$

Por simplicidade, no que segue denotamos simplesmente por α_1 , α_2 e ρ as funções $\alpha_1(\theta, \phi)$, $\alpha_2(\theta, \phi)$ e $\rho(\theta, \phi)$, respectivamente. Adotamos a notação introduzida no Capítulo 5 para as derivadas destas funções, ou seja, $\alpha_1^{(i,j)} = \partial^{i+j} \alpha_1 / \partial \theta^i \partial \phi^j$, $i, j = 1, 2$ e 3 , com notação análoga para as derivadas de α_2 e ρ . As derivadas em relação a θ e a ϕ da função em (A4.1) são dadas, respectivamente, por

$$\frac{\partial \ell}{\partial \theta} = \alpha_1^{(1,0)} \sum_{l=1}^n d_1(y_l) + \alpha_2^{(1,0)} \sum_{l=1}^n d_2(y_l) - n\rho^{(1,0)}$$

e

$$\frac{\partial \ell}{\partial \phi} = \alpha_1^{(0,1)} \sum_{l=1}^n d_1(y_l) + \alpha_2^{(0,1)} \sum_{l=1}^n d_2(y_l) - n\rho^{(0,1)}$$

Das condições de regularidade do logaritmo da função de verossimilhança em (A4.1) valem as seguintes relações,

$$\begin{aligned} \alpha_1^{(1,0)} \beta_1 + \alpha_2^{(0,1)} \beta_2 &= \rho^{(1,0)} , \\ \alpha_1^{(0,1)} \beta_1 + \alpha_2^{(0,1)} \beta_2 &= \rho^{(0,1)} , \end{aligned} \quad (\text{A4.2})$$

onde $\beta_r = E\{d_r(y_l)\}$, $r = 1, 2$, $l = 1, \dots, n$. Derivando ℓ novamente, vem

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \theta^2} &= \alpha_1^{(2,0)} \sum_{l=1}^n d_1(y_l) + \alpha_2^{(2,0)} \sum_{l=1}^n d_2(y_l) - n\rho^{(2,0)} \\ \frac{\partial^2 \ell}{\partial \phi^2} &= \alpha_1^{(0,2)} \sum_{l=1}^n d_1(y_l) + \alpha_2^{(0,2)} \sum_{l=1}^n d_2(y_l) - n\rho^{(2,0)} . \end{aligned}$$

Tomando o valor esperado destas derivadas e utilizando as relações em (A4.2), resulta que a informação para θ e para ϕ são dadas, respectivamente, por

$$\kappa_{\theta\theta} = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} \quad (\text{A4.3})$$

e

$$\kappa_{\phi\phi} = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(0,1)} \quad (\text{A4.4})$$

e, assim, $\kappa^{\theta\theta} = \kappa_{\theta\theta}^{-1}$ e $\kappa^{\phi\phi} = \kappa_{\phi\phi}^{-1}$ são as respectivas inversas. Resolvendo o sistema formado pelas equações em (A4.2), para β_1 e β_2 , resulta

$$\beta_1 = \frac{\rho^{(0,1)} \alpha_2^{(1,0)} - \rho^{(1,0)} \alpha_2^{(0,1)}}{\alpha_1^{(0,1)} \alpha_2^{(1,0)} - \alpha_1^{(1,0)} \alpha_2^{(0,1)}} , \quad \beta_2 = \frac{\rho^{(0,1)} \alpha_1^{(1,0)} - \rho^{(1,0)} \alpha_1^{(0,1)}}{\alpha_2^{(0,1)} \alpha_1^{(1,0)} - \alpha_2^{(1,0)} \alpha_1^{(0,1)}} . \quad (\text{A4.5})$$

Para estes modelos é fácil mostrar, também, que

$$\kappa_{\theta\phi} = -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(0,1)} , \quad \kappa_{\phi\theta} = -n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,0)} . \quad (\text{A4.6})$$

Conforme desenvolvimento apresentado no Capítulo 5, que assume θ e ϕ globalmente ortogonais, ou seja, $\kappa_{\theta\phi} = \kappa_{\phi\theta} = 0$, a estatística da razão de verossimilhança, para o teste da hipótese nula $H : \theta = \theta^{(0)}$ contra $A : \theta \neq \theta^{(0)}$, tem valor esperado, até ordem n^{-1} , expresso por

$$E(\omega) = 1 + A_\theta + C_{\theta,\phi} , \quad (\text{A4.7})$$

sendo que $A_\theta = \ell_{\theta\theta\theta\theta} - \ell_{\theta\theta\theta\theta\theta\theta}$ e $C_{\theta\phi} = \sum'_{\theta,\phi} (\ell_{rstu} - \ell_{rstuvw})$, com

$$\ell_{rstu} = \kappa^{rs} \kappa^{tu} \left[\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rs}^{(tu)} \right]$$

e

$$\ell_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left[\kappa_{rtv} \left(\frac{1}{6} \kappa_{suw} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{1}{4} \kappa_{svw} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right] .$$

A correção de Bartlett para este teste é determinada, então, pela relação $c = E(\omega)$.

Apresentamos, a seguir, os cumulantes envolvidos no cálculo de A_θ , que envolvem somente o parâmetro θ .

$$\begin{aligned} \kappa_{\theta\theta} &= -n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,0)} , \\ \kappa_{\theta\theta}^{(\theta)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(2,0)} \beta_r^{(1,0)} + \alpha_r^{(1,0)} \beta_r^{(2,0)} \right) , \\ \kappa_{\theta\theta}^{(\theta\theta)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(3,0)} \beta_r^{(1,0)} + 2\alpha_r^{(2,0)} \beta_r^{(2,0)} + \alpha_r^{(1,0)} \beta_r^{(3,0)} \right) , \\ \kappa_{\theta\theta\theta} &= -n \sum_{r=1}^2 \left(2\alpha_r^{(2,0)} \beta_r^{(1,0)} + \alpha_r^{(1,0)} \beta_r^{(2,0)} \right) , \\ \kappa_{\theta\theta\theta}^{(\theta)} &= -n \sum_{r=1}^2 \left(2\alpha_r^{(3,0)} \beta_r^{(1,0)} + 3\alpha_r^{(2,0)} \beta_r^{(2,0)} + \alpha_r^{(1,0)} \beta_r^{(3,0)} \right) , \\ \kappa_{\theta\theta\theta\theta} &= -n \sum_{r=1}^2 \left(3\alpha_r^{(3,0)} \beta_r^{(1,0)} + 3\alpha_r^{(2,0)} \beta_r^{(2,0)} + \alpha_r^{(1,0)} \beta_r^{(3,0)} \right) . \end{aligned} \quad (A4.8)$$

Não é difícil verificar que valem as seguintes relações:

$$\kappa_{\theta\theta}^{(\theta)} = \kappa_{\theta\theta\theta} + n \sum_{r=1}^2 \alpha_r^{(2,0)} \beta_r^{(1,0)} , \quad \kappa_{\theta\theta}^{(\theta\theta)} = \kappa_{\theta\theta\theta\theta} + n \sum_{r=1}^2 \left(2\alpha_r^{(3,0)} \beta_r^{(1,0)} + \alpha_r^{(2,0)} \beta_r^{(2,0)} \right) . \quad (A4.9)$$

A substituição dos cumulantes acima nas fórmulas de $\ell_{\theta\theta\theta\theta}$ e $\ell_{\theta\theta\theta\theta\theta\theta}$ resulta na expressão de A_θ em (5.16).

A quantidade mista $C_{\theta,\phi}$ é obtida pela diferença das duas parcelas seguintes:

$$C_{1,\theta,\phi} = \sum_{\theta,\phi} ' \ell_{rstu} \quad \text{e} \quad C_{2,\theta,\phi} = \sum_{\theta,\phi} ' \ell_{rstuvw} .$$

Devido a ortogonalidade entre os parâmetros θ e ϕ , muitos dos cumulantes mistos necessários ao cálculo de $C_{1\theta,\phi}$ e $C_{2\theta,\phi}$ são nulos. Apresentamos, a seguir, os cumulantes não nulos envolvidos nestas expressões:

$$\begin{aligned}
 \kappa_{\theta\theta}^{(\phi)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(1,1)} \beta_r^{(1,0)} + \alpha_r^{(1,0)} \beta_r^{(1,1)} \right), & \kappa_{\theta\theta}^{(\theta)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(1,1)} \beta_r^{(0,1)} + \alpha_r^{(0,1)} \beta_r^{(1,1)} \right), \\
 \kappa_{\theta\theta\phi} &= \kappa_{\theta\phi\theta} = \kappa_{\phi\theta\theta} = -n \sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(1,0)}, & \kappa_{\theta\phi\phi} &= \kappa_{\phi\theta\phi} = \kappa_{\phi\phi\theta} = -n \sum_{r=1}^2 \alpha_r^{(1,1)} \beta_r^{(0,1)}, \\
 \kappa_{\theta\theta\phi}^{(\phi)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(1,2)} \beta_r^{(1,0)} + \alpha_r^{(1,1)} \beta_r^{(1,1)} \right), & \kappa_{\phi\phi\theta}^{(\theta)} &= -n \sum_{r=1}^2 \left(\alpha_r^{(2,1)} \beta_r^{(0,1)} + \alpha_r^{(1,1)} \beta_r^{(1,1)} \right), \\
 \kappa_{\theta\theta\phi\phi} &= \kappa_{\phi\phi\theta\theta} = -n \sum_{r=1}^2 \left(\alpha_r^{(1,2)} \beta_r^{(1,0)} + \alpha_r^{(2,1)} \beta_r^{(0,1)} + \alpha_r^{(1,1)} \beta_r^{(1,1)} \right). & &
 \end{aligned} \tag{A4.10}$$

Para estes cumulantes, facilmente provamos as seguintes relações:

$$\begin{aligned}
 \kappa_{\phi\phi}^{(\theta)} &= \kappa_{\phi\phi\theta} - n \sum_{r=1}^2 \alpha_r^{(0,1)} \beta_r^{(1,1)}, & \kappa_{\theta\theta}^{(\phi)} &= \kappa_{\theta\theta\phi} - n \sum_{r=1}^2 \alpha_r^{(1,0)} \beta_r^{(1,1)}, \\
 \kappa_{\theta\theta\phi}^{(\phi)} &= \kappa_{\theta\theta\phi\phi} + n \sum_{r=1}^2 \alpha_r^{(2,1)} \beta_r^{(0,1)}, & \kappa_{\phi\phi\theta}^{(\theta)} &= \kappa_{\phi\phi\theta\theta} + n \sum_{r=1}^2 \alpha_r^{(1,2)} \beta_r^{(1,0)}. & &
 \end{aligned} \tag{A4.11}$$

Substituindo os cumulantes (A4.10) juntamente com as relações (A4.11) em $C_{1\theta,\phi}$ e $C_{2\theta,\phi}$ resultam as expressões (5.17') e (5.17'').

Se, por outro lado, o interesse é o teste sobre o parâmetro ϕ , basta considerar $(\phi, \theta)^\top$ como o vetor de parâmetros desconhecidos e, assim, as derivadas envolvidas nas expressões (5.16), (5.17') e (5.17'') são tomadas em relação a este vetor ao invés do vetor $(\theta, \phi)^\top$, para produzir $c = 1 + A_\phi + C_{\theta,\phi}$.

A4.2 DERIVADAS ENVOLVIDAS NO CÁLCULO DA CORREÇÃO DE BARTLETT

Apresentamos, agora, as derivadas das funções α_1 , α_2 , ρ , β_1 e β_2 para os modelos biparamétricos da família exponencial, considerados na Seção 5.4. Dividimos a apresentação em duas partes: primeiro, mostramos as derivadas utilizadas para o teste do parâmetro θ , ou seja,

quando se considera $(\theta, \phi)^\top$ o vetor de parâmetros e, na segunda parte, aquelas usadas no teste do parâmetro ϕ , neste caso o vetor de parâmetros desconhecidos é $(\phi, \theta)^\top$.

A4.2.1 TESTE DAS HIPÓTESES $H : \theta = \theta^{(0)}$ VERSUS $A : \theta \neq \theta^{(0)}$

1. Modelo normal $N(\theta, \phi)$

$$\pi(y; \theta, \phi) = (2\pi\phi)^{-1/2} \exp \left\{ -\frac{1}{2} \frac{(y - \theta)^2}{\phi} \right\}, \quad \phi > 0, \quad \theta, y \in \mathbb{R}. \quad (\text{A4.12})$$

$$\begin{aligned} \alpha_1 &= \frac{\theta}{\phi}, \quad \alpha_2 = -\frac{1}{2\phi} \quad \text{e} \quad \rho = \frac{\theta^2}{2\phi} + \frac{1}{2} \log \phi, \\ \rho^{(1,0)} &= \frac{\theta}{\phi}, \quad \rho^{(0,1)} = \frac{1}{2\phi} - \frac{\theta^2}{2\phi^2}, \\ \alpha_1^{(1,0)} &= \frac{1}{\phi}, \quad \alpha_1^{(0,1)} = -\frac{\phi}{\phi^2}, \quad \alpha_2^{(0,1)} = \frac{1}{2\phi^2}. \end{aligned}$$

Então, por (5.21), temos

$$\beta_1 = \theta \quad \text{e} \quad \beta_2 = \theta^2 + \phi.$$

Temos, ainda,

$$\begin{aligned} \alpha_1^{(1,1)} &= -\frac{1}{\phi^2}, \quad \alpha_1^{(1,2)} = \frac{2}{\phi^3}, \quad \alpha_1^{(2,0)} = \alpha_1^{(2,1)} = \alpha_1^{(3,0)} = 0, \quad \alpha_1^{(0,2)} = \frac{2\theta}{\phi^3}, \\ \beta_1^{(1,0)} &= 1, \quad \beta_1^{(2,0)} = \beta_1^{(0,j)} = 0 = \beta_1^{(i,j)}, \quad i, j = 1, 2, \\ \alpha_2^{(0,2)} &= -\frac{1}{\phi^3}, \quad \beta_2^{(0,1)} = 1, \quad \beta_2^{(1,1)} = \beta_2^{(0,2)} = \beta_2^{(1,2)} = 0, \quad \beta_2^{(1,0)} = 2\theta, \quad \beta_2^{(2,0)} = 2. \end{aligned}$$

O fator de correção para o teste considerado pode ser obtido substituindo-se as derivadas nas expressões (5.23) e (5.23') - (5.23''). Estes cálculos fornecem $c = 1 + \frac{3}{2n}$.

2. Modelo normal inverso $NI(\theta, \phi)$

$$\pi(y; \theta, \phi) = \frac{\phi^{1/2}}{(2\pi y^3)^{1/2}} \exp \left\{ -\frac{\phi(y - \theta)^2}{2\theta^2 y} \right\}, \quad y, \theta, \phi > 0. \quad (\text{A4.13})$$

$$\begin{aligned} \alpha_1 &= -\frac{\phi}{2\theta^2}, \quad \alpha_2 = -\frac{\phi}{2}, \quad \rho = -\left(\frac{\phi}{\theta} + \frac{1}{2} \log \phi \right), \\ \rho^{(1,0)} &= \frac{\phi}{\theta^2}, \quad \rho^{(0,1)} = -\left(\frac{1}{\theta} + \frac{1}{2\phi} \right), \quad \alpha^{(1,0)} = \frac{\phi}{\theta^3}, \quad \alpha_1^{(0,1)} = -\frac{\theta}{\phi^2}, \quad \alpha_2^{(0,1)} = -\frac{1}{2}. \end{aligned}$$

De (5.21) vem

$$\beta_1 = \theta \quad \text{e} \quad \beta_2 = \frac{1}{\theta} + \frac{1}{\phi}.$$

As outras derivadas necessárias ao cálculo de (5.22) e (5.23) - (5.23'') são:

$$\begin{aligned} \alpha_1^{(1,1)} &= \frac{1}{\theta^3}, \quad \alpha_1^{(1,2)} = 0 = \alpha_1^{(0,2)} = \alpha_1^{(0,3)}, \\ \alpha_1^{(2,0)} &= -\frac{3\phi}{\theta^4}, \quad \alpha_1^{(2,1)} = -\frac{3}{\theta^4}, \quad \alpha_1^{(3,0)} = 12\frac{\phi}{\theta^5}, \\ \beta_1^{(1,0)} &= 1, \quad \beta_1^{(2,0)} = \beta_1^{(0,i)} = \beta_1^{(i,j)} = 0, \quad i, j = 1, 2, \\ \alpha_2^{(0,2)} &= 0, \\ \beta_2^{(0,1)} &= -\frac{1}{\phi^2}, \quad \beta_2^{(1,1)} = \beta_2^{(1,2)} = 0, \quad \beta_2^{(0,2)} = \frac{2}{\phi^3}. \end{aligned}$$

A substituição das derivadas acima nas fórmulas (5.21), (5.23) - (5.23'') fornece $c = 1 + \frac{3}{2n}$.

3. Modelo gama $G(\theta, \phi)$ (com variância θ^2/ϕ)

$$\pi(y; \theta, \phi) = \left(\frac{\phi}{\theta}\right)^\phi \frac{1}{\pi(\phi)} y^{\phi-1} \exp\left\{-y\frac{\phi}{\theta}\right\}, \quad y, \phi, \theta > 0. \quad (\text{A4.14})$$

$$\begin{aligned} \alpha_1 &= -\frac{\phi}{\theta}, \quad \alpha_2 = \phi, \quad \rho = \phi \log \theta - \phi \log \phi + \log \Gamma(\phi), \\ \rho^{(1,0)} &= \frac{\phi}{\theta}, \quad \rho^{(0,1)} = \log \theta - \log \phi - 1 + \psi(\phi), \\ \alpha^{(1,0)} &= \frac{\phi}{\theta^2}, \quad \alpha_1^{(0,1)} = -\frac{1}{\theta}, \quad \alpha_2^{(0,1)} = 1. \end{aligned}$$

Através de (5.21) temos,

$$\beta_1 = \theta \quad \text{e} \quad \beta_2 = \log \theta - \log \phi + \psi(\phi).$$

Para o cálculo de (5.22) e (5.23), temos, ainda,

$$\begin{aligned} \alpha_1^{(1,1)} &= \frac{1}{\theta^2}, \quad \alpha_1^{(1,2)} = 0, \quad \alpha_1^{(2,0)} = -\frac{2\phi}{\theta^3}, \quad \alpha_1^{(2,1)} = -\frac{2}{\theta^3}, \\ \alpha_1^{(3,0)} &= \frac{6\phi}{\theta^4}, \quad \alpha_1^{(0,2)} = 0 = \alpha_1^{(0,3)}, \\ \beta_1^{(1,0)} &= 1, \quad \beta_1^{(0,2)} = \beta_1^{(0,i)}, \quad \beta_1^{(i,j)} = 0, \quad i, j = 1, 2, \\ \alpha_2^{(0,2)} &= 1, \quad \beta_2^{(0,1)} = \psi'(\phi) - \frac{1}{\phi}, \quad \beta_2^{(1,1)} = 0, \quad \beta_2^{(0,2)} = \psi''(\phi) + \frac{1}{\phi^2}. \end{aligned}$$

Para obtermos a correção de Bartlett, neste problema, basta substituírmos as derivadas acima em (5.22), (5.23) - (5.23''), para produzir $c = 1 + \frac{1}{12n\phi} \left[2 - \frac{3(2\phi^2\psi'' + \phi\psi' + 1)}{(\phi\psi' - 1)^2} \right]$.

4. Modelo log-gama $\text{LG}(\theta, \phi)$

$$\pi(y; \theta, \phi) = \frac{\phi^\phi}{\Gamma(\phi)} \exp \{ \phi[y - \theta - \exp(y - \theta)] \} , \quad \phi > 0, y, \theta \in \mathbb{R} . \quad (\text{A4.15})$$

$$\begin{aligned} \alpha_1 &= -\phi e^{-\theta} , \quad \alpha_2 = \phi , \quad \rho = \theta\phi - \phi \log \phi + \log \Gamma(\phi) , \\ \rho^{(1,0)} &= \phi , \quad \rho^{(0,1)} = \theta + \log \phi + \psi(\phi) - 1 , \\ \alpha_1^{(1,0)} &= \phi e^{-\theta} , \quad \alpha_1^{(0,1)} = -e^{-\theta} , \quad \alpha_2^{(0,1)} = 1 . \end{aligned}$$

De (5.21) temos

$$\beta_1 = e^\theta , \quad \beta_2 = \theta + \psi(\theta) - \log \phi .$$

As outras derivadas envolvidas nas expressões (5.22) e (5.23) têm os seguintes valores:

$$\begin{aligned} \alpha_1^{(1,1)} &= e^{-\theta} , \quad \alpha_1^{(1,2)} = 0 , \quad \alpha_1^{(2,0)} = -\phi e^{-\theta} , \quad \alpha_1^{(2,1)} = -e^{-\theta} , \\ \alpha_1^{(3,0)} &= \phi e^{-\theta} , \quad \alpha_1^{(0,1)} = -e^{-\theta} , \quad \alpha_1^{(0,2)} = 0 , \\ \beta_1^{(1,0)} &= e^\theta = \beta_1^{(2,0)} = \beta_1^{(3,0)} , \quad \beta_1^{(1,1)} = \beta_1^{(1,2)} = \beta_1^{(0,1)} = \beta_1^{(0,2)} = 0 , \\ \alpha_2^{(0,1)} &= 1 , \\ \beta_1^{(0,1)} &= \psi' - \frac{1}{\phi} , \quad \beta_2^{(0,2)} = \psi'' + \frac{1}{\phi^2} , \quad \beta_2^{(1,0)} = 1 , \quad \beta_2^{(1,1)} = 0 . \end{aligned}$$

Substituindo estas derivadas em (5.22) e (5.23) - (5.23'') encontramos

$$c = 1 + \frac{1}{12n\phi} \left[2 - \frac{3(2\phi^2\psi'' + \phi\psi' + 1)}{(\phi\psi' - 1)^2} \right] .$$

A4.2.2 TESTE DAS HIPÓTESES $H : \phi = \phi^{(0)}$ VERSUS $A : \phi \neq \phi^{(0)}$

1. Modelo normal $N(\theta, \phi)$

Para o teste sobre o parâmetro ϕ , devemos tomar derivadas em relação ao vetor de parâmetros $(\phi, \theta)^\top$ das funções $\alpha_1, \alpha_2, \beta_1$ e β_2 associadas à densidade (A4.12). As parcelas A_ϕ e $C_{\theta, \phi}$ da

correção de Bartlett $c = 1 + A_\phi + C_{\theta,\phi}$, são obtidas de (5.16) e (5.24) - (5.25), respectivamente.

As derivadas necessárias ao cálculo destas quantidades estão apresentadas a seguir.

$$\begin{aligned}\alpha_1^{(1,0)} &= -\frac{\theta}{\phi^2}, \quad \alpha_1^{(1,1)} = -\frac{1}{\phi^2}, \quad \alpha_1^{(1,2)} = 0, \quad \alpha_1^{(2,0)} = \frac{2\theta}{\phi^3}, \quad \alpha_1^{(2,1)} = \frac{2}{\phi^3}, \quad \alpha_1^{(3,0)} = -\frac{6\theta}{\phi^4}, \\ \alpha_1^{(0,1)} &= \frac{1}{\phi}, \quad \alpha_1^{(0,2)} = 0, \\ \beta_1^{(0,1)} &= 1, \quad \beta_1^{(0,2)} = \beta_1^{(i,0)} = 0 = \beta_1^{(i,j)}, \quad i, j = 1, 2, \\ \alpha_2^{(1,0)} &= \frac{1}{2\phi^2}, \quad \alpha_2^{(2,0)} = -\frac{1}{\phi^3}, \quad \alpha_2^{(3,0)} = \frac{3}{\phi^4}, \quad \beta_2^{(1,0)} = 1, \quad \beta_2^{(1,1)} = \beta_2^{(2,0)} = 0.\end{aligned}$$

A substituição destas derivadas em (5.16), (5.24) e (5.25) produz $c = 1 + A_\phi + (C_{1\theta,\phi} - C_{2\theta,\phi}) = 1 + \frac{1}{3n} + \left(\frac{2}{n} - \frac{3}{2n}\right) = 1 + \frac{11}{6n}$.

2. Modelo normal inverso NI(θ, ϕ)

Considerando a densidade (A4.13), as derivadas das correspondentes funções $\alpha_1, \alpha_2, \beta_1$ e β_2 em relação ao vetor $(\phi, \theta)^\top$ são apresentadas a seguir.

$$\begin{aligned}\alpha_1^{(1,0)} &= -\frac{1}{2\theta^2}, \quad \alpha_1^{(1,1)} = \frac{1}{\theta^3}, \quad \alpha_1^{(1,2)} = -\frac{3}{\theta^4}, \quad \alpha_1^{(2,0)} = 0 = \alpha_1^{(2,1)} = \alpha_1^{(3,0)}, \\ \alpha_1^{(0,1)} &= \frac{\phi}{\theta^3}, \quad \alpha_1^{(0,2)} = -\frac{3\phi}{\theta^4}, \\ \beta_1^{(0,1)} &= 1, \quad \beta_1^{(0,2)} = \beta_1^{(i,0)} = 0 = \beta_1^{(i,j)}, \quad i, j = 1, 2, \\ \alpha_2^{(1,0)} &= -\frac{1}{2}, \quad \alpha_2^{(2,0)} = 0 = \alpha_2^{(3,0)}, \quad \beta_2^{(1,0)} = -\frac{1}{\phi^2}, \quad \beta_2^{(1,1)} = 0, \quad \beta_2^{(2,0)} = \frac{2}{\phi^3}.\end{aligned}$$

Substituindo as derivadas acima em (5.16), (5.24) e (5.25) obtemos $A_\phi = \frac{1}{3n}$, $C_{1\theta,\phi} = 0$, $C_{2\theta,\phi} = -\frac{3}{2n}$, que resulta $c = 1 + \frac{11}{6n}$.

3. Modelo gama G(θ, ϕ)

Para a distribuição gama, com variância $\frac{\theta^2}{\phi}$, cuja densidade é expressa em (A4.14), as derivadas das funções $\alpha_1, \alpha_2, \beta_1$ e β_2 correspondentes, em relação ao vetor $(\phi, \theta)^\top$, são dadas por

$$\alpha_1^{(1,0)} = -\frac{1}{\theta}, \quad \alpha_1^{(1,1)} = \frac{1}{\theta^2}, \quad \alpha_1^{(1,2)} = -\frac{2}{\theta^3}, \quad \alpha_1^{(2,0)} = \alpha_1^{(3,0)} = \alpha_1^{(2,1)} = 0,$$

$$\begin{aligned}\alpha_1^{(0,1)} &= \frac{\phi}{\theta^2}, \quad \alpha_1^{(0,2)} = -\frac{2\phi}{\theta^3}, \\ \beta_1^{(0,1)} &= 1, \quad \beta_1^{(0,2)} = \beta_1^{(i,0)} = 0 = \beta_1^{(i,j)}, \quad i, j = 1, 2, \\ \alpha_2^{(1,0)} &= 1, \quad \alpha_2^{(2,0)} = 0, \quad \beta_2^{(1,0)} = \psi'(\phi) - \frac{1}{\phi}, \quad \beta_2^{(1,1)} = 0, \quad \beta_2^{(2,0)} = \psi''(\phi) + \frac{1}{\phi^2}.\end{aligned}$$

Para obtermos $c = 1 + A_\phi + C_{\theta,\phi}$, substituímos as derivadas acima nas expressões (5.16), (5.24) e (5.25), produzindo

$$c = 1 + \frac{1}{12n\phi} \left[\frac{5(\phi^2\psi'' + 1)^2}{(\phi\psi' - 1)^3} - \frac{3\phi^2(\phi\psi''' + 2\psi'')}{(\phi\psi' - 1)^2} - \frac{3}{(\phi\psi' - 1)} \right].$$

4. Modelo log-gama LG(θ, ϕ)

As derivadas das funções α_1 , α_2 , β_1 e β_2 em relação ao vetor $(\phi, \theta)^\top$, correspondentes à função densidade em (A4.15), são

$$\begin{aligned}\alpha_1^{(1,0)} &= -e^{-\theta}, \quad \alpha_1^{(1,1)} = e^{-\theta}, \quad \alpha_1^{(1,2)} = -e^{-\theta}, \quad \alpha_1^{(0,1)} = \phi e^{-\theta} = \alpha_1^{(0,3)}, \\ \alpha_1^{(0,2)} &= -\phi e^{-\theta}, \quad \alpha_1^{(2,0)} = 0 = \alpha_1^{(2,1)} = \alpha_1^{(3,0)}, \\ \beta_1^{(1,0)} &= 0 = \beta_1^{(1,1)} = \beta_1^{(2,0)} = \beta_1^{(3,0)}, \quad \beta_1^{(0,1)} = e^\theta = \beta_1^{(0,2)} = \beta_1^{(0,3)}, \\ \alpha_2^{(1,0)} &= 1, \quad \alpha_2^{(2,0)} = 0, \quad \beta_2^{(1,0)} = \psi' - \frac{1}{\phi}, \quad \beta_2^{(2,0)} = \psi'' + \frac{1}{\phi^2}, \quad \beta_2^{(3,0)} = \psi''' - \frac{2}{\phi^3}, \quad \beta_2^{(1,1)} = 0.\end{aligned}$$

Por substituição destas derivadas nas expressões (5.16), (5.24) e (5.25) temos que a correção de Bartlett neste problema é obtida por

$$c = 1 + \frac{1}{12n\phi} \left[\frac{5(\phi^2\psi'' + 1)^2}{(\phi\psi' - 1)^3} - \frac{3\phi^2(\phi\psi''' + 2\psi'')}{(\phi\psi' - 1)^2} - \frac{3}{(\phi\psi' - 1)} \right].$$

REFERÊNCIAS BIBLIOGRÁFICAS

- ABELL, M.L.; BRASELTON, J.P. (1994). *The maple handbook*. New York: AP Professional.
- ABRAMOWITZ, M.; STEGUN, I.A. (1972). *Handbook of mathematical functions*. New York, Addison-Wesley. 640p.
- ANDERSON, J.A.; RICHARDSON, S.C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, **21**, 71–78.
- ATTFIELD, C.L.F. (1991). A Bartlett adjustment to the likelihood ratio test for homoskedasticity in the linear model. *Economics Letters*, **37**, 119–123.
- BARNDORFF-NIELSEN, O.E. (1978). *Information and exponential families in statistical theory*. Chichester, John Wiley (Wiley Series in Probability and Mathematical Statistics - Tracts).
- BARNDORFF-NIELSEN, O.E. (1990). Likelihood theory. In: Hinkley, D.V., Reid, N. and Snell, E.J., eds. *Statistical theory and modelling: in honor to Sir David Cox, FRS*. London, Chapman and Hall, 232–264.
- BARNDORFF-NIELSEN, O.E.; BLAESILD, P. (1986). A note on the calculation of Bartlett adjustments. *Journal of the Royal Statistical Society, B*, **48**(3), 353–358.
- BARNDORFF-NIELSEN, O.E.; COX, D.R. (1984a). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society, B*, **46**, 483–495.
- BARNDORFF-NIELSEN, O.E.; COX, D.R. (1984b). The effect of sampling rules on likelihood statistics. *International Statistical Review*, **52**, 309–326.

- BARTLETT, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, A*, **160**, 268–282.
- BARTLETT, M.S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, **34**, 33–40.
- BARTLETT, M.S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Supplement*, **9**, 176–197.
- BARTLETT, M.S. (1953). Approximate confidence intervals II. *Biometrika*, **40**, 306–317.
- BARTLETT, M.S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society, B*, **16**, 296–298.
- BICKEL, P.J.; DOKSUM, K.A. (1977). *Mathematical statistics: basic ideal and selected topics*. Oakland, Holden-Day.
- BIRCH, M.W. (1963). An introduction for the logarithmic series distribution. *Biometrics*, **19**, 651–652.
- BOTTER, D.A. (1993). *Correções de Bartlett e poderes de testes em algumas classes de modelos de regressão*. São Paulo. 207p. Tese (Doutorado) - Instituto de Matemática e Estatística da Universidade de São Paulo.
- BURR, I.W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, **13**, 215–232.
- CHANDRA, T.K. (1985). Asymptotic expansions of perturbed chi-squared variables. *Sankhyā A*, **47**, 100–110.
- CHANDRA, T.K.; GHOSH, J.K. (1979). Valid asymptotic expansions for the likelihood ratio statistics and other perturbed chi-squared variables. *Sankhyā A*, **41**, 22–47.
- CHANDRA, T.K.; MUKERJEE, R. (1991). Bartlett-type modification for Rao's efficient score statistic. *Journal of Multivariate Analysis*, **36**, 103–112.
- CHANDRASEKHAR, S. (1943). Stochastic problems in physics and astronomy. *Review of Modern Physics*, **15**, 2–89.
- CHAPMAN, D.G. (1956). Estimating the parameters of a truncated gamma distribution. *Annals of Mathematical Statistics*, **27**, 498–506.
- CHAR, B.W.; GEDDES, K.O.; GONNET, M.B.; MONAGAN, M.B.; WATT, S.M. (1988). *Maple reference manual*. Waterloo, WATCOM Publications Ltd., 5.ed.

- CHATFIELD, C. (1986). Discrete distributions and purchasing models. *Communications in Statistics: Theory and Methods*, **15**, 697–708.
- CHATFIELD, C.; EHRENBURG, A.S.C.; GOODHARDT, G.J. (1966). Progress on a simplified model of stationary purchasing behavior. *Journal of the Royal Statistical Society, A*, **129**, 317–367.
- COOK, R.D.; TSAI, C.L.; WEI, B.C. (1986). Bias in nonlinear regression. *Biometrika*, **13**, 615–623.
- COOK, R.D.; WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1–10.
- COPAS, J.B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society, B*, **48**, 225–265.
- CORDEIRO, G.M. (1982). *Improved likelihood ratio statistics for generalized linear models*. London. 227p. Tese (Doutorado) - Imperial College of Science and Technology - University of London.
- CORDEIRO, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society, B*, **45**, 404–413.
- CORDEIRO, G.M. (1985). The null expected deviance for an extended class of generalized linear models. *Proceedings of the International Conference on Generalized Linear Models, Lecture Notes in Statistics*, **32**, 27–34. Berlin, Springer-Verlag.
- CORDEIRO, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**, 265–274.
- CORDEIRO, G.M. (1992). *Introdução à teoria de verossimilhança*. Rio de Janeiro, UFRJ. 174p. Minicurso apresentado no 10o. Simpósio Nacional de Probabilidade e Estatística, Rio de Janeiro, 1992.
- CORDEIRO, G.M. (1993a). General matrix formulae for computing Bartlett corrections. *Statistics and Probability Letters*, **16**, 11–18.
- CORDEIRO, G.M. (1993b). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, A*, **22**, 169–188.
- CORDEIRO, G.M. (1995). Performance of a Bartlett-type modification for the deviance. *Journal of Statistical Computation and Simulation*, **51**, 385–403.

- CORDEIRO, G.M.; FERRARI, S.L.P. (1991). A modified score statistic having chi-squared distribution to order n^{-1} . *Biometrika*, **78**, 573–582.
- CORDEIRO, G.M.; KLEIN, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, **19**, 169–176.
- CORDEIRO, G.M.; McCULLAGH, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, B*, **53**, 629–643.
- CORDEIRO, G.M.; PAULA, G.A. (1989). Improved likelihood statistics for exponential nonlinear models. *Biometrika*, **76**, 93–100.
- CORDEIRO, G.M.; PAULA, G.A.; BOTTER, D.A. (1994). Improved likelihood tests for dispersion models. *International Statistical Review*, **62**, 2, 257–274.
- CORDEIRO, G.M.; PÉREZ-ABREU, V. (1995). Bartlett corrections for V -statistics. *RE-BRAPE: Revista Brasileira de Probabilidade e Estatística*, **9**. A aparecer.
- COX, D.R. (1988). Some aspects of conditional and asymptotic inference: a review. *Sankhyā A*, **50**(3), 314–337.
- COX, D.R.; HINKLEY, D.V. (1974). *Theoretical statistics*. New York, Wiley. 511p.
- COX, D.R.; REID, N. (1987). Parameters orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society, B*, **49**, 1–39.
- COX, D.R.; SNELL, E.J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society, B*, **42**, 248–275.
- CRIBARI-NETO, F.; CORDEIRO, G.M. (1995). Bartlett and Bartlett-type corrections. *Econometric Reviews*. Submetido.
- CRIBARI-NETO, F.; FERRARI, S.L.P. (1995). An improved Lagrange multiplier test for heteroskedasticity. *Communication in Statistics: Simulation and Computation*, **24**, 31–44.
- EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics*, **6**, 1189–1242.
- ENGLE, R.F. (1984). Wald, likelihood ratio and Lagrange multiplier tests in econometrics. In: Griliches, Z. and Intriligator, M., eds. *Handbook of Econometrics*. Amsterdam, North-Holland, 776–826. (Handbook of Econometrics, 2).
- EVANS, R.D. (1955). *The atomic nucleus*. New York, McGraw-Hill Book Co.

- FELLER, W. (1957). *An introduction to probability theory and its applications*. V.1, 2.ed. New York, John Wiley and Sons.
- FERRARI, S.L.P. (1991). *Aperfeiçoamento de testes escore, aplicações e extensões*. São Paulo, 185p. Tese (Doutorado) - Instituto de Matemática e Estatística da Universidade de São Paulo.
- FERRARI, S.L.P.; CORDEIRO, G.M. (1994). Matrix formulae for computing improved score tests. *Journal of Statistical Computation and Simulation*, **49**, 195–206.
- FISHER, R.A.; CORBET, A.S.; WILLIAMS, C.B. (1943). The relation between the number of species and number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.
- FRYDENBERG, M.; JENSEN, J.L. (1989). Is the “improved likelihood statistic” really improved in the discrete case? *Biometrika*, **76**, 655–661.
- GALBRAITH, R.F.; GALBRAITH, J.I. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Statistics*, **11**, 63–71.
- GLEJSER, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical Association*, **64**, 316–323.
- GOLDFELD, S.M.; QUANDT, R.E. (1965). Some tests for heteroscedasticity. *Journal of the American Statistical Association*, **69**, 539–547.
- GREENWOOD, J.A.; DURAND, D. (1960). Aids for fitting gamma distribution by maximum likelihood. *Technometrics*, **2**, 55–65.
- GRUNTZ, D.; MONAGAN, M. (1993). Introduction to Gauss. *Maple Technical Newsletter*, **9**, 23–35.
- HAIGHT, F.A. (1967). *Handbook of the Poisson distribution*. Publications in operations research No. 11. New York, John Wiley.
- HALDANE, J.B.S.; SMITH, S.M. (1956). The sampling distribution of a maximum likelihood estimate. *Biometrika*, **63**, 96–103.
- HARRIS, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistics. *Biometrika*, **72**(3), 653–659.
- HARRIS, P. (1986). A note on Bartlett adjustments to likelihood ratio tests. *Biometrika*, **73**, 735–737.

- HARRIS, P. (1987). Discussion of the paper by Professors COX and REID. *Journal of the Royal Statistical Society, B*, 49(1), 30.
- HARRISON, M.J.; McCABE, B.P.M. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *Journal of the American Statistical Association*, 74, 494–500.
- HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics, A*, 29, 359–78, and *Corrigenda in Annals of the Institute of Statistical Mathematics*, 39, 681 (1987).
- HAYAKAWA, T.; PURI, M.L. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics*, 37, 95–108.
- HEARN, A.C., ed. (1984). *Reduce user's manual*. Version 3.1. (The Rand Corporation).
- HILL, G.W.; DAVIS, A.W. (1968). Generalized asymptotic expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics*, 39, 1246–1273.
- JARQUE, C.M. (1981). A test for heteroscedasticity in a limited dependent variable model. *The Australian Journal of Statistics*, 23, 159–163.
- JENSEN, J.L. (1993). A historical sketch and some new results on the improved log likelihood ratio statistics. *Scandinavian Journal of Statistics*, 20, 1–15.
- JOHNSON, N.L.; KOTZ, S. (1970a). *Continuous univariate distributions*. V.1. Boston, Houghton Mifflin.
- JOHNSON, N.L.; KOTZ, S. (1970b). *Continuous univariate distributions*. V.2. Boston, Houghton Mifflin.
- JØRGENSEN, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society, B*, 49, 127–162.
- KENDALL, M.G.; STUART, A. (1969). *The advanced theory of statistics*. V.1: distribution theory. London, Charles Griffin.
- KOENKER, R.; BASSET, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrika*, 50, 43–64.
- LAWLEY, D.N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, 43, 295–303.
- LETAC, G.; MORA, M. (1990). Natural real exponential families with cubic variance function. *Annals of Statistics*, 10, 1–37.

- LINDLEY, D.V. (1965). *Introduction to probability and statistics from a Bayesian view point. Part I, Probability*. Cambridge, Cambridge University Press.
- LJUNG, G.M.; BOX, G.E.P. (1979). The likelihood function of stationary autorregressive-moving average models. *Biometrika*, **66**, 265–270.
- MARDIA, K.V. (1972). *Statistics of directional data*. London, Academic Press, 357p.
- MARDIA, K.V.; MARSHALL, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71**, 135–146.
- MASUYAMA, M.; KUROIWA, W. (1952). Table for the likelihood solutions of gamma distribution and its medical applications. *Reports of Statistical Application Research (JUSE)*, **1**, 18–23.
- McCULLAGH, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika*, **71**, 461–476.
- McCULLAGH, P. (1985). On the asymptotic distribution of Pearson's statistic in linear exponential-family models. *International Statistical Review*, **53**, 61–67.
- McCULLAGH, P. (1987). *Tensor methods in statistics*. London, Chapman and Hall. 285p.
- McCULLAGH, P. (1989). Some statistical properties of a family of continuous univariate distribution. *Journal of the American Statistical Association*, **84**, 125–141.
- McCULLAGH, P.; COX, D.R. (1986). Invariants and the likelihood ratio statistic. *The Annals of Statistics*, **14**, 1419–1430.
- McCULLAGH, P.; NELDER, J.A. (1989). *Generalized linear models*. 2.ed. London, Chapman and Hall. 511p.
- McLACHLAN, P. (1980). A note on bias corrections in maximum likelihood estimation with logistic discrimination. *Technometrics*, **22**, 621–627.
- MØLLER, J. (1986). Bartlett adjustments for structured covariances. *Scandinavian Journal of Statistics*, **13**, 1–15.
- MORRIS, C.N. (1982). Natural exponential families with quadratic variance functions. *Annals of Statistics*, **10**, 65–80.
- MOULTON, L.H.; WEISSFELD, L.A.; LAURENT, R.T. (1993). Bartlett correction factors in logistic regression models. *Computational Statistics and Data Analysis*, **15**, 1–11.

- MUNROE, M.E. (1951). *Theory of probability*. New York, McGraw-Hill Book Co.
- NELDER, J.A.; WEDDERBURN, R.W.N. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A*, **135**, 370–384.
- NEYMAN, J.; PEARSON, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20**(A), 175–240 e 263–294.
- PATTERSON, S.J. (1988). *An introduction to the theory of the Riemann zeta-function*. New York, Cambridge University Press.
- PORTEOUS, B.T. (1985). Improved likelihood ratio statistics for covariance selection models. *Biometrika*, **72**, 97–101.
- PRESS, S.J. (1982). *Applied multivariate analysis including Melbourne Bayesian and frequentist methods of inference*. Melbourne, Krieger Publishing Company.
- RAYLEIGH, J.W.S. (1919). Random vibrations and random flights in one, two or three dimensions. *Philosophical Magazine*, S.6, **37**, 321–347.
- RAYNA, G. (1987). *REDUCE: software for algebraic computation*. New York, Springer. 329p. (Symbolic Computation, Artificial Intelligence).
- REID, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, **3**, 213–238.
- ROSS, W.H. (1987). The expectation of the likelihood ratio criterion. *International Statistical Review*, **55**, 315–330.
- ROTHENBERG, T.J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica*, **52**, 827–842.
- SCHAFFER, R.L. (1983). Bias correction in maximum likelihood logistic regression. *Statistics in Medicine*, **2**, 71–78.
- SHAMAN, P. (1973). On the inverse of the covariance matrix for an autorregressive-moving average process. *Biometrika*, **60**, 193–196.
- SHENTON, L.R.; BOWMAN, K. (1963). Higher moments of a maximum likelihood estimate. *Journal of the Royal Statistical Society, B*, **25**, 305–317.
- SHENTON, L.R.; BOWMAN, K. (1969). Maximum likelihood estimate moments for the two-parameter gamma distribution. *Sankhyā B*, **31**, 379–396.

- SHENTON, L.R.; BOWMAN, K. (1977). *Maximum likelihood estimation in small samples*. London, Griffin.
- SHENTON, L.R.; WALLINGTON, P.A. (1962). The bias of moment estimators with an application to the relative binomial distribution. *Biometrika*, **49**, 193-204.
- TANAKA, K. (1984). An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *Journal of the Royal Statistical Society, B*, **46**, 58-67.
- TANIGUCHI, M. (1991). Third-order asymptotic properties of a class of test statistics under a local alternative. *Journal of Multivariate Analysis*, **37**, 223-238.
- WILKS, S.S. (1962). *Mathematical statistics*. New York, John Wiley.
- WILLIAMS, C.B. (1944). The logarithm series and its applications to biological problems. *Journal of Ecology*, **34**, 253-272.
- WILLIAMS, C.B. (1964). *Patterns in balance of nature*. London, Academic Press.
- WILLIAMS, D.A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika*, **16**, 141-148.
- WOLFRAM, S. (1991). *Mathematica: a system for doing mathematics by computer*. 2.ed. New York, Addison-Wesley.