

Ajustes para a verossimilhança perfilada em modelos lineares generalizados

Fernando Lucambio Pérez

Tese apresentada ao
Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo
para a obtenção do grau de
Doutor em Estatística

Área de Concentração: Estatística

Orientadora: Prof. Dra. Silvia Lopes de Paula Ferrari

São Paulo, Brasil

Junho de 2003

Ajustes para a verossimilhança perfilada em modelos lineares generalizados

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Fernando Lucambio Pérez e aprovada pela comissão julgadora.

São Paulo, junho de 2003.

Banca examinadora

- Profa. Dra. Silvia Lopes de Paula Ferrari (Orientadora), IME, USP
- Prof. Dr. Francisco Cribari Neto, UFPE
- Prof. Dr. Gauss Moutinho Cordeiro, UFBA
- Profa. Dra. Denise Aparecida Botter, IME, USP
- Prof. Dr. Filidor Edilson Vilca Labra, IMECC, UNICAMP

Summary

Inference about a multiparameter of interest in the presence of nuisance parameters is oftentimes based on the profile likelihood function. However, it does not behave as a true likelihood function and several adjustments to the profile likelihood function have been proposed. In this dissertation, we consider an additive adjustment that aims at reducing the score and information biases of the profile score function from $O(1)$ to order $O(n^{-1})$. The adjustment was originally proposed by Stern (1997) but we show that his formula is in error. One of the goals of this dissertation is to obtain the correct expression for Stern's adjustment. It is noteworthy that this adjustment is applicable in wide generality since it allows both the interest and nuisance parameters to be vector-valued. Our second goal is to derive a Bartlett correction for the adjusted profile likelihood ratio statistic. We also obtain simple closed-form expressions for Stern's adjustment and its Bartlett correction in the class of the generalized linear models. A simulation study compares the performance of the usual profile likelihood ratio test, the adjusted profile likelihood ratio test and the associated Bartlett-corrected tests.

Agradeço o apoio financeiro do CNPq e da FINEP.

Agradeço à minha família brasileira.

A mi familia cubana. A mis padres.

Sumário

1	Introdução	8
1.1	Definições gerais	9
1.1.1	A função de verossimilhança perfilada	12
1.1.2	Notação	16
1.2	Função de verossimilhança perfilada ajustada	18
1.3	Modelos lineares generalizados	22
1.3.1	Definição e algumas propriedades	22
1.3.2	Estimação e teste da razão de verossimilhanças	25
2	Ajuste para a verossimilhança perfilada	28
2.1	Função de verossimilhança perfilada ajustada	29
2.1.1	Ajuste para o vício da esperança da função escore perfilada	29
2.1.2	Ajuste para o vício da informação da função escore perfilada	35
2.1.3	Exemplo: modelo de regressão linear normal	42
2.2	Correção de Bartlett para o teste da razão de verossimilhanças ajustado	48
2.2.1	Exemplo: modelo de regressão linear normal	54
3	Ajuste para a verossimilhança perfilada em modelos lineares generalizados	59
3.1	Função de verossimilhança perfilada ajustada nos modelos lineares generalizados	60
3.1.1	Dispersão como parâmetro de interesse	60
3.1.2	Exemplo: modelo de regressão normal linear	63

3.1.3	Parte do vetor de parâmetros de regressão como de interesse e parâmetro de dispersão conhecido	68
3.1.4	Parte do vetor de parâmetros de regressão como de interesse e parâmetro de dispersão desconhecido	75
3.1.5	Exemplo: todos os parâmetros de regressão de interesse	81
3.1.6	Exemplo: distribuição normal; média como parâmetro de interesse	83
3.2	Correção de Bartlett para o teste da razão de verossimilhanças perfiladas ajustadas nos modelos lineares generalizados	87
3.2.1	Teste do parâmetro de dispersão	87
3.2.2	Teste dos parâmetros de regressão	90
3.2.3	Exemplo: modelo de regressão normal	99
3.2.4	Exemplo: modelo de regressão gama	102
4	Considerações finais e conclusões	110
A	Algumas relações entre cumulantes	112
A.1	Forma alternativa da soma $\sigma_{ab}\lambda^{br}$	112
A.2	Matriz de cumulantes ν	113
A.3	Obtenção de A_r^c	114
B	Cumulantes de ordem superior nos modelos lineares generalizados	116
B.1	Obtenção dos cumulantes $\lambda_{rst/u}$ e λ_{rstu}	120
C	Propriedades do produto Hadamard e da função traço	122
	Referências Bibliográficas	125

Lista de Figuras

1.1	Gráfico da função $d_1''(\phi) = 1/\phi - d^2 \log \Gamma(\phi)/d\phi^2$ no modelo gama.	24
3.1	Comportamento médio, em cem amostras, do logaritmo das funções de verossimilhança perfilada e perfilada ajustada. Parâmetro de interesse $\phi = 1$, $n = 10$ e vetor de parâmetros de regressão $\beta = (1, 1, 1)^\top$	65
3.2	Média, em cem amostras, da discrepância dos quantis amostrais das estatísticas da razão de verossimilhanças perfiladas, perfiladas ajustadas e as correspondentes estatísticas com correção de Bartlett no modelo de regressão gama. Vetor de parâmetros de interesse $\beta = (0, 0, 0, 0)$ ($p = 4$), $n = 20$ e parâmetros de perturbação $\beta = (1, 1, 1)$ e $\phi = 1$, estes desconhecidos.	105
3.3	Média, em cem amostras, da discrepância dos quantis amostrais das estatísticas da razão de verossimilhanças perfiladas, perfiladas ajustadas e as correspondentes estatísticas com correção de Bartlett no modelo de regressão gama. Vetor de parâmetros de interesse $\beta = (0, 0, 0, 0)$ ($p = 4$), $n = 25$ e vetor de parâmetros de regressão $\beta = (1, 1, 1)$ e $\phi = 1$, estes desconhecidos.	105

Lista de Tabelas

1.1	Ligações canônicas de diferentes modelos e segunda derivada da função $d_1(\phi)$.	23
3.1	Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(1))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(1))$, $\alpha = 0,15; 0,10; 0,05; 0,01; 0,005$ e $0,0005$. Valores percentuais em 100.000 réplicas para testar $H_0 : \phi = 1$ no modelo de regressão linear normal com β sendo o vetor de parâmetros de perturbação.	89
3.2	Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(1))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(1))$, $\alpha = 0,15; 0,10; 0,05; 0,01; 0,005$ e $0,0005$. Valores percentuais em 100.000 réplicas para testar $H_0 : \phi = 1$ no modelo de regressão linear normal com β sendo o vetor de parâmetros de perturbação.	90
3.3	Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$, valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com ϕ como o parâmetro de perturbação ($\phi = 1$) no modelo de regressão linear normal, número de parâmetros de interesse $p = 1, 2, 3$ e 4 e tamanho de amostra $n = 10$ e 15 .	100

- 3.4 Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$, valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com ϕ como o parâmetro de perturbação ($\phi = 1$) no modelo de regressão linear normal, número de parâmetros de interesse $p = 1, 2, 3, 4$ e 5 e tamanho de amostra $n = 20$ e $n = 25$ 102
- 3.5 Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4 , número de parâmetros de perturbação $q = 2$ e 3 , $\phi = 1$ e tamanho de amostra $n = 15$ 103
- 3.6 Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4 , número de parâmetros de perturbação $q = 2, 3, 4$ e 5 , $\phi = 1$ e tamanho de amostra $n = 20$ 106
- 3.7 Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4 , número de parâmetros de perturbação $q = 2, 3, 4$ e 5 , $\phi = 1$ e tamanho de amostra $n = 25$ 107

3.8 Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_{\alpha}^2(p))$, $P(\mathcal{W}^* \geq \chi_{\alpha}^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_{\alpha}^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_{\alpha}^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 0, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4 , número de parâmetros de perturbação $q = 2, 3, 4$ e 5 , $\phi = 1$ e tamanho de amostra $n = 30$ 108

Capítulo 1

Introdução

Esta tese tem por objetivo estudar uma situação dentro da teoria da estimação estatística muito considerada recentemente. Referimo-nos ao estudo de inferência baseada em verossimilhanças perfiladas. Como exemplo de trabalhos recentes neste sentido podemos mencionar Barndorff-Nielsen (1986), Cox & Reid (1987), Davison (1988), McCullagh & Tibshirani (1990), Barndorff-Nielsen (1992, 1994), DiCiccio, Martin, Stern & Young (1996), Stern (1997), Severini (1998b), Davison & Stafford (1998) e outros aos quais faremos referência posteriormente.

As verossimilhanças perfiladas são utilizadas quando parte do vetor de parâmetros de um modelo estatístico não é de interesse direto do pesquisador e é, portanto, considerado como de perturbação. A utilização de verossimilhanças perfiladas induz certas dificuldades e algumas das soluções propostas serão objeto de estudo.

Os resultados obtidos nesta tese são de ampla aplicabilidade no sentido de que somente exigem modelos estatísticos nos quais o estimador de máxima verossimilhança exista e possua as propriedades assintóticas usuais. A aplicação dos resultados em modelos de regressão, em particular nos modelos lineares generalizados (McCullagh & Nelder, 1989), será discutida amplamente.

Definições gerais necessárias ao desenvolvimento deste trabalho serão apresentadas a seguir.

1.1 Definições gerais

Consideremos $Y = (Y_1, \dots, Y_n)$ um vetor de variáveis aleatórias com função densidade conjunta $f(y; \vartheta)$, sendo $\vartheta \in \Omega$, $\vartheta = (\vartheta^1, \dots, \vartheta^{p+q})^\top$ o vetor de parâmetros desconhecidos, de dimensão $p + q$, que descreve o problema a ser estudado, Ω , o espaço paramétrico, considerado um subconjunto de \mathbb{R}^{p+q} . O vetor de variáveis aleatórias Y também pode ser discreto, e desta forma ter associada uma função de probabilidade, mas não faremos distinção entre as situações discretas e absolutamente contínuas e por esta razão faremos referência a $f(y; \vartheta)$, genericamente, como uma função densidade.

Definição 1.1.1 *A função de verossimilhança do parâmetro ϑ será escrita como $\ell(\vartheta)$ e definida como*

$$\ell(\vartheta) = \ell(\vartheta; y) = \prod_{i=1}^n f_i(y_i; \vartheta),$$

se as variáveis aleatórias Y_1, \dots, Y_n forem independentes, onde $f_i(y_i; \vartheta)$ é a função densidade de Y_i avaliada no ponto y_i , $i = 1, \dots, n$.

Em situações mais gerais diremos somente que $\ell(\vartheta) = f(y; \vartheta)$ é a função de verossimilhança de ϑ . ■

Nos exemplos considerados neste trabalho a função de verossimilhança é escrita como $\ell(\vartheta) = \prod_{i=1}^n f(y_i; \vartheta^i)$, já que neles a função de densidade é a mesma, somente avaliada em pontos amostrais e valores dos parâmetros diferentes.

Observemos que, embora $\ell(\vartheta)$ seja uma função de variáveis aleatórias, estaremos interessados em ressaltar sua dependência do vetor de parâmetros e não seu caráter aleatório. Por isso, no argumento desta e de outras funções de verossimilhança a serem definidas, consideraremos somente o vetor de parâmetros que define o modelo.

Para obter o estimador de máxima verossimilhança $\hat{\vartheta}$ de ϑ , e suas propriedades estatísticas, utilizaremos o logaritmo da função de verossimilhança, denotado por $L(\vartheta)$ e definido como

$$L(\vartheta) = \log \ell(\vartheta).$$

A partir desta função, obtemos a função de estimação $U(\vartheta)$ (Godambe, 1960), definida como

$$U(\vartheta) = \frac{\partial}{\partial \vartheta} L(\vartheta)$$

e conhecida como função escore. O estimador de máxima verossimilhança do vetor de parâmetros ϑ é obtido da equação de estimação

$$U(\hat{\vartheta}) = 0,$$

a qual consideraremos que possui solução única.

Assumiremos que a função de verossimilhança e, portanto, a função escore, satisfazem condições de regularidade para garantir que o estimador de máxima verossimilhança de ϑ seja consistente e com distribuição limite conhecida. Estas condições são amplamente conhecidas na literatura, mas utilizaremos tais condições aplicadas à função escore segundo foram apresentadas em Godambe (1960). Mais recentemente encontramos estas condições de regularidade escritas de maneira mais geral, por exemplo, em Sweeting (1980) e Knudsen (1998).

Este trabalho trata de problemas conhecidos na literatura estatística como de teoria assintótica. Nele admitimos que o número de observações n cresce mas a dimensão do vetor de parâmetros não, mantendo-se constante.

Utilizaremos amplamente o conceito de ordem de magnitude de seqüências estocásticas ou $O_P(\cdot)$ e $o_P(\cdot)$, também conhecido como notação de Mann-Wald; recomendamos a leitura de Bishop, Fienberg & Holland (1975), McCullagh (1987), Cox, Hinkley, Reid & Snell (1991), Sen & Singer (1993) e Barndorff-Nielsen & Cox (1994). Recordemos que $X_n = O_P(n^a)$ se a seqüência $\{n^{-a}X_n\}_{n \geq 1}$ for limitada em probabilidade, quando $n \rightarrow \infty$. Um exemplo muito importante de variável aleatória de ordem $O_P(n^{1/2})$ é a função escore. Recordemos também que se $\{n^{-a}X_n\}_{n \geq 1}$ converge em probabilidade para zero, quando $n \rightarrow \infty$, então $X_n = o_P(n^a)$.

Outros conceitos que serão de grande utilização neste trabalho são a matriz de informação observada $j(\vartheta)$, onde

$$j(\vartheta) = -\frac{\partial^2 L(\vartheta)}{\partial \vartheta \partial \vartheta^T}$$

e a matriz de informação esperada ou matriz de informação de Fisher, definida como

$$K = \mathbb{E} \left\{ \frac{\partial L}{\partial \vartheta} \frac{\partial L}{\partial \vartheta^\top} \right\},$$

a qual, por condições de regularidade, pode ser calculada alternativamente como

$$K = \mathbb{E}\{j(\vartheta)\}.$$

Lembremos que a matriz de informação esperada é importante para resultados assintóticos. Por exemplo, o inverso desta matriz é a variância da distribuição limite do estimador de máxima verossimilhança (Wilks, 1962; Gnedenko & Kolmogorov, 1962; Rao, 1973; Sen & Singer, 1993).

Durante este trabalho assumiremos que o vetor de parâmetros completo ϑ pode ser decomposto como $\vartheta = (\psi, \zeta)$, onde $\psi = (\psi^1, \dots, \psi^p)^\top$ representa o vetor de parâmetros de interesse e $\zeta = (\zeta^1, \dots, \zeta^q)^\top$ o vetor de parâmetros de perturbação, sendo estes vetores de dimensões p e q , respectivamente.

Utilizando a partição (ψ, ζ) do vetor de parâmetros em um determinado problema, podemos definir a estatística do teste da razão de verossimilhanças $\mathcal{W}(\psi)$ como

$$\mathcal{W}(\psi) = 2 \left\{ L(\widehat{\psi}, \widehat{\zeta}) - L(\psi, \widehat{\zeta}(\psi)) \right\}, \quad (1.1)$$

onde ψ é o vetor de parâmetros de interesse e $\widehat{\psi}$ e $\widehat{\zeta}(\psi)$ são os estimadores de máxima verossimilhança dos vetores ψ e ζ , respectivamente, este último considerando ψ fixado. Sob condições de regularidade, $\mathcal{W}(\psi)$ tem distribuição χ^2 com número de graus de liberdade igual à dimensão do vetor ψ .

O ponto $(\psi, \widehat{\zeta}(\psi))$ será denotado, eventualmente, como $\widetilde{\vartheta}(\psi)$ e durante todo o trabalho assumiremos que

$$\widetilde{\vartheta}(\psi) - \vartheta = O_P(n^{-1/2}). \quad (1.2)$$

Esta é uma suposição condizente com a literatura (Cox & Reid, 1987; DiCiccio & Stern, 1994; Severini, 1998a, entre outros) e uma demonstração pode ser encontrada em Lawley (1956), a partir da expansão em série de Taylor da equação $U(\widehat{\vartheta}) = 0$.

Igualmente importante, assumiremos durante todo o trabalho que

$$E\{O_P(n^{-1/2})\} = O(n^{-1}),$$

sendo esta também uma suposição condizente com a literatura. Embora não exista ainda uma demonstração desta relação, existem alguns resultados neste sentido que podem ser encontrados, por exemplo, em Sen & Singer (1993).

1.1.1 A função de verossimilhança perfilada

Em um determinado modelo estatístico podemos estar interessados somente no vetor ψ e não no vetor de parâmetros completo ϑ . Em situações como estas é possível, por diferentes metodologias, construir uma função que dependa somente de ψ e que possamos utilizar para realizar inferências acerca de ψ . Estas funções são conhecidas como funções de pseudo-verossimilhança.

Diversas destas funções têm sido consideradas na literatura e muitos esforços dedicados a uma delas, a função de verossimilhança perfilada. Devemos observar que esta função somente tem sentido quando parte do vetor de parâmetros que define o modelo estatístico em estudo é considerada como de perturbação.

Definição 1.1.2 *Define-se o logaritmo da função de verossimilhança perfilada para ψ como*

$$L_P(\psi) = \max_{\zeta} L(\psi, \zeta),$$

sendo que o máximo é obtido em todo Ω fixando um valor de ψ . ■

Observemos que o processo de maximização ao qual faz referência a definição anterior é realizado quando obtemos $\widehat{\zeta}(\psi)$. Desta forma a função de verossimilhança perfilada pode ser definida como

$$L_P(\psi) = L(\psi, \widehat{\zeta}(\psi)).$$

Provavelmente a obtenção por William S. Gosset em 1908 (recomendamos a leitura de Peters (1987)) da distribuição amostral da média, em um modelo normal, quando a variância é desconhecida, que ficou conhecida como distribuição t -Student, seja um dos primeiros estudos da função de verossimilhança perfilada. Outros estudos famosos são, por exemplo, a obtenção, por Ronald A. Fisher em 1915, da distribuição do coeficiente de correlação amostral em um modelo normal bivariado, a obtenção por John Wishart em 1928 da distribuição que detém seu nome, o teste de homogeneidade de variâncias de Bartlett (Bartlett, 1937) e algumas soluções para o problema de Behrens-Fisher (Anderson, 1958; Rao, 1973).

Todos estes trabalhos tinham por objetivo encontrar a função de densidade perfilada em modelos particulares. Recentemente, Fraser (1989), McCullagh & Tibshirani (1990), DiCiccio & Stern (1994), Barndorff-Nielsen (1994), DiCiccio *et al.* (1996), Stern (1997), Sartori, Bellio & Salvan (1999) e outros autores estudaram a função de verossimilhança perfilada de diversas maneiras. O objetivo agora é considerar um modelo geral e, através de modificações ou ajustes na função de verossimilhança perfilada, conseguir melhorias nos estimadores e testes.

Passaremos a discutir algumas propriedades de funções de verossimilhança perfiladas. Neste sentido, primeiramente provaremos que os máximos das funções $L_P(\psi)$ e $L(\vartheta)$ coincidem. Suponhamos que $\hat{\psi}_P$ maximiza $L_P(\psi)$. Temos então

$$L_P(\hat{\psi}_P) \geq L_P(\psi) \geq L(\psi, \zeta),$$

e dado que $\hat{\vartheta} = (\hat{\psi}, \hat{\zeta})$ é tal que

$$L(\hat{\psi}, \hat{\zeta}) = \max_{\psi, \zeta} L(\psi, \zeta),$$

então

$$L_P(\hat{\psi}_P) \geq L(\hat{\psi}, \hat{\zeta}).$$

Por outro lado, como $\hat{\vartheta}$ é máximo absoluto de $L(\vartheta)$ no espaço paramétrico,

$$L(\hat{\psi}, \hat{\zeta}) \geq L_P(\hat{\psi}_P),$$

já que $\hat{\psi}_P$ é máximo em um subespaço de Ω . Desta forma obtemos que os pontos $L_P(\hat{\psi}_P)$ e $L(\hat{\psi}, \hat{\zeta})$ coincidem.

Esta propriedade nos permite afirmar que a estatística da razão de verossimilhanças, escrita a partir da função de verossimilhança perfilada, é a própria estatística definida em (1.1), já que

$$\begin{aligned}\mathcal{W}(\psi) &= 2 \left\{ L_P(\widehat{\psi}_P) - L_P(\psi) \right\} \\ &= 2 \left\{ L(\widehat{\psi}, \widehat{\zeta}) - L(\psi, \widehat{\zeta}(\psi)) \right\}.\end{aligned}$$

Provemos agora que os pontos de máximo de $L_P(\psi)$ e $L(\vartheta)$ coincidem, ou seja, provemos que $\widehat{\psi}_P = \widehat{\psi}$, onde $\widehat{\psi}$ e $\widehat{\zeta}$ são soluções das equações de máxima verossimilhança $\partial L / \partial \psi|_{\widehat{\psi}, \widehat{\zeta}} = 0$ e $\partial L / \partial \zeta|_{\widehat{\psi}, \widehat{\zeta}} = 0$, e que, para todo ψ fixado, $\widehat{\zeta}(\psi)$ é solução de $\partial L / \partial \zeta|_{\widehat{\zeta}(\psi)} = 0$ e $\widehat{\psi}_P$, o estimador de máxima verossimilhança perfilada de ψ , é solução de $\partial L_P / \partial \psi|_{\widehat{\psi}_P} = 0$.

Expandindo a função escore perfilada em série de Taylor no ponto (ψ, ζ) , temos

$$\begin{aligned}\frac{\partial L_P}{\partial \psi^a} &= \frac{\partial L}{\partial \psi^a} + \sum_{i=1}^q \frac{\partial^2 L}{\partial \psi^a \partial \zeta^i} (\widehat{\zeta}^i(\psi) - \zeta^i) + \\ &\quad + \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^q \frac{\partial^3 L}{\partial \psi^a \partial \zeta^i \partial \zeta^j} (\widehat{\zeta}^i(\psi) - \zeta^i) (\widehat{\zeta}^j(\psi) - \zeta^j) + O_P(\|\widehat{\zeta}(\psi) - \zeta\|^2),\end{aligned}$$

onde ψ^a , representa a a -ésima componente do vetor de interesse ψ , ζ^i representa a i -ésima componente do vetor de perturbação ζ e $\|\cdot\|$ representa uma norma. Avaliando esta expansão em $(\psi, \zeta) = (\widehat{\psi}, \widehat{\zeta}(\widehat{\psi}))$ obtemos

$$\frac{\partial L_P(\widehat{\psi})}{\partial \psi} = 0,$$

já que $\partial L(\widehat{\psi}) / \partial \psi = 0$ e os outros termos da expansão também se anulam, isto devido ao fato de serem potências de $\widehat{\zeta}(\psi) - \zeta$, e esta diferença se anula quando avaliada em $(\widehat{\psi}, \widehat{\zeta}(\widehat{\psi}))$. Desta forma, provamos que $\widehat{\psi}$ pode ser obtido diretamente como solução da equação de máxima verossimilhança perfilada onde o vetor ζ foi eliminado.

A matriz de informação observada perfilada de ψ , $j_P(\psi)$, é definida de forma análoga à matriz de informação observada $j(\vartheta)$ do vetor ϑ , ou seja,

$$j_P(\psi) = -\frac{\partial^2 L_P(\psi)}{\partial \psi \partial \psi^T}.$$

Supondo a matriz $j(\vartheta)$ particionada segundo a partição (ψ, ζ) , temos

$$j(\psi, \zeta) = \begin{pmatrix} j_{\psi\psi} & j_{\psi\zeta} \\ j_{\psi\zeta} & j_{\zeta\zeta} \end{pmatrix},$$

onde $j_{\psi\psi} = -\partial L(\psi, \zeta)/\partial\psi\partial\psi^\top$, $j_{\psi\zeta} = -\partial L(\psi, \zeta)/\partial\psi\partial\zeta^\top$ e $j_{\zeta\zeta} = -\partial L(\psi, \zeta)/\partial\zeta\partial\zeta^\top$.

Vamos mostrar que $j_P(\widehat{\psi}) = j_{\psi\psi}$, esta avaliada em $(\widehat{\psi}, \widehat{\zeta})$ e, por resultados assintóticos, isto nos permitirá calcular a matriz de covariância assintótica estimada de $\widehat{\psi}$ como $j_P(\widehat{\psi})^{-1}$. Salientemos a vantagem operacional deste cálculo, pois invertemos uma matriz de ordem q , enquanto o seu cálculo através da matriz de informação $j(\vartheta)$ envolve a inversão de uma matriz de ordem $p + q$. Esta vantagem é apreciável quando o modelo apresenta muitos parâmetros de perturbação.

A partir da expansão em série de Taylor de $\partial^2 L_P(\psi)/\partial\psi\partial\psi^\top$ no ponto (ψ, ζ) , podemos ver que

$$\begin{aligned} \frac{\partial^2 L_P}{\partial\psi^a\partial\psi^b} &= \frac{\partial^2 L}{\partial\psi^a\partial\psi^b} + \sum_{i=1}^q \frac{\partial^3 L}{\partial\psi^a\partial\psi^b\partial\zeta^i} (\widehat{\zeta}^i(\psi) - \zeta^i) + \\ &+ \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^q \frac{\partial^4 L}{\partial\psi^a\partial\psi^b\partial\zeta^i\partial\zeta^j} (\widehat{\zeta}^i(\psi) - \zeta^i)(\widehat{\zeta}^j(\psi) - \zeta^j) + O_P(\|\widehat{\zeta}(\psi) - \zeta\|^2), \end{aligned}$$

avaliando em $(\psi, \zeta) = (\widehat{\psi}, \widehat{\zeta}(\widehat{\psi}))$ obtemos $\partial^2 L_P/\partial\psi^a\partial\psi^b = \partial^2 L/\partial\psi^a\partial\psi^b$ e, portanto,

$$\frac{\partial^2 L_P}{\partial\psi\partial\psi^\top} = \frac{\partial^2 L}{\partial\psi\partial\psi^\top},$$

como se queria demonstrar.

Estas propriedades constituem vantagens da utilização da função $L_P(\psi)$, a qual poderá ser utilizada como qualquer outra verossimilhança, desfrutando por isso de uma irrestrita aplicabilidade, fundamentalmente em problemas com parâmetros de perturbação. Infelizmente ela não usufrui de todas as propriedades de uma verossimilhança genuína. É por esta razão que se faz necessário ajustá-la de alguma forma.

A função escore perfilada

$$U_P(\psi) = \frac{\partial}{\partial\psi} L_P(\psi)$$

é, em geral, viciada, ou seja

$$E\{U_P(\psi)\} \neq 0;$$

ver McCullagh & Tibshirani (1990). Pode ocorrer também um vício na informação, ou seja,

$$E\{U_P(\psi)U_P(\psi)^\top\} \neq -E\{\partial U_P(\psi)/\partial\psi^\top\}.$$

1.1.2 Notação

As derivadas do logaritmo da função de verossimilhança com relação aos componentes do vetor ψ serão denotadas por

$$L_a = \frac{\partial L}{\partial\psi^a}, \quad L_{ab} = \frac{\partial^2 L}{\partial\psi^a\partial\psi^b}, \quad \dots,$$

e com relação aos componentes do vetor ζ por

$$L_i = \frac{\partial L}{\partial\zeta^i}, \quad L_{ij} = \frac{\partial^2 L}{\partial\zeta^i\partial\zeta^j}, \quad \dots,$$

seguindo a notação estabelecida em Stern (1997). Além disso,

$$L_{ai} = \frac{\partial^2 L}{\partial\psi^a\partial\zeta^i}$$

representa a derivada de L com relação ao a -ésimo componente de ψ e com relação ao i -ésimo componente de ζ .

Os subíndices a, b, c, \dots assumirão valores em $1, \dots, p$, ou seja, nos componentes do vetor de parâmetros de interesse ψ . Os subíndices i, j, k, \dots assumirão valores $p+1, \dots, p+q$, isto é, nos componentes do vetor de parâmetros de perturbação ζ e os subíndices r, s, t, \dots assumirão valores em $1, \dots, p+q$ abrangendo os componentes de todo o vetor de parâmetros ϑ .

Desenvolveremos a seguir nossa notação para os cumulantes de derivadas do logaritmo da função de verossimilhança, visando unificar aquelas adotadas nos diferentes trabalhos aos quais faremos referência posteriormente.

Os cumulantes serão denotados pela letra grega λ e são definidos como

$$\lambda_r = E\{L_r\}, \quad \lambda_{rs} = E\{L_{rs}\}, \quad \lambda_{r,s} = E\{L_r L_s\}, \quad \lambda_{rst} = E\{L_{rst}\}, \quad \dots,$$

onde $L_r = \partial L/\partial\vartheta^r$, $L_{rs} = \partial L/\partial\vartheta^r\partial\vartheta^s$, etc. e assumiremos que $\lambda_r, \lambda_{rs}, \lambda_{rst}$, etc. são de ordem $O(n)$, que é uma suposição condizente com a literatura (DiCiccio & Stern, 1994; DiCiccio

et al., 1996; Stern, 1997). Derivadas de cumulantes serão amplamente utilizadas e definidas como

$$\lambda_{rs/t} = \partial\lambda_{rs}/\partial\vartheta^t, \quad \lambda_{rst/u} = \partial\lambda_{rst}/\partial\vartheta^u, \quad \lambda_{rs/tu} = \partial^2\lambda_{rs}/\partial\vartheta^t\partial\vartheta^u, \quad \dots,$$

as quais também assumiremos que sejam de ordem $O(n)$.

Estamos agora em condições de definir a matriz de cumulantes $\lambda = (\lambda_{rs})$, a inversa da matriz de cumulantes $\lambda^{-1} = (\lambda^{rs})$ e de observar que a matriz de informação de Fisher pode ser escrita como $K = (\lambda_{r,s})$, de inversa $K^{-1} = (\lambda^{r,s})$.

Utilizaremos repetidas vezes a matriz de cumulantes $(\sigma_{ab}) = (\lambda^{ab})^{-1}$ e algumas relações entre cumulantes, as quais são conhecidas como *identidades de Bartlett* (McCullagh, 1987). Dentre estas, merecem especial atenção

$$\lambda^{rst} = \sum_{u,v,w=1}^{p+q} \lambda^{r,u} \lambda^{s,v} \lambda^{t,w} \lambda_{uvw}, \quad (1.3)$$

dada em McCullagh & Tibshirani (1990), e

$$\lambda_{rs,t} = -\lambda_{rst} + \lambda_{rs/t},$$

em DiCiccio & Stern (1994). Também utilizaremos as funções de cumulantes

$$\tau^{rs} = \sum_{a,b=1}^p \lambda^{ra} \lambda^{sb} \sigma_{ab} \quad (1.4)$$

e

$$\nu^{st} = \lambda^{rs} - \tau^{rs}. \quad (1.5)$$

Entenderemos que quando aparecem índices repetidos como subscritos e sobrescritos em uma expressão, como nas que definem as funções de cumulantes λ^{rst} e τ^{rs} , significa que estamos somando nestes índices, e isto nos permitirá escrever de maneira mais simples estes cumulantes. Assim, a expressão em (1.3) pode ser escrita como

$$\lambda^{rst} = \lambda^{r,u} \lambda^{s,v} \lambda^{t,w} \lambda_{uvw}$$

e a expressão em (1.4), como

$$\tau^{rs} = \lambda^{ra} \lambda^{sb} \sigma_{ab}.$$

No Apêndice A mostramos que a matriz ν , cujos elementos definimos em (1.5), assume a forma

$$\nu = \begin{pmatrix} 0 & 0 \\ 0 & (\lambda_{ij})^{-1} \end{pmatrix}.$$

Serão de ampla utilização durante o trabalho as variáveis aleatórias

$$l_r = L_r, \quad l_{rs} = L_{rs} - \lambda_{rs}, \quad l_{rst} = L_{rst} - \lambda_{rst}, \quad \dots,$$

as quais assumiremos que sejam de esperança zero e ordem $O_P(n^{1/2})$. Por exemplo, para demonstrar (1.2) utilizamos a expansão

$$\widehat{\vartheta}^r - \vartheta^r = -\lambda^{rs} l_s + \lambda^{rs} \lambda^{tu} l_{st} l_u - \frac{1}{2} \lambda^{rs} \lambda^{tv} \lambda^{uw} \lambda_{stu} l_v l_w + O_P(n^{-3/2}),$$

dada por Lawley (1956). Utilizando a suposição acerca da ordem das variáveis aleatórias, cumulantes e funções de cumulantes na expressão acima, obtemos (1.2).

1.2 Função de verossimilhança perfilada ajustada

Quando discursamos acerca da verossimilhança perfilada colocamos ao leitor as vantagens e desvantagens da utilização desta função; diversos trabalhos mencionados tiveram por objetivo introduzir melhorias, em determinado sentido. Uma destas melhorias foi apresentada por Stern (1997).

As propostas de melhorias na verossimilhança perfilada apresentadas antes do trabalho de Stern (1997) foram desenvolvidas em modelos específicos, de maneira que são difíceis ou impossíveis de aplicar quando o vetor de parâmetros de interesse é multidimensional.

Neste sentido, a proposta de Stern (1997) vem resolver esta dificuldade. O objetivo então é construir uma função de verossimilhança perfilada ajustada, de ampla aplicabilidade, mesmo quando o vetor de interesse for multidimensional, cuja definição é apresentada a seguir.

Definição 1.2.1 *O logaritmo da função de verossimilhança perfilada ajustada é definido como*

$$\bar{L}_P(\psi) = L_P(\psi) + \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi), \quad (1.6)$$

onde $\Lambda(\vartheta) = A^a L_a$, sendo A^a uma função de ϑ de ordem $O(n^{-1})$, $\Theta(\vartheta) = \frac{1}{2} B^{ab} L_a L_b$, com B^{ab} uma função de ϑ de ordem $O(n^{-2})$, $\tilde{\Lambda}(\psi) = \Lambda(\tilde{\vartheta}(\psi)) = \Lambda(\psi, \hat{\zeta}(\psi))$ e $\tilde{\Theta}(\psi) = \Theta(\tilde{\vartheta}(\psi)) = \Theta(\psi, \hat{\zeta}(\psi))$. Faremos referência a $\bar{L}_P(\psi)$ simplesmente como função de verossimilhança perfilada ajustada. ■

Ressaltamos que o objetivo das funções de ajuste $\Lambda(\vartheta)$ e $\Theta(\vartheta)$ é corrigir os vícios da esperança e da informação, ou seja, a função de verossimilhança perfilada ajustada definida em (1.6) deve satisfazer

$$E\{U_P(\psi)\} = O(n^{-1})$$

e

$$E\{U_P(\psi)U_P(\psi)^\top\} + E\{\partial U_P(\psi)/\partial\psi^\top\} = O(n^{-1}).$$

Desta forma, se estes vícios não são eliminados, pelo menos são reduzidos até uma ordem aceitável, sendo nulos assintoticamente.

Segundo é mostrado em Stern (1997), as expressões que correspondem às funções A^a e B^{ab} são

$$A^a = \lambda^{ar} \nu^{st} \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right)$$

e

$$\begin{aligned} B^{ab} = & \lambda^{ar} \lambda^{bs} \lambda^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rts} \lambda_{uvw} - \lambda_{rst} \lambda_{uv/w} + \lambda_{rt/s} \lambda_{uv/w} + \lambda_{st/r} \lambda_{uv/w} \right. \\ & \left. - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} - \frac{1}{2} \lambda_{st/r} \lambda_{uvw} \right) \\ & + \lambda^{ar} \lambda^{bs} \lambda^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} + \lambda_{rt/u} \lambda_{sw/v} \right. \\ & \left. + \lambda_{rt/u} \lambda_{vw/s} + \lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{rtu} \lambda_{vw/s} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) \\ & - \lambda^{ar} \lambda^{bs} \lambda^{tu} \left(\frac{1}{2} \lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su} + \lambda_{st/ru} - \frac{1}{2} \lambda_{rtu/s} - \frac{1}{2} \lambda_{stu/r} \right). \end{aligned} \quad (1.7)$$

Contudo, mostraremos que este último resultado contém erro.

A idéia original desta tese era aplicar os resultados publicados em Stern (1997) na classe de modelos lineares generalizados. Porém, quando começamos a estudar tais resultados, percebemos que não foi possível reproduzir o Exemplo 1 (continuação), da pg. 663, do referido artigo. Em comunicação com o autor, ele nos afirmou que realmente algumas das expressões nesse exemplo continham erro e que concordava com os nossos resultados.

A situação em conflito é o modelo de regressão normal linear múltiplo com σ , o desvio padrão, sendo o parâmetro de interesse e o vetor β , dos parâmetros de regressão, os de perturbação, ou seja, $\psi = \sigma$ e $\zeta = (\beta^1, \dots, \beta^q)$. Assim, assumiremos que $Y = X\beta + \epsilon$, onde $\epsilon \sim N(0, \sigma^2)$, X é uma matriz $n \times q$ de covariadas e β é o vetor de parâmetros de regressão de dimensão q .

Neste modelo

$$L_P(\sigma) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{n\hat{\sigma}^2}{2\sigma^2},$$

onde

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2$$

é o estimador de máxima verossimilhança para σ^2 , $\hat{\mu}_i$ é o i -ésimo componente do vetor $\hat{\mu} = X\hat{\beta}$ e $\hat{\beta} = (X^T X)^{-1} X^T Y$ é o estimador de máxima verossimilhança do vetor de perturbação.

A função de verossimilhança perfilada ajustada, segundo Stern (1997, Exemplo 1, pg. 663), é da forma

$$\bar{L}_P(\sigma) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{n\hat{\sigma}^2}{2\sigma^2} + \frac{q}{2} \left(1 - \frac{\hat{\sigma}}{\sigma}\right) + \frac{q}{4} \left(1 - \frac{\hat{\sigma}}{\sigma}\right)^2$$

e, portanto, a função escore perfilada ajustada $\bar{U}_P(\sigma) = d\bar{L}_P(\sigma)/d\sigma$ será

$$\bar{U}_P(\sigma) = -\frac{n}{\sigma} + \frac{q\hat{\sigma}}{\sigma^2} + \frac{n\hat{\sigma}^2}{\sigma^3} \left(1 - \frac{q}{2n}\right).$$

Sabemos que $n\hat{\sigma}^2/\sigma^2 \sim \chi^2(n - q)$ e, assim, $\sqrt{n}\hat{\sigma}/\sigma \sim \chi(n - q)$, esta última chamada de distribuição qui com $n - q$ graus de liberdade, que satisfaz (Stuart & Ord, 1987)

$$E\{\hat{\sigma}\} = \frac{\sigma\mu'}{\sqrt{n}},$$

onde

$$\mu' = \sqrt{2} \frac{\Gamma(\frac{1}{2}(n - q + 1))}{\Gamma(\frac{1}{2}(n - q))}$$

e

$$E\{\hat{\sigma}^2\} = \left(1 - \frac{q}{n}\right) \sigma^2,$$

sendo $\Gamma(\cdot)$ a função gama¹.

Destes resultados obtemos

$$E\{\bar{U}_P(\sigma)\} = \frac{q\mu'}{\sqrt{n}\sigma} - \frac{3q}{2\sigma} + \frac{q^2}{2n\sigma}.$$

Por outro lado, pela aproximação de Stirling

$$\frac{\mu'}{\sqrt{n}} = \sqrt{1 - \frac{q}{n}} - \frac{1}{4\sqrt{n(n-q)}} + \dots$$

e expandindo o primeiro termo em série de Taylor obtemos

$$\sqrt{1 - \frac{q}{n}} = 1 - \frac{q}{2n} + \frac{q^2}{8n^2} - \frac{q^3}{16n^3} + \dots$$

Substituindo acima, chegamos a

$$\frac{\mu'}{\sqrt{n}} = 1 + O(n^{-1}).$$

Desta forma, obtemos

$$E\{\bar{U}_P(\sigma)\} = -\frac{q}{2\sigma} + O(n^{-1}),$$

e, portanto, o ajuste não corrige o vício da função escore e $E\{\bar{U}_P(\sigma)\} = O(1)$. Concluimos então que há necessidade de corrigir o erro em (1.7).

No capítulo seguinte nos dedicaremos a obter a expressão correta para a função B^{ab} e nos desenvolvimentos seguintes provaremos de diversas maneiras que nosso resultado está correto. Além disso, aplicaremos a função de verossimilhança perfilada ajustada nos modelos lineares generalizados.

¹ $\Gamma(\cdot)$ é a função gama, definida como $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, para $\alpha > 0$.

1.3 Modelos lineares generalizados

1.3.1 Definição e algumas propriedades

Os modelos lineares generalizados (Nelder & Wedderburn, 1972) desempenham um papel importante na Estatística uma vez que generalizam o modelo tradicional de regressão normal linear, permitindo muitas opções para a distribuição da variável resposta e dando maior flexibilidade para a ligação entre a média e a parte sistemática do modelo, estendendo a estrutura conhecida para a regressão normal linear. A grande vantagem é a possibilidade do estudo conjunto das propriedades de diferentes modelos de regressão. Existe uma vasta literatura no assunto, mas a referência mais completa é McCullagh & Nelder (1989).

Suponhamos que Y_1, \dots, Y_n sejam variáveis aleatórias independentes, cada uma com densidade na forma

$$\pi(y_i; \theta^i, \phi) = \exp[\phi\{y_i\theta^i - b(\theta^i) + c(y_i)\} + a(y_i, \phi)], \quad (1.8)$$

com $E\{Y_i\} = \mu^i$, $\text{var}\{Y_i\} = \phi^{-1}V_i$, $V_i = d\mu^i/d\theta^i$ é a função de variância, θ é o parâmetro canônico e ϕ é o parâmetro de dispersão ($\phi > 0$) e $i = 1, 2, \dots, n$.

Os modelos lineares generalizados são definidos por (1.8) e pelo componente sistemático

$$g(\mu^i) = \eta^i, \quad (1.9)$$

onde $\eta^i = X_i\beta$ é o preditor linear, $\beta = (\beta^1, \dots, \beta^m)^\top$, $m < n$, é um vetor de parâmetros a serem estimados, $X_i = (x_{i1}, \dots, x_{im})$ representa os valores de m variáveis explicativas e $g(\cdot)$ é uma função monótona e duplamente diferenciável, denominada função de ligação. As distribuições normal, binomial, Poisson, gama, binomial negativa e normal inversa são algumas das distribuições conhecidas pertencentes a esta família de modelos. Consideraremos $a(\cdot, \cdot)$, $b(\cdot)$ e $c(\cdot)$ em (1.8) como funções conhecidas.

Nos modelos binomial, binomial negativo e Poisson ϕ assume o valor 1, o que implica que $a(y, \phi) = a(y)$ e em outras situações temos que $a(y, \phi) = d_1(\phi) + d_2(y)$. Na distribuição normal com variância ϕ^{-1} , $d_1(\phi) = \log(\phi/2\pi)/2$ e $d_2(y) = 0$, na distribuição gama $d_1(\phi) = \phi \log \phi -$

$\log \Gamma(\phi)$ e $d_2(y) = -\log y$, se a distribuição é a normal inversa com $\phi = E\{Y\}^3/\text{var}\{Y\}$, $d_1(\phi) = \log \sqrt{\phi}$ e $d_2(y) = -\log(2\pi y^3)/2$.

Dado que $\mu = b'(\theta)$ e pela relação (1.9) podemos considerar $\theta = \theta(\eta)$ e $\eta = \eta(\beta)$. Nestas condições, assumindo que $a(y, \phi) = d_1(\phi) + d_2(y)$ o logaritmo da função de verossimilhança é da forma $L(\phi, \beta) = \phi \sum_{i=1}^n \{y_i \theta^i - b(\theta^i) + c(y_i)\} + nd_1(\phi) + \sum_{i=1}^n d_2(y_i)$.

Uma situação particular importante ocorre quando o parâmetro canônico coincide com o preditor linear, ou seja, quando $\theta = \eta$, o que implica que

$$L(\phi, \beta) = \phi \sum_{l=1}^n \left\{ y_l \sum_{i=1}^m x_{li} \beta^i - b \left(\sum_{i=1}^m x_{li} \beta^i \right) + c(y_l) \right\} + nd_1(\phi) + \sum_{l=1}^n d_2(y_l).$$

As ligações que correspondem a esta verossimilhança são chamadas de ligações canônicas e aparecem na Tabela 1.1. Estas ligações garantem que L é uma função côncava em β e, nesta situação, Wedderburn (1976) provou que o estimador de máxima verossimilhança $\hat{\beta}$ existe e é único. Conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente.

Tabela 1.1: Ligações canônicas de diferentes modelos e segunda derivada da função $d_1(\phi)$.

modelo	ligação canônica	$d_1''(\phi)$
Poisson	$\log \mu = \eta$	-
Binomial	$\log \mu - \log(1 - \mu) = \eta$	-
Binomial negativa	$\log \mu - \log(1 - \mu) = \eta$	-
Normal	$\mu = \eta$	$-\frac{1}{2\phi^2}$
Gama	$\mu^{-1} = \eta$	$\frac{1}{\phi} - \frac{d^2}{d\phi^2} \log \Gamma(\phi)$
Normal inversa	$\mu^{-2} = \eta$	$-\frac{1}{2\phi^2}$

Em relação ao parâmetro de dispersão, vemos que

$$\frac{\partial L(\beta, \phi)}{\partial \phi} = \sum_{i=1}^n \{y_i \theta^i - b(\theta^i) + c(y_i)\} + nd_1'(\phi) \quad (1.10)$$

e

$$\frac{\partial^2 L(\beta, \phi)}{\partial \phi^2} = nd_1''(\phi) \quad (1.11)$$

independentemente da função de ligação assumida.

Por propriedade de derivadas superiores sabemos que L será côncava em ϕ se a segunda derivada for negativa (Spivak, 1970). Podemos notar de (1.11) que o sinal da segunda derivada de L com respeito a ϕ coincide com o de $d_1''(\phi)$. Na Tabela 1.1 também apresentamos as expressões de $d_1''(\phi)$ para os modelos mais utilizados. Na regressão normal e normal inversa não há dúvidas de que a segunda derivada da função d_1 é negativa no intervalo $(0, \infty)$, já o modelo gama deve ser estudado com maiores detalhes.

Para analisar o modelo gama observemos a Figura 1.1. Nela mostramos o comportamento da segunda derivada da função $d_1(\phi)$. Observamos que esta função é negativa, fato que provaremos a seguir, podendo concluir então que a função de verossimilhança é côncava.

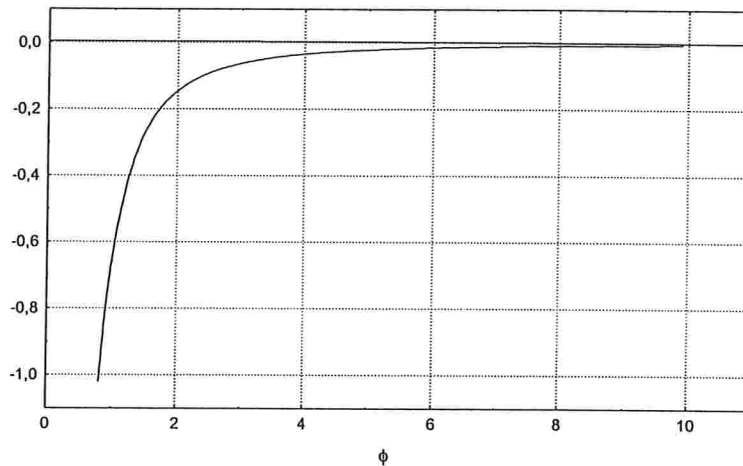


Figura 1.1: Gráfico da função $d_1''(\phi) = 1/\phi - d^2 \log \Gamma(\phi)/d\phi^2$ no modelo gama.

Provaremos agora que a função $d_1''(\phi)$ é negativa. Neste sentido, observemos (Whittaker & Watson, 1952, pg. 250) que

$$\frac{d^2}{d\phi^2} \log \Gamma(\phi) = \frac{1}{2\phi^2} + \frac{1}{\phi} + \int_0^\infty \frac{4\phi t}{(\phi^2 + t^2)^2} \frac{1}{e^{2\pi t} - 1} dt,$$

e, portanto,

$$d_1''(\phi) = -\frac{1}{2\phi^2} - \int_0^\infty \frac{4\phi t}{(\phi^2 + t^2)^2} \frac{1}{e^{2\pi t} - 1} dt.$$

A função sob o sinal de integração é estritamente positiva no intervalo $t \in (0, \infty)$, de onde concluímos que $d_1''(\phi)$, no modelo gama, é estritamente negativa.

Desta forma vemos que, em todas as situações consideradas, a utilização da ligação canônica nos garante bons resultados. Embora os resultados que serão apresentados a seguir e nos próximos capítulos sejam gerais, no sentido de serem obtidos para qualquer função de ligação, os exemplos demonstrativos do comportamento da função de verossimilhança perfilada ajustada tratam do caso de ligação canônica.

1.3.2 Estimação e teste da razão de verossimilhanças

Encontremos a seguir a função escore para o vetor de parâmetros de regressão β , ou seja, encontremos $U(\beta)$. Neste sentido, observemos que

$$\frac{\partial L}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^n \phi [y_i \theta^i - b(\theta^i) + c(y_i)] + \frac{\partial}{\partial \beta} \sum_{i=1}^n a(y_i, \phi) = \phi \sum_{i=1}^n \left[y_i \frac{d\theta^i}{d\beta} - \frac{db(\theta^i)}{d\beta} \right],$$

na qual

$$\frac{d\theta^i}{d\beta} = \frac{d\theta^i}{d\mu^i} \frac{d\mu^i}{d\eta^i} \frac{d\eta^i}{d\beta} = \frac{1}{V_i} \frac{d\mu^i}{d\eta^i} X_i^\top$$

e

$$\frac{db(\theta^i)}{d\beta} = \frac{db(\theta^i)}{d\theta^i} \frac{d\theta^i}{d\beta} = \mu^i \frac{1}{V_i} \frac{d\mu^i}{d\eta^i} X_i^\top.$$

Então, podemos escrever a função escore para β como

$$U(\beta) = \phi X T V^{-1} (y - \mu),$$

onde $X = (X_1^\top, \dots, X_n^\top)$, $T = \text{diag}\{T_1, \dots, T_n\}$, com $T_i = d\mu^i/d\eta^i$, $V = \text{diag}\{V_1, \dots, V_n\}$, $y = (y_1, \dots, y_n)^\top$ e $\mu = (\mu^1, \dots, \mu^n)^\top$.

Derivando mais uma vez L com relação a β^\top , obtemos

$$\frac{\partial^2 L}{\partial \beta \partial \beta^\top} = -\phi X T V^{-1} T X^\top = -\phi X W X^\top,$$

onde $W \text{diag}\{w_1, \dots, w_n\}$ com $w_i = (d\mu^i/d\eta^i)^2/V_i$ e, portanto, a matriz de informação observada para o vetor β é da forma

$$j(\beta) = \phi X^\top W X.$$

Quanto ao parâmetro de dispersão ϕ , encontramos em (1.10) a função escore e em (1.11) a informação observada. Também,

$$\frac{\partial^2 L(\beta, \phi)}{\partial \beta \partial \phi} = XTV^{-1}(y - \mu),$$

que tem esperança nula. Então, a matriz de informação esperada ou informação de Fisher assume a forma

$$K = \begin{pmatrix} \phi X^T W X & 0 \\ 0 & -nd_1''(\phi) \end{pmatrix}, \quad (1.12)$$

com inversa

$$K^{-1} = \begin{pmatrix} \frac{1}{\phi}(X^T W X)^{-1} & 0 \\ 0 & \frac{-1}{nd_1''(\phi)} \end{pmatrix}.$$

Expandindo a função escore $U(\beta)$ em torno de um valor inicial, obtemos o processo iterativo de Newton-Raphson como sendo

$$\beta^{(m)} = \beta^{(m-1)} + \{-U'(\beta^{(m)})\}^{-1}U(\beta^{(m)}),$$

onde $U'(\beta)$ denota a primeira derivada de $U(\beta)$ com relação a β . A matriz $-U'(\beta)$ pode não ser positiva definida e em situações como esta a convergência do algoritmo é comprometida. Uma solução é substituir a matriz $-U'(\beta)$ pelo correspondente valor esperado, resultando no processo iterativo scoring de Fisher da forma

$$\beta^{(m)} = \beta^{(m-1)} + K^{-1}(\beta^{(m)})U(\beta^{(m)}).$$

Sob condições de regularidade (Fahrmeir & Kaufmann, 1985; Sen & Singer, 1993), tem-se que $\hat{\beta}$ é um estimador consistente de β e que $\sqrt{n}(\hat{\beta} - \beta)$ converge em distribuição para a $N(0, \Sigma(\beta)^{-1})$, quando $n \rightarrow \infty$, onde $\Sigma(\beta) = \lim_{n \rightarrow \infty} K(\beta)/n$, sendo $\Sigma(\beta)$ uma matriz definida positiva.

Quanto ao parâmetro de dispersão, já observamos que β e ϕ são ortogonais, no sentido de que $E\{\partial^2 L(\phi, \beta)/\partial \beta \partial \phi\} = 0$. Uma consequência é a independência assintótica entre $\hat{\beta}$ e $\hat{\phi}$. Mostra-se também que, sob condições de regularidade, $\sqrt{n}(\hat{\phi} - \phi)$ converge em distribuição

para a $N(0, \sigma_\phi^2)$, quando $n \rightarrow \infty$, onde $\sigma_\phi^2 = -1/d_1''(\phi)$. Portanto, um estimador consistente para $\text{var}\{\widehat{\phi}\}$ é dado $-1/nd_1''(\widehat{\phi})$.

Acerca do teste da razão de verossimilhanças para os parâmetros de regressão, podemos considerar a partição $\beta = (\beta_1, \beta_2)$, β_1 de dimensão p e β_2 de dimensão q , e a hipótese nula $H_0 : \beta_1 = \beta_1^0$, a ser testada contra $H_A : \beta_1 \neq \beta_1^0$. Para este caso temos $\mathcal{W}(\beta_1^0) = 2\{L(\widehat{\beta}) - L(\widehat{\beta}^0)\}$, onde $\widehat{\beta}^0$ é a estimativa de máxima verossimilhança do vetor de parâmetros de regressão sob H_0 . Para obtê-la, escrevemos a parte sistemática do modelo de regressão como $\eta = \eta_1^0 + \eta_2$, onde $\eta_1^0 = \sum_{j=1}^p x_j \beta_1^{0j}$ e $\eta_2 = \sum_{j=p+1}^{p+q} x_j \beta_2^j$ se ϕ for conhecido. Caso o parâmetro de dispersão seja estimado, escrevemos $\mathcal{W}(\beta_1^0) = 2\{L(\widehat{\beta}, \widehat{\phi}) - L(\widehat{\beta}^0, \widehat{\phi}^0)\}$, onde $\widehat{\phi}^0$ é a estimativa de máxima verossimilhança de ϕ sob a hipótese nula. Se o objetivo for realizar inferências acerca do parâmetro de dispersão podemos considerar a hipótese $H_0 : \phi = \phi^0$ a ser testada contra $H_A : \phi \neq \phi^0$ e a estatística da razão de verossimilhanças assume a forma $\mathcal{W}(\phi) = 2\{L(\widehat{\beta}, \widehat{\phi}) - L(\widehat{\beta}, \widehat{\phi}^0)\}$.

Capítulo 2

Ajuste para a verossimilhança perfilada

Uma das propostas fundamentais de nosso trabalho é apresentada neste capítulo. Utilizando as idéias de Stern (1997) desenvolvemos um ajuste para a função de verossimilhança perfilada com o objetivo de corrigir o vício da função escore perfilada e o vício da informação. O desenvolvimento desta função de ajuste será realizado em duas etapas, na primeira corrigiremos o vício da função escore perfilada e posteriormente nos dedicaremos a corrigir o vício da informação, sendo este o principal resultado deste capítulo.

Ressaltemos que a necessidade de realizar estas demonstrações foi justificada pelo erro apontado anteriormente no artigo de Stern (1997). Na Seção 2.1.1 nos dedicaremos a corrigir somente o vício da função escore perfilada para posteriormente, na Seção 2.1.2, dedicarmos-nos ao vício da informação.

Mesmo com a melhoria que esta função de ajuste induz na verossimilhança perfilada, percebemos que as estatísticas de teste construídas a partir dela, também necessitam de ajuste. É com este objetivo que desenvolvemos a correção de Bartlett para o teste da razão de verossimilhanças perfilada ajustada. A Seção 2.2 dedica-se a obter este resultado.

2.1 Função de verossimilhança perfilada ajustada

O ajuste para o logaritmo da função de verossimilhança perfilada foi definido na Seção 1.2, porém, segundo observamos, o resultado apresentado no artigo de Stern (1997) contém erro no termo que denotamos por B^{ab} e apresentamos, na forma dada por Stern, em (1.7). Dedicaremos esta e as próximas seções a obter a expressão correta da função B^{ab} .

2.1.1 Ajuste para o vício da esperança da função escore perfilada

Nesta seção demonstraremos que a expressão da função de ajuste $\Lambda(\vartheta)$ corresponde à forma integral do ajuste de McCullagh & Tibshirani (1990) para o vício da esperança da função escore perfilada.

Teorema 2.1.1 *Seja*

$$\Lambda(\vartheta) = A^a(\vartheta)L_a(\vartheta),$$

onde

$$A^a = \lambda^{ar} \nu^{st} \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right).$$

Então, a função de verossimilhança perfilada ajustada, definida como

$$\bar{L}_P(\psi) = L_P(\psi) + \tilde{\Lambda}(\psi),$$

tem vício de função escore perfilada ajustada de ordem $O(n^{-1})$. Aqui, $\tilde{\Lambda}(\psi) = \tilde{A}^a(\psi)\tilde{L}_a(\psi)$.

Antes de demonstrar este resultado observemos que a afirmação deste teorema significa que

$$E\{\bar{U}_P(\psi)\} = O(n^{-1}),$$

onde $\bar{U}_P(\psi) = \partial \bar{L}_P(\psi) / \partial \psi$ é a função escore perfilada ajustada. Observemos também que

$$\tilde{\Lambda}(\psi) = A^a(\tilde{\vartheta}(\psi))L_a(\tilde{\vartheta}(\psi)) = \tilde{A}^a(\psi, \hat{\zeta}(\psi))\tilde{L}_a(\psi, \hat{\zeta}(\psi)),$$

onde $\tilde{\vartheta}(\psi) = (\psi, \hat{\zeta}(\psi))$, e que A^a é uma função de ordem $O(n^{-1})$.

Apresentamos a seguir a demonstração do teorema. McCullagh & Tibshirani (1990) demonstraram que o a -ésimo componente da expansão em série de Taylor da função escore perfilada assume a seguinte forma:

$$U_P^a(\psi) = L_a + L_{ai} \left\{ \lambda_{i,j}^{-1} L_j + \lambda_{i,j}^{-1} \lambda_{k,l}^{-1} (L_{jk} - \lambda_{jk}) L_l + \frac{1}{2} \lambda^{ijk} L_j L_k \right\} + \frac{1}{2} \lambda_{i,k}^{-1} \lambda_{j,l}^{-1} L_{aij} L_k L_l + O_P(n^{-1/2}). \quad (2.1)$$

Utilizaremos relações entre cumulantes e funções destes para escrever de maneira mais simples a aproximação à função escore perfilada acima. Primeiramente, vemos que

$$\nu = \begin{pmatrix} 0 & 0 \\ 0 & \lambda_{ij}^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -\lambda_{i,j}^{-1} \end{pmatrix},$$

o que nos permite escrever

$$U_P^a(\psi) = L_a + L_{as} \left\{ -\nu^{st} L_t + \nu^{st} \nu^{uw} (L_{tu} - \lambda_{tu}) L_w \right\} + \frac{1}{2} \lambda^{ijk} L_{ai} L_j L_k + \frac{1}{2} \nu^{st} \nu^{uw} L_{asu} L_t L_w + O_P(n^{-1/2}),$$

ou, alternativamente,

$$U_P^a(\psi) = L_a - \nu^{st} L_{as} L_t + \nu^{st} \nu^{uw} L_{as} L_{tu} L_w - \nu^{st} \nu^{uw} \lambda_{tu} L_{as} L_w + \frac{1}{2} \lambda^{ijk} L_{ai} L_j L_k + \frac{1}{2} \nu^{st} \nu^{uw} L_{asu} L_t L_w + O_P(n^{-1/2}).$$

McCullagh & Tibshirani (1990) provaram que $\lambda^{ijk} = -\nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw}$. Portanto,

$$U_P^a(\psi) = L_a - \nu^{st} L_{as} L_t + \nu^{st} \nu^{uw} L_{as} L_{tu} L_w - \nu^{st} \nu^{uw} \lambda_{tu} L_{as} L_w - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} L_{ar} L_s L_t + \frac{1}{2} \nu^{st} \nu^{uw} L_{asu} L_t L_w + O_P(n^{-1/2}).$$

Lembrando que $l_r = L_r$, $l_{rs} = L_{rs} - \lambda_{rs}$ e que $l_{rst} = L_{rst} - \lambda_{rst}$, a expressão acima pode ser escrita como

$$U_P^a(\psi) = l_a - \nu^{st} (l_{as} + \lambda_{as}) l_t + \nu^{st} \nu^{uw} (l_{as} + \lambda_{as}) (l_{tu} + \lambda_{tu}) l_w - \nu^{st} \nu^{uw} \lambda_{tu} (l_{as} + \lambda_{as}) l_w - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} (l_{ar} + \lambda_{ar}) l_s l_t + \frac{1}{2} \nu^{st} \nu^{uw} (l_{asu} + \lambda_{asu}) l_t l_w + O_P(n^{-1/2}).$$

Desenvolvendo, vemos que

$$\begin{aligned}
U_P^a(\psi) = & l_a - \nu^{st} l_{as} l_t - \nu^{st} \lambda_{as} l_t + \nu^{st} \nu^{uw} l_{as} l_{tu} l_w + \nu^{st} \nu^{uw} l_{as} \lambda_{tu} l_w + \nu^{st} \nu^{uw} \lambda_{as} l_{tu} l_w \\
& + \nu^{st} \nu^{uw} \lambda_{as} \lambda_{tu} l_w - \nu^{st} \nu^{uw} \lambda_{tu} l_{as} l_w - \nu^{st} \nu^{uw} \lambda_{tu} \lambda_{as} l_w - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} l_{ar} l_s l_t \\
& - \frac{1}{2} \nu^{ru} \nu^{st} \nu^{tw} \lambda_{uvw} \lambda_{ar} l_s l_t + \frac{1}{2} \nu^{st} \nu^{uw} l_{asu} l_t l_w + \frac{1}{2} \nu^{st} \nu^{uw} \lambda_{asu} l_t l_w + O_P(n^{-1/2}).
\end{aligned}$$

Provamos, no Apêndice A, que $\sigma_{ab} \lambda^{br} l_r = l_a - \nu^{st} \lambda_{as} l_t$. Então,

$$\begin{aligned}
U_P^a(\psi) = & \sigma_{ab} \lambda^{br} l_r - \nu^{st} l_{as} l_t + \nu^{st} \nu^{uw} l_{as} l_{tu} l_w + \nu^{st} \nu^{uw} \lambda_{as} l_{tu} l_w - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} l_{ar} l_s l_t \\
& - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} \lambda_{ar} l_s l_t + \frac{1}{2} \nu^{st} \nu^{uw} l_{asu} l_t l_w + \frac{1}{2} \nu^{st} \nu^{uw} \lambda_{asu} l_t l_w + O_P(n^{-1/2}).
\end{aligned}$$

Observando que as variáveis aleatórias $\nu^{st} \nu^{uw} l_{as} l_{tu} l_w$, $\nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} l_{ar} l_s l_t$ e $\nu^{st} \nu^{uw} l_{asu} l_t l_w$ são de ordem $O_P(n^{-1/2})$, obtemos

$$U_P^a(\psi) = \sigma_{ab} \lambda^{br} l_r - \nu^{st} l_{as} l_t + \nu^{st} \nu^{uw} \lambda_{as} l_{tu} l_w - \frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} \lambda_{ar} l_s l_t + \frac{1}{2} \nu^{st} \nu^{uw} \lambda_{asu} l_t l_w + O_P(n^{-1/2}).$$

Também podemos escrever (ver Apêndice A)

$$\begin{aligned}
-\nu^{st} l_{as} l_t + \nu^{st} \nu^{uw} \lambda_{as} l_{tu} l_w &= \nu^{st} l_{as} l_t + \nu^{st} \nu^{uw} \lambda_{au} l_{ws} l_t \\
&= -(\nu^{st} l_{as} l_t - \nu^{st} (\nu^{uw} \lambda_{au}) l_{ws} l_t) \\
&= -\sigma_{ab} \lambda^{br} \nu^{st} l_{rs} l_t
\end{aligned}$$

e

$$-\frac{1}{2} \nu^{ru} \nu^{sv} \nu^{tw} \lambda_{uvw} \lambda_{ar} l_s l_t + \frac{1}{2} \nu^{st} \nu^{uw} \lambda_{asu} l_t l_w = \frac{1}{2} \sigma_{ab} \lambda^{br} \nu^{su} \nu^{tv} \lambda_{rst} l_u l_v.$$

Desta forma obtemos

$$U_P^a(\psi) = \sigma_{ab} \lambda^{br} l_r - \sigma_{ab} \lambda^{br} \nu^{st} l_{rs} l_t + \frac{1}{2} \sigma_{ab} \lambda^{br} \nu^{su} \nu^{tv} \lambda_{rst} l_u l_v + O_P(n^{-1/2}). \quad (2.2)$$

Esta expressão aparece mencionada em DiCiccio & Stern (1994) sem demonstração. O fato de ser mais simples do que (2.1) nos permitirá desenvolver o ajuste de Stern. Tomando esperança em (2.2), obtemos a expressão do vício da função escore. Com esse objetivo utilizaremos a identidade $E\{l_{rs} l_t\} = -\lambda_{rst} + \lambda_{rs}/t$. Lembrando que $E\{l_r\} = 0$ e que $E\{l_r l_s\} = -\lambda_{rs}$, obtemos

$$E\{U_P^a(\psi)\} = \sigma_{ab} \lambda^{br} \nu^{st} \lambda_{rst} - \sigma_{ab} \lambda^{br} \nu^{st} \lambda_{rs}/t - \frac{1}{2} \sigma_{ab} \lambda^{br} \nu^{su} \nu^{tv} \lambda_{rst} \lambda_{uv} + O(n^{-1}).$$

Podemos escrever esta esperança de forma diferente. Utilizando a relação $\nu^{su}\nu^{tv}\lambda_{uv} = \lambda_{su}^{-1}\lambda_{uv}\lambda_{tv}^{-1} = \nu^{st}$, obtemos que $\sigma_{ab}\lambda^{br}\nu^{su}\nu^{tv}\lambda_{rst}\lambda_{uv} = \sigma_{ab}\lambda^{br}\nu^{st}\lambda_{rst}$, chegando assim à expressão do vício da função escore perfilada:

$$E\{U_P^a(\psi)\} = -\sigma_{ab}\lambda^{br}\nu^{st} \left(\lambda_{rs/t} - \frac{1}{2}\lambda_{rst} \right) + O(n^{-1}). \quad (2.3)$$

Por outro lado,

$$\frac{\partial \bar{L}_P(\psi)}{\partial \psi^a} = \frac{\partial L_P(\psi)}{\partial \psi^a} + \frac{\partial \tilde{\Lambda}(\psi)}{\partial \psi^a} = U_P^a(\psi) + \frac{\partial \tilde{\Lambda}(\psi)}{\partial \psi^a};$$

provaremos a seguir que $E\{\partial \tilde{\Lambda}(\psi)/\partial \psi^a\} = -E\{U_P^a(\psi)\} + O(n^{-1})$.

Primeiramente, observemos que

$$\frac{\partial \tilde{\Lambda}(\psi)}{\partial \psi^a} = \frac{\partial \tilde{A}^b}{\partial \psi^a} \tilde{L}_b + \tilde{A}^b \frac{\partial \tilde{L}_b}{\partial \psi^a}.$$

Agora, pela regra da cadeia

$$\frac{\partial \tilde{A}^b}{\partial \psi^a} \tilde{L}_b = \frac{\partial \tilde{A}^b}{\partial \tilde{\vartheta}^r} \frac{\partial \tilde{\vartheta}^r}{\partial \psi^a} \tilde{L}_b$$

e

$$\tilde{A}^b \frac{\partial \tilde{L}_b}{\partial \psi^a} = \tilde{A}^b \frac{\partial \tilde{L}_b}{\partial \tilde{\vartheta}^r} \frac{\partial \tilde{\vartheta}^r}{\partial \psi^a}.$$

A ordem de cada termo pode ser obtida através de expansão em série de Taylor. Temos inicialmente que

$$\tilde{A}^b(\psi) = A^b(\vartheta) + A_r^b(\vartheta)(\tilde{\vartheta}(\psi) - \vartheta)^r + O_P(n^{-3/2}),$$

onde $A_r^b(\vartheta) = \partial A^b(\vartheta)/\partial \vartheta^r$, que é de ordem $O(n^{-1})$. Lembrando que $\tilde{\vartheta}(\psi) - \vartheta = O_P(n^{-1/2})$, temos $\tilde{A}^b = A^b(\vartheta) + O_P(n^{-3/2})$, podendo concluir que $\tilde{A}^b(\psi)$ e $\partial A^b(\vartheta)/\partial \vartheta^a$ são de ordem $O_P(n^{-1})$. Também

$$\tilde{L}_b(\psi) = L_b(\vartheta) + L_{br}(\vartheta)(\tilde{\vartheta}(\psi) - \vartheta)^r + O_P(1),$$

concluindo que $\tilde{L}_b(\psi) = L_b(\vartheta) + O_P(1)$ é de ordem $O_P(n^{1/2})$ e, assim,

$$\frac{\partial \tilde{A}^b}{\partial \psi^a} \tilde{L}_b = A_a^b(\vartheta) L_a(\vartheta) + O_P(n^{-1}),$$

cuja esperança é $O(n^{-1})$, já que $\mathbb{E}\{L_a(\vartheta)\} = 0$. Desta forma, vemos que

$$\mathbb{E} \left\{ \frac{\partial}{\partial \psi^a} \tilde{\Lambda}(\psi) \right\} = \mathbb{E} \left\{ \tilde{A}^b \frac{\partial \tilde{L}_b}{\partial \psi^a} \right\} + O(n^{-1}).$$

Em segundo lugar, $L_i(\tilde{\vartheta}(\psi))$ é zero, porém $L_{bi}(\tilde{\vartheta}(\psi))$ não. Então, pela regra da cadeia,

$$\frac{\partial}{\partial \psi^a} L_i(\tilde{\vartheta}(\psi)) = \frac{\partial \tilde{L}_i}{\partial \tilde{\vartheta}^r} \frac{\partial \tilde{\vartheta}^r}{\partial \psi^a} = \tilde{L}_{ir}(\psi) \tilde{\vartheta}_a^r(\psi) = 0,$$

onde $\tilde{L}_{ir}(\psi) = L_{ir}(\tilde{\vartheta}(\psi))$. Dado que

$$\tilde{L}_{ir}(\psi) \tilde{\vartheta}_a^r(\psi) = \tilde{L}_{ib}(\psi) \tilde{\vartheta}_a^b(\psi) + \tilde{L}_{ij}(\psi) \tilde{\vartheta}_a^j(\psi),$$

onde $\tilde{\vartheta}_a^b(\psi) = \delta_a^b$ é o delta de Kronecker, ou seja, $\delta_a^b = 1$ se $a = b$ e zero caso contrário, podemos escrever

$$0 = \tilde{L}_{ia}(\psi) + \tilde{L}_{ij}(\psi) \tilde{\vartheta}_a^j(\psi),$$

obtendo-se

$$\tilde{\vartheta}_a^j(\psi) = -\tilde{L}_{ij}^{-1}(\psi) \tilde{L}_{ia}(\psi). \quad (2.4)$$

Utilizando o resultado acerca da inversa de matriz simétrica particionada (Rao, 1973, pg.33), podemos escrever que o elemento (i, j) da matriz (L_{ij}^{-1}) é da forma

$$L_{ij}^{-1}(\psi) = L^{ij}(\psi) - L^{bj}(\psi) L^{ab}(\psi)^{-1} L^{ai}(\psi).$$

Substituindo esta relação em (2.4) vemos que

$$\tilde{\vartheta}_a^j(\psi) = -\tilde{L}^{ij}(\psi) \tilde{L}_{ia}(\psi) + \tilde{L}^{bj}(\psi) \tilde{L}^{ab}(\psi)^{-1} \tilde{L}^{ai}(\psi) \tilde{L}_{ia}(\psi).$$

Dado que $\tilde{L}^{ab}(\psi) \tilde{L}_{ab}(\psi) + \tilde{L}^{ai}(\psi) \tilde{L}_{ia}(\psi) = 1$, então

$$\tilde{L}^{ai}(\psi) \tilde{L}_{ia}(\psi) = 1 - \tilde{L}^{ab}(\psi) \tilde{L}_{ab}(\psi),$$

e, portanto,

$$\tilde{\vartheta}_a^j(\psi) = -\tilde{L}^{ij}(\psi) \tilde{L}_{ia}(\psi) + \tilde{L}^{bj}(\psi) \tilde{L}^{ab}(\psi)^{-1} - \tilde{L}^{bj}(\psi) \tilde{L}^{ab}(\psi)^{-1} \tilde{L}^{ab}(\psi) \tilde{L}_{ab}(\psi).$$

Como $\tilde{L}^{ab}(\psi)\tilde{L}^{ab}(\psi)^{-1} = 1$, então

$$\tilde{\vartheta}_a^j(\psi) = \tilde{L}^{ij}(\psi)\tilde{L}^{ia}(\psi) + \tilde{L}^{bj}(\psi)\tilde{L}^{ab}(\psi)^{-1} - \tilde{L}^{bj}(\psi)\tilde{L}^{ab}(\psi) = \tilde{L}^{bj}(\psi)\tilde{L}^{ab}(\psi)^{-1},$$

mas $\tilde{L}_{ab}(\psi)\tilde{L}^{bj}(\psi) + \tilde{L}_{ia}(\psi)\tilde{L}^{ij}(\psi) = 0$, concluindo-se que $\tilde{\vartheta}_a^j(\psi) = \tilde{L}^{bj}(\psi)\tilde{L}^{ab}(\psi)^{-1}$.

Podemos escrever, então,

$$\tilde{\vartheta}_a^r(\psi) = \tilde{L}^{br}(\psi)\tilde{L}^{ab}(\psi)^{-1}, \quad (2.5)$$

já que $\tilde{\vartheta}_a^a(\psi) = 1$ e $\tilde{\vartheta}_a^j(\psi) = \tilde{L}^{bj}(\psi)\tilde{L}^{ab}(\psi)^{-1} = 1$, como foi provado acima. Pelo Método Delta (Sen & Singer, 1993) obtemos que $\tilde{L}_{br}(\psi) = \lambda_{br}(\vartheta) + O_P(n^{-1/2})$, $\tilde{L}^{rb}(\psi) = \lambda^{rb} + O_P(n^{-1/2})$ e que $L^{ab}(\psi)^{-1} = \sigma^{ab} + O_P(n^{-1/2})$. Substituindo estas expressões em (2.5) obtemos

$$\tilde{\vartheta}_a^r(\psi) = \tilde{L}^{br}(\psi)\tilde{L}^{ab}(\psi)^{-1} = \lambda^{rb}\sigma_{ab} + O_P(n^{-1/2}).$$

Assim,

$$\begin{aligned} \tilde{A}^b \frac{\partial \tilde{L}_b}{\partial \psi^a} &= \tilde{A}^b \tilde{L}_{br} \tilde{\vartheta}_a^r + O_P(n^{-1/2}) \\ &= \lambda^{br} \nu^{st} \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right) \lambda_{br} \lambda^{cr} \sigma_{ac} + O_P(n^{-1/2}). \end{aligned}$$

Provaremos a seguir que $\sigma_{ab} = \lambda_{br} \lambda^{cr} \sigma_{ac}$. Por definição,

$$(\sigma_{ab}) = (\lambda^{ab})^{-1}$$

e

$$\begin{pmatrix} (\lambda_{bd}) & (\lambda_{bj}) \\ (\lambda_{id}) & (\lambda_{ij}) \end{pmatrix}^{-1} = \begin{pmatrix} (\lambda^{bd}) & (\lambda^{bj}) \\ (\lambda^{id}) & (\lambda^{ij}) \end{pmatrix},$$

então $\lambda^{bu} \lambda_{bu} = \lambda^{bd} \lambda_{bd} + \lambda^{bi} \lambda_{bi} = 1$, de onde temos que

$$\lambda^{bu} \lambda_{bu} \sigma_{ab} = \lambda^{bd} \lambda_{bd} \sigma_{ab} + \lambda^{di} \lambda_{id} \sigma_{ab}.$$

Assim,

$$\frac{\partial}{\partial \psi^a} \tilde{\Lambda}(\psi) = \sigma_{ab} \lambda^{br} \nu^{st} \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right) + O_P(n^{-1/2}),$$

como queríamos demonstrar.

2.1.2 Ajuste para o vício da informação da função escore perfilada

O diferencial do ajuste de Stern é apresentado no teorema a seguir. A idéia é encontrar uma função que adicionada ao logaritmo da função de verossimilhança perfilada corrija também o vício da informação, problema este comentado na Seção 1.1.1.

Teorema 2.1.2 *Seja*

$$\Theta(\vartheta) = \frac{1}{2}B^{ab}(\vartheta)L_a(\vartheta)L_b(\vartheta),$$

onde B^{ab} é dado por

$$\begin{aligned} B^{ab} &= \lambda^{ar}\lambda^{bs}\lambda^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rts}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rt/s}\lambda_{uv/w} + \lambda_{st/r}\lambda_{uv/w} \right. \\ &\quad \left. - \frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \frac{1}{2}\lambda_{st/r}\lambda_{uvw} \right) \\ &\quad + \lambda^{ar}\lambda^{bs}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} + \lambda_{rt/u}\lambda_{sv/v} \right. \\ &\quad \left. + \lambda_{rt/u}\lambda_{vw/s} + \lambda_{tu/r}\lambda_{sv/w} - \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} - \frac{1}{2}\lambda_{tu/r}\lambda_{svw} \right) \\ &\quad - \lambda^{ar}\lambda^{bs}\lambda^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su} + \lambda_{st/ru} - \frac{1}{2}\lambda_{rtu/s} - \frac{1}{2}\lambda_{stu/r} \right). \end{aligned} \quad (2.6)$$

Então, a função de verossimilhança perfilada ajustada, definida como

$$\bar{L}_P(\psi) = L_P(\psi) + \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi),$$

tem vício de função escore perfilada ajustada e vício de informação de ordem $O(n^{-1})$. Aqui, $\tilde{\Lambda}(\psi) = \tilde{A}^a(\psi)\tilde{L}_a(\psi)$ e $\tilde{\Theta}(\psi) = \frac{1}{2}\tilde{B}^{ab}(\psi)\tilde{L}_a(\psi)\tilde{L}_b(\psi)$.

Como vimos na Definição 1.2.1, a proposta de Stern (1997) é ajustar o logaritmo da função de verossimilhança perfilada na forma

$$\bar{L}_P(\psi) = L_P(\psi) + \tilde{A}^a\tilde{L}_a + \frac{1}{2}\tilde{B}^{ab}\tilde{L}_a\tilde{L}_b,$$

e aqui nos dedicaremos a encontrar a expressão correta para a função B^{ab} .

Observemos que se $B^{ab} = 0$ temos $\Theta(\vartheta) = 0$ e, desta forma, podemos afirmar que $\bar{L}_P(\psi)$ acima é uma generalização daquela obtida no Teorema 2.1.1. Observemos também que B^{ab} é uma função de ordem $O(n^{-2})$ e que a afirmação do teorema significa que

$$E\{\bar{U}_P(\psi)\} = O(n^{-1})$$

e

$$E\{\partial\bar{U}_P(\psi)/\partial\psi^\top\} + E\{\bar{U}_P(\psi)\bar{U}_P(\psi)^\top\} = O(n^{-1}).$$

Apresentamos a seguir a demonstração do teorema. Devemos provar primeiramente que a vantagem obtida na esperança da função escore perfilada ajustada não é perdida pelo acréscimo da função $\Theta(\vartheta)$, ou seja, devemos verificar que a esperança da função escore perfilada ajustada permanece de ordem $O(n^{-1})$. Neste sentido, calculemos $\bar{U}_P^a(\psi) = \partial\bar{L}_P(\psi)/\partial\psi^a$:

$$\bar{U}_P^a(\psi) = U_P^a(\psi) + \frac{\partial}{\partial\psi^a}\tilde{\Lambda}(\psi) + \frac{\partial}{\partial\psi^c}\tilde{\Theta}(\psi).$$

Na seção anterior foi provado que

$$E\left\{U_P^a(\psi) + \frac{\partial}{\partial\psi^a}\tilde{\Lambda}(\psi)\right\} = O(n^{-1}).$$

Por outro lado,

$$\frac{\partial}{\partial\psi^a}\left\{\frac{1}{2}\tilde{B}^{cd}\tilde{L}_c\tilde{L}_d\right\} = \frac{1}{2}\frac{\partial\tilde{B}^{cd}}{\partial\psi^a}\tilde{L}_c\tilde{L}_d + \frac{1}{2}\tilde{B}^{cd}\left\{\frac{\partial\tilde{L}_c}{\partial\psi^a}\tilde{L}_d + \tilde{L}_c\frac{\partial\tilde{L}_d}{\partial\psi^a}\right\},$$

que pode ser escrita como

$$\frac{\partial}{\partial\psi^a}\left\{\frac{1}{2}\tilde{B}^{cd}\tilde{L}_c\tilde{L}_d\right\} = \frac{1}{2}\tilde{B}_a^{cd}\tilde{L}_c\tilde{L}_d + \frac{1}{2}\tilde{B}^{cd}\left\{\tilde{L}_{cr}\tilde{\vartheta}_a^r\tilde{L}_d + \tilde{L}_c\tilde{L}_{ds}\tilde{\vartheta}_a^s\right\}.$$

Por expansão em série de Taylor, temos que

$$\tilde{B}^{ab}(\psi) = B^{ab}(\vartheta) + B_r^{ab}(\vartheta)(\tilde{\vartheta} - \vartheta)^r + O_P(n^{-5/2}),$$

onde $B_r^{ab}(\vartheta) = \partial B^{ab}(\vartheta)/\partial\vartheta^r$ e, portanto,

$$\tilde{B}^{ab}(\psi) = B^{ab}(\vartheta) + O_P(n^{-3/2}),$$

obtendo-se que $\tilde{B}^{ab}(\psi) = O_P(n^{-2})$.

Desta maneira, concluímos que $\tilde{\Theta}(\psi) = O_P(n^{-1})$ e, portanto,

$$E \{ \bar{U}_P^a(\psi) \} = O(n^{-1}).$$

Provemos a seguir que o vício da informação da função escote perfilada ajustada também é de ordem $O(n^{-1})$. Por simples diferenciação, temos que

$$\begin{aligned} \frac{\partial^2}{\partial \psi^a \partial \psi^b} \{ \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi) \} &= \frac{\partial}{\partial \psi^a} \left\{ \tilde{A}_b^c \tilde{L}_c + \tilde{A}^c \tilde{L}_{bc} + \frac{1}{2} \tilde{B}_b^{cd} \tilde{L}_c \tilde{L}_d + \frac{1}{2} \tilde{B}^{cd} \tilde{L}_{bc} \tilde{L}_d + \frac{1}{2} \tilde{B}^{cd} \tilde{L}_c \tilde{L}_{bd} \right\} \\ &= \tilde{A}_{ab}^c \tilde{L}_c + \tilde{A}_b^c \tilde{L}_{cr} \tilde{\vartheta}_a^r + \tilde{A}_a^c \tilde{L}_{br} \tilde{\vartheta}_c^r + \tilde{A}^c \tilde{L}_{abc} + \frac{1}{2} \left\{ \tilde{B}_{ab}^{cd} \tilde{L}_c \tilde{L}_d \right. \\ &\quad + \tilde{B}_b^{cd} \tilde{L}_{cr} \tilde{\vartheta}_a^r \tilde{L}_d + \tilde{B}_b^{cd} \tilde{L}_c \tilde{L}_{dr} \tilde{\vartheta}_a^r + \tilde{B}_a^{cd} \tilde{L}_{br} \tilde{\vartheta}_c^r \tilde{L}_d + \tilde{B}^{cd} \tilde{L}_{abc} \tilde{L}_d \\ &\quad \left. + \tilde{B}^{cd} \tilde{L}_{br} \tilde{\vartheta}_c^r \tilde{L}_{ds} \tilde{\vartheta}_a^s + \tilde{B}_a^{cd} \tilde{L}_c \tilde{L}_{br} \tilde{\vartheta}_d^r + \tilde{B}^{cd} \tilde{L}_{cr} \tilde{\vartheta}_a^r \tilde{L}_{bs} \tilde{\vartheta}_d^s + \tilde{B}^{cd} \tilde{L}_c \tilde{L}_{abd} \right\}. \end{aligned}$$

Agora, pela regra da cadeia

$$\tilde{L}_{abc} = \tilde{L}_{abr} \tilde{\vartheta}_c^r = \tilde{L}_{abr} \tilde{\vartheta}_c^r = \tilde{L}_{ars} \tilde{\vartheta}_b^s \tilde{\vartheta}_c^r = \tilde{L}_{rst} \tilde{\vartheta}_a^t \tilde{\vartheta}_b^s \tilde{\vartheta}_c^r,$$

e, portanto, $\tilde{L}_{abc} = \lambda_{rst} \lambda^{rd} \sigma_{cd} \lambda^{se} \sigma_{be} \lambda^{tf} \sigma_{af} + O_P(n^{-1/2})$. Em outras palavras, $\tilde{L}_{abc} = O_P(n)$.

Também pela regra da cadeia temos $\tilde{A}_c^a = \tilde{A}_r^a \tilde{\vartheta}_c^r$ e, portanto, $\tilde{A}_c^a = A_r^a \lambda^{rd} \sigma_{cd} + O_P(n^{-1/2})$.

Estamos em condições de perceber que os termos $\tilde{A}_{ab}^c \tilde{L}_c$, $\tilde{B}_{ab}^{cd} \tilde{L}_c \tilde{L}_d$, $\tilde{B}_b^{cd} \tilde{L}_{cr} \tilde{\vartheta}_a^r \tilde{L}_d$, $\tilde{B}_b^{cd} \tilde{L}_c \tilde{L}_{dr} \tilde{\vartheta}_a^r$, $\tilde{B}_a^{cd} \tilde{L}_{br} \tilde{\vartheta}_c^r \tilde{L}_d$, $\tilde{B}^{cd} \tilde{L}_{abc} \tilde{L}_d$, $\tilde{B}_a^{cd} \tilde{L}_c \tilde{L}_{br} \tilde{\vartheta}_d^r$ e $\tilde{B}^{cd} \tilde{L}_c \tilde{L}_{abd}$ são de ordem $O_P(n^{-1/2})$, enquanto que os outros são de ordem $O_P(1)$. Então,

$$\begin{aligned} \frac{\partial^2}{\partial \psi^a \partial \psi^b} \{ \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi) \} &= \tilde{A}_b^c \tilde{L}_{cr} \tilde{\vartheta}_a^r + \tilde{A}_a^c \tilde{L}_{br} \tilde{\vartheta}_c^r + \tilde{A}^c \tilde{L}_{abc} \\ &\quad + \frac{1}{2} \left\{ \tilde{B}^{cd} \tilde{L}_{br} \tilde{\vartheta}_c^r \tilde{L}_{ds} \tilde{\vartheta}_b^s + \tilde{B}^{cd} \tilde{L}_{cr} \tilde{\vartheta}_a^r \tilde{L}_{bs} \tilde{\vartheta}_d^s \right\} + O_P(n^{-1/2}) \\ &= A_s^c \lambda^{se} \sigma_{be} \lambda_{cr} \lambda^{rd} \sigma_{ad} + A_s^c \lambda^{se} \sigma_{ae} \lambda_{cr} \lambda^{rd} \sigma_{ad} \\ &\quad + A^c \lambda_{rst} \lambda^{rd} \sigma_{cd} \lambda^{se} \sigma_{be} \lambda^{tf} \sigma_{af} + \frac{1}{2} \left\{ B^{cd} \lambda_{br} \lambda^{re} \sigma_{ce} \lambda_{ds} \lambda^{sf} \sigma_{af} \right. \\ &\quad \left. + B^{cd} \lambda_{cr} \lambda^{re} \sigma_{ae} \lambda_{bs} \lambda^{sf} \sigma_{df} \right\} + O_P(n^{-1/2}). \end{aligned}$$

Definindo

$$\Psi_{ab} = E \left\{ \frac{\partial^2}{\partial \psi^a \partial \psi^b} \left\{ \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi) \right\} \right\},$$

obtemos

$$\begin{aligned} \Psi_{ab} = & A_r^c \lambda^{re} \sigma_{be} \sigma_{ac} + A_r^c \lambda^{re} \sigma_{ae} \sigma_{ac} + A^c \lambda_{rst} \lambda^{rd} \sigma_{cd} \lambda^{se} \sigma_{be} \lambda^{tf} \sigma_{af} \\ & + \frac{1}{2} \{ B^{cd} \sigma_{bc} \sigma_{ad} + B^{cd} \sigma_{ac} \sigma_{bd} \} + O(n^{-1}), \end{aligned} \quad (2.7)$$

recordando que, na demonstração do Teorema 2.1.1, provamos a relação $\sigma_{ab} = \lambda_{br} \lambda^{cr} \sigma_{ac}$.

Na expressão acima desconhecemos a forma das funções A_r^c e B^{ab} . Na Seção A.3, do Apêndice A, demonstramos que A_r^c é da forma

$$\begin{aligned} A_r^c = & -\lambda^{cs} \lambda^{tu} \nu^{vw} \lambda_{st/r} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) - \lambda^{cu} \nu^{tv} \nu^{sw} \lambda_{st/r} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) \\ & + \lambda^{cu} \nu^{vw} \left(\lambda_{uv/rw} - \frac{1}{2} \lambda_{uvw/r} \right). \end{aligned}$$

Substituindo A_r^c em (2.7) e observando que $\tau^{rs} = \lambda^{ra} \lambda^{sb} \sigma_{ab}$, obtemos

$$\begin{aligned} \Psi_{ab} = & \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \tau^{tu} \nu_{vw} \lambda_{rst} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) \\ & - (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \lambda^{tv} \nu^{uw} \lambda_{rt/s} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) \\ & - (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \lambda_{tu/r} \left(\lambda_{sv/w} - \frac{1}{2} \lambda_{svw} \right) \\ & + (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\lambda_{st/su} - \frac{1}{2} \lambda_{rtu/s} \right) \\ & + \frac{1}{2} \sigma_{ac} \sigma_{bd} (B^{cd} + B^{dc}) + O(n^{-1}). \end{aligned}$$

DiCiccio *et al.* (1996) provaram que, para qualquer função de ajuste adicionada à função escore perfilada, o vício da informação será de ordem $O(n^{-1})$ se

$$\zeta_a^b = \Delta_{ac} \lambda^{bc} - \rho_a \rho_b \lambda^{bc} + \rho_{a/r} \lambda^{br} + \rho_{c/r} \sigma_{ad} \lambda^{bc} \lambda^{dr} - \Psi_{ac} \lambda^{bc} + O(n^{-1}), \quad (2.8)$$

onde $\Delta_{ab} = E\{\tilde{L}_a(\psi) \tilde{L}_b(\psi)\} + E\{\tilde{L}_{ab}(\psi)\}$, $E\{\tilde{L}_a(\psi)\} = -\rho_a + O(n^{-1})$ e ζ_a^b é uma variável não-aleatória de ordem $O(n^{-1})$. Também DiCiccio *et al.* (1996) informam que, para obter a função de ajuste que faz o vício da informação ser de ordem $O(n^{-1})$ e que satisfaz (2.3), devemos fazer $\zeta_a^b = 0$ em (2.8).

As expressões de $\rho_{a/r}$ e Δ_{ab} foram obtidas por DiCiccio *et al.* (1996) e são dadas por

$$\begin{aligned} \rho_{a/r} = & \sigma_{ab} \lambda^{bs} \nu^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{st/r} \lambda_{uvw} - \lambda_{st/r} \lambda_{uv/w} + \frac{1}{2} \lambda_{tu/r} \lambda_{svw} - \lambda_{tu/r} \lambda_{sv/w} \right) \\ & - \sigma_{ab} \lambda^{bs} \nu^{tu} \left(\frac{1}{2} \lambda_{stu/r} - \lambda_{st/ru} \right) \end{aligned} \quad (2.9)$$

e

$$\begin{aligned} \Delta_{ab} = & -\sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\frac{1}{2} \lambda_{rstu} - \lambda_{rst/u} - \frac{1}{2} \lambda_{stu/r} - \frac{1}{2} \lambda_{rtu/s} + \lambda_{st/ru} + \lambda_{rt/su} \right) \\ & + \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rst} \lambda_{uvw} - \lambda_{rst} \lambda_{uv/w} - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} - \frac{1}{2} \lambda_{st/r} \lambda_{uvw} \right. \\ & + \lambda_{rt/s} \lambda_{uv/w} + \lambda_{st/r} \lambda_{uv/w} + \frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} \\ & - \frac{1}{2} \lambda_{rtu} \lambda_{vw/s} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} + \lambda_{rt/u} \lambda_{vw/s} + \lambda_{rt/u} \lambda_{sw/v} + \lambda_{tu/r} \lambda_{sv/w} \\ & \left. + \frac{1}{4} \lambda_{rtv} \lambda_{suw} - \frac{1}{2} \lambda_{rtv} \lambda_{su/w} - \frac{1}{2} \lambda_{rt/v} \lambda_{suw} + \lambda_{rt/v} \lambda_{su/w} \right) + O(n^{-1}). \end{aligned}$$

Se assumimos que a função B^{ab} é simétrica, no sentido de $B^{cd} = B^{dc}$, temos que

$$\begin{aligned} \Psi_{ab} = & \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \tau^{tu} \nu_{vw} \lambda_{rst} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) \\ & - (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \lambda^{tv} \nu^{uw} \lambda_{rt/s} \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) \\ & - (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \lambda_{tu/r} \left(\lambda_{sv/w} - \frac{1}{2} \lambda_{svw} \right) \\ & + (\sigma_{ac} \sigma_{bd} + \sigma_{bc} \sigma_{ad}) \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\lambda_{st/su} - \frac{1}{2} \lambda_{rtu/s} \right) \\ & + \sigma_{ac} \sigma_{bd} B^{cd} + O(n^{-1}). \end{aligned} \quad (2.10)$$

A vantagem de se exigir simetria na função B^{ab} é que as derivadas parciais $\partial \bar{U}_P^a / \partial \psi^b$ e $\partial \bar{U}_P^b / \partial \psi^a$ serão iguais. Isto garante, por exemplo, que a matriz de informação esperada será simétrica.

Substituindo as expressões de Δ_{ab} , ρ_a , $\rho_{a/r}$ e Ψ_{ab} em (2.8) quando $\zeta_a^b = 0$, temos

$$\begin{aligned}
0 = & -\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} - \frac{1}{2}\lambda_{stu/r} - \frac{1}{2}\lambda_{rtu/s} + \lambda_{st/ru} + \lambda_{rt/su}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\frac{1}{2}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \frac{1}{2}\lambda_{st/r}\lambda_{uvw}\right. \\
& + \lambda_{rt/s}\lambda_{uv/w} + \lambda_{st/r}\lambda_{uv/w} + \frac{1}{2}\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} - \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} \\
& - \frac{1}{2}\lambda_{tu/r}\lambda_{svw} + \lambda_{rt/u}\lambda_{vw/s} + \lambda_{rt/u}\lambda_{sw/v} + \lambda_{tu/r}\lambda_{sv/w} + \frac{1}{4}\lambda_{rtu}\lambda_{suw} - \frac{1}{2}\lambda_{rtu}\lambda_{su/w} \\
& \left. - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} + \lambda_{rt/v}\lambda_{su/w}\right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\nu^{vw}\left(\lambda_{rt/u}\lambda_{sv/w} - \frac{1}{2}\lambda_{rt/u}\lambda_{svw} - \frac{1}{2}\lambda_{rtu}\lambda_{sv/w} + \frac{1}{4}\lambda_{rtu}\lambda_{svw}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \lambda_{tu/s}\lambda_{rv/w}\right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\frac{1}{2}\lambda_{st/r}\lambda_{uvw} - \lambda_{st/r}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/r}\lambda_{svw} - \lambda_{tu/r}\lambda_{sv/w}\right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\frac{1}{2}\lambda_{stu/r} - \lambda_{st/ru}\right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\nu^{vw}\lambda_{rst}\left(\lambda_{uv/w} - \frac{1}{2}\lambda_{uvw}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{vw}\lambda_{rt/s}\left(\lambda_{uv/w} - \frac{1}{2}\lambda_{uvw}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\nu^{vw}\lambda_{st/r}\left(\lambda_{uv/w} - \frac{1}{2}\lambda_{uvw}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\lambda_{tu/r}\left(\lambda_{sv/w} - \frac{1}{2}\lambda_{svw}\right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\lambda_{tu/s}\left(\lambda_{rv/w} - \frac{1}{2}\lambda_{rvw}\right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\lambda_{rt/su} - \frac{1}{2}\lambda_{rtu/s}\right) - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\lambda_{st/ru} - \frac{1}{2}\lambda_{stu/r}\right) \\
& - \sigma_{ac}\sigma_{bd}B^{cd}.
\end{aligned}$$

Então,

$$\begin{aligned}
\sigma_{ac}\sigma_{bd}B^{cd} = & \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} - \frac{1}{2}\lambda_{stu/r} - \frac{1}{2}\lambda_{rtu/s} + \lambda_{st/ru} + \lambda_{rt/su} \right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rt/s}\lambda_{uvw} \right. \\
& - \frac{1}{2}\lambda_{st/r}\lambda_{uvw} + \lambda_{rt/s}\lambda_{uv/w} + \lambda_{st/r}\lambda_{uv/w} + \frac{1}{2}\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} \\
& - \lambda_{rt/u}\lambda_{svw} - \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} - \frac{1}{2}\lambda_{tu/r}\lambda_{svw} + \lambda_{rt/u}\lambda_{vw/s} + \lambda_{rt/u}\lambda_{sv/w} \\
& + \lambda_{tu/r}\lambda_{sv/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} + \lambda_{rt/v}\lambda_{su/w} \\
& + \frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \lambda_{tu/s}\lambda_{rv/w} + \frac{1}{2}\lambda_{st/r}\lambda_{uvw} \\
& - \lambda_{st/r}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/r}\lambda_{svw} - \lambda_{tu/r}\lambda_{sv/w} + \lambda_{tu/r}\lambda_{sv/w} - \frac{1}{2}\lambda_{tu/r}\lambda_{svw} \\
& \left. + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\lambda^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} \right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} \right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\lambda^{tv}\nu^{uw} \left(\lambda_{rt/s}\lambda_{uv/w} - \frac{1}{2}\lambda_{rt/s}\lambda_{uvw} \right) \\
& + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\lambda^{tv}\nu^{uw} \left(\lambda_{st/r}\lambda_{uv/w} - \frac{1}{2}\lambda_{st/r}\lambda_{uvw} \right).
\end{aligned}$$

Desta equação obtemos a expressão de B^{ab} dada em (2.6). Observemos que esta expressão é diferente daquela publicada em Stern (1997), a diferença estando no coeficiente $\lambda^{ar}\lambda^{bs}\nu^{tv}\nu^{uw}$. O termo correspondente no artigo mencionado é $\lambda^{ar}\lambda^{bs}\lambda^{tv}\nu^{uw}$. Na seção seguinte obtemos uma função de verossimilhança perfilada ajustada sem fazer uso da equação (2.6) e no Capítulo 3 utilizamos nosso resultado para B^{ab} em diferentes exemplos.

2.1.3 Exemplo: modelo de regressão linear normal

Seja $Y \sim N(X\beta, \phi^{-1}I)$, onde X é uma matriz $n \times q$, β é um vetor coluna de dimensão q , $\phi^{-1} > 0$ e I é a matriz identidade de dimensão n . Consideraremos como parâmetro de interesse o de dispersão ϕ e o vetor de parâmetros de regressão β , o de perturbação.

Neste modelo sabemos que $\hat{\phi}$, o estimador de máxima verossimilhança de ϕ , é n/S^2 , onde

$$S^2 = \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2,$$

e $\hat{\mu} = X\hat{\beta}$, sendo $\hat{\beta} = (X^T X)^{-1} X^T Y$ o estimador de máxima verossimilhança de β .

O logaritmo da função de verossimilhança deste modelo é

$$L(\beta, \phi) = -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \phi - \frac{\phi}{2} \sum_{i=1}^n (Y_i - \mu_i)^2,$$

e o logaritmo da função de verossimilhança perfilada é

$$L_P(\phi) = -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \phi - \frac{\phi}{2} S^2,$$

com a qual obtemos que a função escore perfilada é

$$U_P(\phi) = \frac{n}{2\phi} - \frac{S^2}{2}.$$

A esperança desta função é

$$E \left\{ \frac{n}{2\phi} - \frac{S^2}{2} \right\} = \frac{n}{2\phi} - \frac{n-q}{2\phi} = \frac{q}{2\phi},$$

que não é nula. Utilizamos aqui o fato de que $S^2 \sim \phi^{-1} \chi^2(n-q)$, que implica em $E\{S^2\} = (n-q)\phi^{-1}$ e $E\{S^2\}^2 = \text{var}\{S^2\} + E^2\{S^2\} = 2(n-q)\phi^{-2} + (n-q)^2\phi^{-2}$.

Vamos obter agora a função de verossimilhança perfilada ajustada no modelo de regressão linear normal múltipla quando o interesse é o parâmetro ϕ sem recorrer às expressões gerais, ou seja, não utilizaremos as expressões das funções $\Lambda(\vartheta)$ e $\Theta(\vartheta)$, dos Teoremas 2.1.1 e 2.1.2, para encontrar a função de verossimilhança perfilada ajustada de Stern. O leitor estaria esperando a utilização dos teoremas mencionados no modelo de regressão normal, porém utilizaremos

amplamente os resultados nestes teoremas no Capítulo 3. Interessa-nos aqui aproveitar o fato de podermos obter expressões fechadas para as funções $\Lambda(\vartheta)$ e $\Theta(\vartheta)$ e, posteriormente, no Capítulo 3 considerar este exemplo como uma situação particular dos modelos lineares generalizados.

Segundo a Definição 1.2.1,

$$\tilde{\Lambda}(\phi) = \tilde{A}(\phi)U_P(\phi),$$

onde $\tilde{\Lambda}(\phi) = \Lambda(\tilde{\vartheta}(\phi))$, este termo sendo tal que

$$\mathbb{E} \left\{ \frac{d}{d\phi} \tilde{\Lambda}(\phi) \right\} = \rho(\phi) + O(n^{-1}),$$

com

$$\rho(\phi) = -\mathbb{E} \{U_P(\phi)\} + O(n^{-1}). \quad (2.11)$$

Encontraremos em seguida a expressão de $\rho(\phi)$ e para isso devemos achar a esperança de $d\tilde{\Lambda}(\phi)/d\phi$, que é dada por

$$\frac{d}{d\phi} \tilde{\Lambda}(\phi) = \left(\frac{d}{d\phi} \tilde{A}(\phi) \right) U_P(\phi) + \tilde{A}(\phi) \left(\frac{d}{d\phi} U_P(\phi) \right)$$

e

$$\frac{d}{d\phi} U_P(\phi) = -\frac{n}{2\phi^2}.$$

Logo,

$$\frac{d}{d\phi} \tilde{\Lambda}(\phi) = \left(\frac{d}{d\phi} \tilde{A}(\phi) \right) U_P(\phi) + \tilde{A}(\phi) \left(-\frac{n}{2\phi^2} \right)$$

e, portanto,

$$\begin{aligned} \mathbb{E} \left\{ \frac{d}{d\phi} \tilde{\Lambda}(\phi) \right\} &= \mathbb{E} \left\{ \left(\frac{d}{d\phi} \tilde{A}(\phi) \right) U_P(\phi) \right\} + \tilde{A}(\phi) \left(-\frac{n}{2\phi^2} \right) \\ &= \frac{q}{2\phi} \mathbb{E} \left\{ \frac{d}{d\phi} \tilde{A}(\phi) \right\} - \frac{n}{2\phi^2} \tilde{A}(\phi). \end{aligned}$$

Observemos que se $A(\phi) = q\phi/n$, $\tilde{A}(\phi) = A(\phi)$, então $dA(\phi)/d\phi = q/n$ e, desta forma,

$$\mathbb{E} \left\{ \frac{d}{d\phi} \tilde{\Lambda}(\phi) \right\} = \frac{q^2}{2n\phi} - \frac{q}{2\phi} = -\frac{q}{2\phi} + O(n^{-1}).$$

Escolhendo

$$\rho(\phi) = -\frac{q}{2\phi},$$

(2.11) é satisfeita, concluindo-se então que

$$\Lambda(\phi) = \frac{q\phi}{n} U_P(\phi) = \frac{q\phi}{n} \frac{n}{2\phi} \left(1 - \frac{\phi}{\widehat{\phi}}\right)$$

e obtendo-se

$$\Lambda(\phi) = \frac{q}{2} \left(1 - \frac{\phi}{\widehat{\phi}}\right).$$

Definimos

$$\begin{aligned} \widetilde{\Theta}(\phi) &= \frac{1}{2} \widetilde{B}^{ab} \widetilde{L}_a \widetilde{L}_b = \frac{1}{2} \widetilde{B}^{\phi\phi} U_P(\phi) U_P(\phi) \\ &= \frac{1}{2} \widetilde{B}^{\phi\phi} \left(\frac{n}{2\phi}\right)^2 \left(1 - \frac{\phi}{\widehat{\phi}}\right)^2, \end{aligned}$$

onde $\widetilde{\Theta}(\phi) = \Theta(\widetilde{\vartheta}(\phi))$.

Fazendo $\zeta_a^b = 0$, onde ζ_a^b vem da equação (2.8), obtemos

$$\Delta_{\phi\phi} \lambda^{\phi\phi} - \rho_{\phi} \rho_{\phi} \lambda^{\phi\phi} + \rho_{\phi/r} \lambda^{\phi r} + \sigma_{\phi\phi} \rho_{\phi/r} \lambda^{\phi\phi} \lambda^{\phi r} - \Psi_{\phi\phi} \lambda^{\phi\phi} = O(n^{-2}),$$

e dado que $\lambda^{\phi r} = 0$ se $r \neq \phi$ então

$$\Delta_{\phi\phi} \lambda^{\phi\phi} - \rho_{\phi}^2 \lambda^{\phi\phi} + \rho_{\phi/\phi} \lambda^{\phi\phi} + \sigma_{\phi\phi} \rho_{\phi/\phi} (\lambda^{\phi\phi})^2 - \Psi_{\phi\phi} \lambda^{\phi\phi} = O(n^{-2}).$$

Note-se que $\lambda^{\phi\phi} = 1/\lambda_{\phi\phi}$ e que

$$\lambda_{\phi\phi} = \mathbb{E} \left\{ \frac{\partial^2}{\partial \phi^2} \left(-\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \phi - \frac{\phi}{2} \sum_{i=1}^n (Y_i - \mu_i)^2 \right) \right\}.$$

Observemos que

$$\frac{\partial}{\partial \phi} \left(-\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \phi - \frac{\phi}{2} \sum_{i=1}^n (Y_i - \mu_i)^2 \right) = \frac{n}{2\phi} - \frac{1}{2} \sum_{i=1}^n (Y_i - \mu_i)^2.$$

Então,

$$\frac{\partial^2}{\partial \phi^2} \left(-\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \phi - \frac{\phi}{2} \sum_{i=1}^n (Y_i - \mu_i)^2 \right) = -\frac{n}{2\phi^2}.$$

Logo, $\lambda_{\phi\phi} = -n/2\phi^2$, $\lambda^{\phi\phi} = -2\phi^2/n$ e $\sigma_{\phi\phi} = -n/2\phi^2$. Assim,

$$-\frac{2\phi^2}{n}\Delta_{\phi\phi} - \left(-\frac{q}{2\phi}\right)^2 \left(-\frac{2\phi^2}{n}\right) + \rho_{\phi/\phi} \left(-\frac{2\phi^2}{n}\right) + \sigma_{\phi\phi}\rho_{\phi/\phi} \left(-\frac{2\phi^2}{n}\right)^2 - \Psi_{\phi\phi} \left(-\frac{2\phi^2}{n}\right) = O(n^{-2}). \quad (2.12)$$

Por outro lado,

$$\rho_{\phi/\phi} = \frac{d}{d\phi}\rho_{\phi} = \frac{d}{d\phi} \left(-\frac{q}{2\phi}\right) = \frac{q}{2\phi^2},$$

e substituindo $\sigma_{\phi\phi}$ e $\rho_{\phi/\phi}$ em (2.12), obtemos

$$-\frac{2\phi^2}{n}\Delta_{\phi\phi} + \frac{q^2}{2n} + \left(\frac{q}{2\phi^2}\right) \left(-\frac{2\phi^2}{n}\right) + \left(-\frac{n}{2\phi^2}\right) \left(\frac{q}{2\phi^2}\right) \frac{4q\phi^4}{n^2} + \frac{2\phi^2}{n}\Psi_{\phi\phi} = O(n^{-2}),$$

ou seja,

$$-\frac{2\phi^2}{n}\Delta_{\phi\phi} + \frac{q^2}{2n} - \frac{2q}{n} + \frac{2\phi^2}{n}\Psi_{\phi\phi} = O(n^{-2}).$$

Nesta situação,

$$\begin{aligned} \Delta_{\phi\phi} = & -\sigma_{\phi\phi}\sigma_{\phi\phi}\lambda^{\phi\phi}\lambda^{\phi\phi}\nu^{tu} \left(\frac{1}{2}\lambda_{\phi\phi tu} - \lambda_{\phi\phi t/u} - \frac{1}{2}\lambda_{\phi tu/\phi} - \frac{1}{2}\lambda_{\phi tu/\phi} + \lambda_{\phi t/\phi u} + \lambda_{\phi t/\phi u} \right) \\ & + \sigma_{\phi\phi}\sigma_{\phi\phi}\lambda^{\phi\phi}\lambda^{\phi\phi}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{\phi\phi t}\lambda_{uvw} - \lambda_{\phi\phi t}\lambda_{uv/w} - \frac{1}{2}\lambda_{\phi t/\phi}\lambda_{uvw} - \frac{1}{2}\lambda_{\phi t/\phi}\lambda_{uvw} \right. \\ & + \lambda_{\phi t/\phi}\lambda_{uv/w} + \lambda_{\phi t/\phi}\lambda_{uv/w} + \frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} - \lambda_{rtu}\lambda_{\phi v/w} - \lambda_{\phi t/u}\lambda_{\phi vw} \\ & - \frac{1}{2}\lambda_{\phi tu}\lambda_{uv/\phi} - \frac{1}{2}\lambda_{tu/\phi}\lambda_{\phi vw} + \lambda_{\phi t/u}\lambda_{uv/\phi} + \lambda_{\phi t/u}\lambda_{\phi v/w} + \lambda_{tu/\phi}\lambda_{\phi v/w} \\ & \left. + \frac{1}{4}\lambda_{\phi tv}\lambda_{\phi uw} - \frac{1}{2}\lambda_{\phi tv}\lambda_{\phi u/w} - \frac{1}{2}\lambda_{\phi t/v}\lambda_{\phi uw} + \lambda_{\phi t/v}\lambda_{\phi u/w} \right) + O(n^{-1}), \end{aligned}$$

e os cumulantes $\lambda_{\phi\phi tu}$, $\lambda_{\phi\phi t/u}$, $\lambda_{\phi tu/\phi}$, $\lambda_{\phi t/\phi u}$, $\lambda_{\phi\phi t}$, $\lambda_{\phi t/\phi}$, $\lambda_{\phi v/w}$, $\lambda_{\phi t/u}$, $\lambda_{\phi u/w}$ e $\lambda_{\phi t/v}$ são nulos.

Portanto

$$\Delta_{\phi\phi} = \nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} - \frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} - \frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} + \frac{1}{4}\lambda_{\phi tv}\lambda_{\phi uw} \right) + O(n^{-1}),$$

que é simplesmente

$$\Delta_{\phi\phi} = -\frac{1}{2}\nu^{tv}\nu^{uw}\lambda_{\phi tu}\lambda_{\phi vw} + \frac{1}{4}\nu^{tv}\nu^{uw}\lambda_{\phi tv}\lambda_{\phi uw} + O(n^{-1}),$$

já que $\lambda_{vw/\phi} = \lambda_{\phi vw}$ e $\lambda_{tu/\phi} = \lambda_{\phi tu}$.

A expressão do cumulante λ_{tu} é dada em (B.3), e dado que $\lambda_{\phi tu} = d\lambda_{tu}/d\phi$, obtemos que $\lambda_{\phi tu} = -\sum_{l=1}^n x_{lt}x_{lu}$. Portanto,

$$-\frac{1}{2}\nu^{tv}\nu^{uw}\lambda_{\phi tu}\lambda_{\phi vw} = -\frac{1}{2}\sum_{l,m=1}^n\sum_{t,v=1}^{p+q}\nu^{tv}x_{lt}x_{mv}\sum_{u,w=1}^{p+q}\nu^{uw}x_{lu}x_{mw} = -\frac{q}{2\phi^2}$$

e

$$\frac{1}{4}\nu^{tv}\nu^{uw}\lambda_{\phi tv}\lambda_{\phi uw} = \frac{1}{4}\sum_{l,m=1}^n\sum_{t,v=1}^{p+q}\nu^{tv}x_{lt}x_{lv}\sum_{u,w=1}^{p+q}\nu^{uw}x_{mu}x_{mw} = \frac{q^2}{4\phi^2},$$

e, desta forma, obtemos que

$$\Delta_{\phi\phi} = -\frac{q}{2\phi^2} + \frac{q^2}{4\phi^2}.$$

As expressões dos cumulantes e as somas anteriores são desenvolvidas em detalhes no próximo capítulo e no Apêndice B. Então,

$$\frac{2\phi^2}{n}\Psi_{\phi\phi} = \frac{q}{n} + O(n^{-2})$$

ou

$$\Psi_{\phi\phi} = \frac{q}{2\phi^2} + O(n^{-1}),$$

e desta expressão obteremos a função $B^{\phi\phi}$.

Por definição

$$\Psi_{\phi\phi} = E\left\{\frac{d^2}{d\phi^2}\left(\tilde{\Lambda}(\phi) + \tilde{\Theta}(\phi)\right)\right\},$$

onde

$$\frac{d^2}{d\phi^2}\tilde{\Lambda}(\phi) = 0,$$

já que

$$\frac{d}{d\phi}\frac{q}{2}\left(1 - \frac{\phi}{\hat{\phi}}\right) = -\frac{q}{2\hat{\phi}}.$$

Logo,

$$\Psi_{\phi\phi} = E\left\{\frac{d^2}{d\phi^2}\tilde{\Theta}(\phi)\right\}$$

e

$$\tilde{\Theta}(\phi) = \frac{1}{2}B^{\phi\phi}U_P^2(\phi) = \frac{1}{2}B^{\phi\phi}\left(\frac{n}{2\phi} - \frac{S^2}{2}\right)^2.$$

Portanto,

$$\begin{aligned}\frac{d}{d\phi}\tilde{\Theta}(\phi) &= \frac{1}{2}\left(\frac{d}{d\phi}B^{\phi\phi}\right)\left(\frac{n}{2\phi}-\frac{S^2}{2}\right)^2 + \frac{1}{2}B^{\phi\phi}\frac{d}{d\phi}\left(\frac{n}{2\phi}-\frac{S^2}{2}\right)^2 \\ &= \frac{1}{2}\left(\frac{d}{d\phi}B^{\phi\phi}\right)\left(\frac{n}{2\phi}-\frac{S^2}{2}\right)^2 + B^{\phi\phi}\left(\frac{nS^2}{4\phi^2}-\frac{n^2}{4\phi^3}\right)\end{aligned}$$

e

$$\begin{aligned}\frac{d^2}{d\phi^2}\tilde{\Theta}(\phi) &= \frac{1}{2}\left(\frac{d^2}{d\phi^2}B^{\phi\phi}\right)\left(\frac{n}{2\phi}-\frac{S^2}{2}\right)^2 + 2\left(\frac{d}{d\phi}B^{\phi\phi}\right)\left(\frac{nS^2}{4\phi^2}-\frac{n^2}{4\phi^3}\right) \\ &\quad + B^{\phi\phi}\left(\frac{3n^2}{4\phi^4}-\frac{nS^2}{2\phi^3}\right).\end{aligned}$$

Podemos, então, escrever

$$\begin{aligned}\mathbb{E}\left\{\frac{d^2}{d\phi^2}\tilde{\Theta}(\phi)\right\} &= \frac{1}{2}\left(\frac{d^2}{d\phi^2}B^{\phi\phi}\right)\mathbb{E}\left\{\frac{n}{2\phi}-\frac{S^2}{2}\right\}^2 - \frac{nq}{2\phi^3}\left(\frac{d}{d\phi}B^{\phi\phi}\right) \\ &\quad + \left(\frac{n^2}{4\phi^4} + \frac{nq}{2\phi^4}\right)B^{\phi\phi} + O(n^{-1})\end{aligned}\tag{2.13}$$

e

$$\mathbb{E}\left\{\frac{n}{2\phi}-\frac{S^2}{2}\right\}^2 = \frac{2n-2q+q^2}{4\phi^2}.$$

Seja

$$B^{\phi\phi} = \frac{2q}{n^2}\phi^2,$$

então $dB^{\phi\phi}/d\phi = 4q/n^2\phi$ e $d^2B^{\phi\phi}/d\phi^2 = 4q/n^2$.

Substituindo $B^{\phi\phi}$, $dB^{\phi\phi}/d\phi$ e $d^2B^{\phi\phi}/d\phi^2$ em (2.13), obtemos que

$$\mathbb{E}\left\{\frac{d^2}{d\phi^2}\tilde{\Theta}(\phi)\right\} = \frac{q}{2\phi^2} + O(n^{-1}),$$

como queríamos demonstrar.

Encontramos que a função de verossimilhança perfilada ajustada para ϕ é da forma

$$\bar{L}_P(\phi) = -\frac{n}{2}\log 2\pi + \frac{n}{2}\log \phi - \frac{\phi}{S^2} + \frac{q}{2}\left(1 - \frac{\phi}{\hat{\phi}}\right) + \frac{q}{4}\left(1 - \frac{\phi}{\hat{\phi}}\right)^2,\tag{2.14}$$

e durante o desenvolvimento deste exemplo mostramos a dificuldade para encontrar as funções de ajuste Λ e Θ . O Capítulo 3 será dedicado a obter estas funções nos modelos lineares generalizados.

2.2 Correção de Bartlett para o teste da razão de verossimilhanças ajustado

Em problemas regulares, a estatística da razão de verossimilhanças $\mathcal{W}(\psi)$ possui distribuição assintótica qui-quadrado com p graus de liberdade (ver, por exemplo Wilks, 1962, p. 410, Teorema 13.4.3) e erro de ordem $O(n^{-1})$. Isto implica que

$$E_{\vartheta}\{\mathcal{W}(\psi)\} = p + O(n^{-1}).$$

A idéia da correção para o teste da razão de verossimilhanças, originalmente devida a Bartlett (1937), visa à construção de uma modificação da estatística \mathcal{W} que tenha distribuição mais próxima da distribuição qui-quadrado de referência.

Seguindo as técnicas de Bartlett (1953a,b, 1955) para construir intervalos de confiança modificados, Lawley (1956) demonstrou, através de um procedimento algébrico complicado, que é possível definir uma estatística modificada \mathcal{W}^* , cuja esperança mantenha-se p a menos de um erro de ordem $O(n^{-2})$.

Lawley obteve o valor esperado de \mathcal{W} até ordem $O(n^{-1})$ como

$$E_{\vartheta}\{\mathcal{W}(\psi)\} = p \{1 + \xi(\vartheta)/p + O(n^{-2})\},$$

onde $\xi(\vartheta)$ (Cordeiro, 1987) é tipicamente de ordem $O(n^{-1})$, depende de cumulantes de derivadas do logaritmo da função de verossimilhança e eventualmente pode depender do parâmetro. Em situações como estas, o argumento dado em Barndorff-Nielsen & Hall (1988) mostra que se substituirmos $\xi(\vartheta)$ por $\xi(\hat{\vartheta})$, a aproximação por χ^2 continua de ordem $O(n^{-1})$.

Lawley mostrou ainda que, definindo

$$\mathcal{W}^*(\psi) = \mathcal{W}(\psi)/\{1 + \xi(\vartheta)/p\},$$

onde $1 + \xi(\vartheta)/p$ é conhecido como fator de correção de Bartlett, todos os cumulantes de \mathcal{W}^* coincidem com os correspondentes da distribuição $\chi^2(p)$, eliminando-se termos de ordem iguais ou inferiores a $O(n^{-2})$. Desta forma, temos

$$E_{\vartheta}\{\mathcal{W}^*(\psi)\} = p + O(n^{-2}).$$

Inúmeros artigos foram publicados apresentando correções de Bartlett em problemas específicos e para modelos amplos, bem como, métodos alternativos de cálculo das correções de Bartlett. A expressão para $\xi(\vartheta)$ fornecida por Lawley (1956) foi posteriormente muito explorada (Cordeiro, 1983, 1987; Cordeiro & Paula, 1989; Cordeiro & McCullagh, 1991; Cordeiro, Paula & Botter, 1994, entre outros) e estendida a outras estatísticas de teste que convergem para uma distribuição qui-quadrado sob a hipótese nula (Cordeiro & Ferrari, 1991; Cordeiro, Ferrari & Paula, 1993; Cribari-Neto & Ferrari, 1995, entre outros).

Definamos a função de verossimilhança perfilada ajustada como $\bar{L}_P(\psi) = L_P(\psi) + M(\psi)$, onde $M(\psi)$ é uma função, chamada de ajuste, de derivadas de ordem $O_P(1)$. Por exemplo, no Teorema 2.1.1 podemos escrever $M(\psi) = \tilde{\Lambda}(\psi)$ e no Teorema 2.1.2, $M(\psi) = \tilde{\Lambda}(\psi) + \tilde{\Theta}(\psi)$.

Para realizar inferências, definimos a estatística da razão de verossimilhanças ajustada como

$$\bar{\mathcal{W}}(\psi) = 2\{\bar{L}_P(\bar{\psi}) - \bar{L}_P(\psi)\},$$

onde

$$\bar{L}_P(\bar{\psi}) = \max_{\psi} \bar{L}_P(\psi).$$

Somando e subtraindo $L_P(\hat{\psi})$ na definição da estatística $\bar{\mathcal{W}}$ vemos que

$$\begin{aligned} \bar{L}_P(\bar{\psi}) - \bar{L}_P(\psi) &= L_P(\bar{\psi}) + M(\bar{\psi}) + L_P(\hat{\psi}) - L_P(\psi) - M(\psi) - L_P(\hat{\psi}) \\ &= L_P(\hat{\psi}) - L_P(\psi) + L_P(\bar{\psi}) + M(\bar{\psi}) - L_P(\hat{\psi}) - M(\psi), \end{aligned}$$

o que nos permite escrever

$$\bar{\mathcal{W}}(\psi) = \mathcal{W}(\psi) + 2\{L_P(\bar{\psi}) - L_P(\hat{\psi}) + M(\bar{\psi}) - M(\psi)\}.$$

Em problemas regulares, $\bar{\mathcal{W}}(\psi)$ tem distribuição nula $\chi^2(p)$ com erro de ordem $O(n^{-1})$ e, segundo foi provado em DiCiccio & Stern (1994), a estatística

$$\bar{\mathcal{W}}^*(\psi) = \bar{\mathcal{W}}(\psi) / \{1 + \bar{\xi}(\vartheta)/p\},$$

tem distribuição $\chi^2(p)$ sob H_0 com erro de ordem $O(n^{-2})$ onde $\bar{\xi}(\vartheta)$ é definido como

$$\bar{\xi}(\vartheta) = \xi(\vartheta) + \xi^\dagger(\vartheta), \tag{2.15}$$

com $\xi(\vartheta)$ representando o fator de correção de Bartlett obtido da verossimilhança usual e

$$\xi^\dagger(\vartheta) = 2E_{\vartheta}\{L_P(\bar{\psi}) - L_P(\widehat{\psi}) + M(\bar{\psi}) - M(\psi)\},$$

o acréscimo devido à consideração do ajuste.

De maneira geral, DiCiccio & Stern (1994) mostraram que

$$\xi^\dagger(\vartheta) = \lambda^{ab} (\Psi_{ab} - \rho_a \rho_b) + 2\lambda^{br} \lambda^{st} \rho_b \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right) - 2\lambda^{as} \rho_{a/s} + O(n^{-3/2}), \quad (2.16)$$

onde $\rho_{a/s} = \partial \rho_a / \partial \vartheta^s$ foi dado em (2.9), Ψ_{ab} sendo dado em (2.10).

Nosso objetivo é obter a expressão de $\xi^\dagger(\vartheta)$ quando o ajuste considerado é o de Stern e para isso identificaremos os termos necessários em (2.16).

Substituindo B^{ab} em (2.10), escrevemos

$$\begin{aligned} \Psi_{ab} = & \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \left(\lambda_{rst} \lambda_{uv/w} - \frac{1}{2} \lambda_{rst} \lambda_{uvw} \right) \\ & - \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \lambda^{tv} \nu^{uw} \left(\lambda_{rt/s} \lambda_{uv/w} - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} \right) \\ & - \sigma_{ad} \sigma_{bc} \lambda^{cr} \lambda^{ds} \lambda^{tv} \nu^{uw} \left(\lambda_{rt/s} \lambda_{uv/w} - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} \right) \\ & - \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \left(\lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) \\ & - \sigma_{ad} \sigma_{bc} \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \left(\lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) \\ & + \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\lambda_{rt/su} - \frac{1}{2} \lambda_{rtu/s} \right) + \sigma_{ad} \sigma_{bc} \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\lambda_{rt/su} - \frac{1}{2} \lambda_{rtu/s} \right) \\ & + \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \lambda^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rst} \lambda_{uvw} - \lambda_{rst} \lambda_{uv/w} \right. \\ & \quad \left. + \lambda_{rt/s} \lambda_{uv/w} + \lambda_{st/r} \lambda_{uv/w} - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} - \frac{1}{2} \lambda_{st/r} \lambda_{uvw} \right) \\ & + \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} + \lambda_{rt/u} \lambda_{sw/v} \right. \\ & \quad \left. + \lambda_{rt/u} \lambda_{vw/s} + \lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{rtu} \lambda_{vw/s} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) \\ & - \sigma_{ac} \sigma_{bd} \lambda^{cr} \lambda^{ds} \nu^{tu} \left(\frac{1}{2} \lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su} + \lambda_{st/ru} - \frac{1}{2} \lambda_{rtu/s} - \frac{1}{2} \lambda_{stu/r} \right), \end{aligned}$$

que se reduz a

$$\begin{aligned}\Psi_{ab} = & -\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw}\right. \\ & + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} \\ & \left. + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{rt/u}\lambda_{vw/s} + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s}\right) \\ & - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u}\right).\end{aligned}$$

Por outro lado,

$$\rho_a\rho_b = \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw}\right)$$

e, lembrando que $\tau^{tv} = \lambda^{tv} - \nu^{tv}$, temos

$$\begin{aligned}2\lambda^{br}\lambda^{st}\rho_b\left(\lambda_{rs/t} - \frac{1}{2}\lambda_{rst}\right) = & 2\sigma_{bd}\lambda^{br}\lambda^{dv}\lambda^{st}\nu^{uw}\left(\lambda_{rs/t}\lambda_{uv/w} - \frac{1}{2}\lambda_{rs/t}\lambda_{uvw}\right. \\ & \left. - \frac{1}{2}\lambda_{rst}\lambda_{uv/w} + \frac{1}{4}\lambda_{rst}\lambda_{uvw}\right).\end{aligned}$$

Também obtemos que

$$\begin{aligned}2\lambda^{as}\rho_{a/s} = & 2\sigma_{ac}\lambda^{as}\lambda^{cr}\nu^{tv}\nu^{uw}\left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \lambda_{tu/s}\lambda_{rv/w}\right) \\ & - 2\sigma_{ac}\lambda^{as}\lambda^{cr}\nu^{tu}\left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su}\right)\end{aligned}$$

e

$$\begin{aligned}\Psi_{ab} - \rho_a\rho_b = & -\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw}\right. \\ & + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} \\ & + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{rt/u}\lambda_{vw/s} + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} \\ & + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw}\left.)\right) \\ & - \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u}\right).\end{aligned}$$

Agrupando convenientemente os termos, vemos que

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\sigma_{ac}\sigma_{bd}\lambda^{ab}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sv/v} - \lambda_{rt/u}\lambda_{vw/s} \\
& \left. + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& - \sigma_{ac}\sigma_{bd}\lambda^{ab}\lambda^{cr}\lambda^{ds}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} \right) \\
& + 2\sigma_{bd}\lambda^{br}\lambda^{dv}\lambda^{st}\nu^{uw} \left(\lambda_{rs/t}\lambda_{uv/w} - \frac{1}{2}\lambda_{rs/t}\lambda_{uvw} - \frac{1}{2}\lambda_{rst}\lambda_{uv/w} + \frac{1}{4}\lambda_{rst}\lambda_{uvw} \right) \\
& - 2\sigma_{ac}\lambda^{as}\lambda^{cr}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{rvw}\lambda_{tu/s} - \lambda_{rv/w}\lambda_{tu/s} \right) \\
& + 2\sigma_{ac}\lambda^{as}\lambda^{cr}\nu^{tu} \left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su} \right).
\end{aligned}$$

A matriz de funções de cumulantes σ é difícil de ser obtida. Por esta razão utilizaremos a relação $\tau^{rs} = \sigma_{ab}\lambda^{ra}\lambda^{sb}$ com o intuito de simplificar a expressão anterior.

Desta forma,

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\sigma_{bd}\tau^{br}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sv/v} - \lambda_{rt/u}\lambda_{vw/s} \\
& \left. + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& - \sigma_{bd}\tau^{br}\lambda^{ds}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} \right) \\
& + 2\tau^{rv}\lambda^{st}\nu^{uw} \left(\lambda_{rs/t}\lambda_{uv/w} - \frac{1}{2}\lambda_{rs/t}\lambda_{uvw} - \frac{1}{2}\lambda_{rst}\lambda_{uv/w} + \frac{1}{4}\lambda_{rst}\lambda_{uvw} \right) \\
& - 2\tau^{rs}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{rvw}\lambda_{tu/s} - \lambda_{rv/w}\lambda_{tu/s} \right) \\
& + 2\tau^{rs}\nu^{tu} \left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su} \right).
\end{aligned}$$

Utilizando agora a relação $\tau^{rs} = \lambda^{rs} - \nu^{rs}$ vemos que $\xi^\dagger(\vartheta)$ pode ser escrita como

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\sigma_{bd}\lambda^{br}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{rt/u}\lambda_{vw/s} \\
& \left. + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& + \sigma_{bd}\nu^{br}\lambda^{ds}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{rt/u}\lambda_{vw/s} \\
& \left. + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& - \sigma_{bd}\lambda^{br}\lambda^{ds}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} \right) + \sigma_{bd}\nu^{br}\lambda^{ds}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} \right) \\
& + 2\tau^{rv}\lambda^{st}\nu^{uw} \left(\lambda_{rs/t}\lambda_{uv/w} - \frac{1}{2}\lambda_{rs/t}\lambda_{uvw} - \frac{1}{2}\lambda_{rst}\lambda_{uv/w} + \frac{1}{4}\lambda_{rst}\lambda_{uvw} \right) \\
& - 2\tau^{rs}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} - \lambda_{tu/s}\lambda_{rv/w} \right) \\
& + 2\tau^{rs}\nu^{tu} \left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su} \right).
\end{aligned}$$

Pela forma da matriz ν , os termos da expressão anterior onde aparece ν^{br} são nulos, e, portanto,

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\tau^{rs}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} + \lambda_{tu/s}\lambda_{rv/w} - \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{rt/u}\lambda_{vw/s} \\
& \left. + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} \right) \\
& - \tau^{rs}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} \right) \\
& + 2\tau^{rs}\lambda^{tv}\nu^{uw} \left(\frac{1}{4}\lambda_{rtv}\lambda_{suw} - \frac{1}{2}\lambda_{rv/t}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \lambda_{rv/t}\lambda_{su/w} \right) \\
& - 2\tau^{rs}\nu^{tv}\nu^{uw} \left(\frac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \lambda_{rt/s}\lambda_{uv/w} - \lambda_{tu/s}\lambda_{rv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right) \\
& + 2\tau^{rs}\nu^{tu} \left(\frac{1}{2}\lambda_{rtu/s} - \lambda_{rt/su} \right).
\end{aligned}$$

Observamos que os coeficientes do primeiro e do quarto termo, na expressão anterior coincidem, assim como os coeficientes do terceiro e do quinto termo. Concluimos então que

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\tau^{rs}\nu^{tv}\nu^{uw} \left(\lambda_{rst}\lambda_{uv/w} - \frac{1}{2}\lambda_{rst}\lambda_{uvw} - \lambda_{tu/s}\lambda_{rv/w} + \frac{1}{2}\lambda_{tu/s}\lambda_{rvw} \right. \\
& - \frac{1}{2}\lambda_{rtu}\lambda_{svw} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sv/v} \\
& - \lambda_{rt/u}\lambda_{vw/s} + \frac{1}{2}\lambda_{rtu}\lambda_{vw/s} + \lambda_{rt/v}\lambda_{su/w} - \frac{1}{2}\lambda_{rt/v}\lambda_{suw} \\
& \left. - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \frac{1}{4}\lambda_{rtv}\lambda_{suw} + \lambda_{rt/s}\lambda_{uvw} - 2\lambda_{rt/s}\lambda_{uv/w} \right) \\
& + 2\tau^{rs}\lambda^{tv}\nu^{uw} \left(\frac{1}{4}\lambda_{rtv}\lambda_{suw} - \frac{1}{2}\lambda_{rv/t}\lambda_{suw} - \frac{1}{2}\lambda_{rtv}\lambda_{su/w} + \lambda_{rv/t}\lambda_{su/w} \right) \\
& - \tau^{rs}\nu^{tu} \left(\frac{1}{2}\lambda_{rstu} - \lambda_{rst/u} - \lambda_{rtu/s} + 2\lambda_{rt/su} \right), \tag{2.17}
\end{aligned}$$

constituindo este o termo adicional da correção de Bartlett para a estatística da razão de verossimilhanças perfilada ajustada de Stern.

2.2.1 Exemplo: modelo de regressão linear normal

Na Seção anterior foi obtida a expressão geral para o termo adicional ξ^\dagger da correção de Bartlett para o teste da razão de verossimilhanças ajustado de Stern. Aqui nos dedicaremos a obter a expressão de ξ^\dagger quando o modelo é o de regressão linear normal múltipla, definido na Seção 2.1.3. Não utilizaremos o resultado em (2.17), mas sim desenvolveremos uma expressão para o valor esperado de $\overline{\mathcal{W}}(\phi)$. Posteriormente, na Seção 3.2.1, obteremos ξ^\dagger para os modelos lineares generalizados a partir de (2.17) e obteremos o resultado correspondente para o modelo de regressão linear normal como um caso particular.

A estatística da razão de verossimilhanças neste modelo é

$$\mathcal{W}(\phi) = n \log n - n \log S^2 - n - n \log \phi + \phi S^2,$$

onde S^2 foi definido na Seção 2.1.3.

Nosso objetivo agora é obter a esperança de $\mathcal{W}(\phi)$. Como $\phi S^2 \sim \chi^2(n - q)$, temos

$E\{S^2\} = \phi^{-1}(n - q)$, $\text{var}\{S^2\} = 2\phi^{-2}(n - q)$ e obtém-se

$$E\{\mathcal{W}(\phi)\} = n \log n - nE\{\log S^2\} - n - n \log \phi + (n - q). \quad (2.18)$$

A esperança de $\mathcal{W}(\phi)$, até ordem $O(n^{-1})$, nos permitirá definir a correção necessária para diminuir o vício da esperança da estatística da razão de verossimilhanças. Com esse objetivo, consideremos a seguinte expansão em série de Taylor no ponto $n\phi^{-1}$:

$$\begin{aligned} \log S^2 &= \log n\phi^{-1} + \frac{1}{n\phi^{-1}}(S^2 - n\phi^{-1}) + \frac{1}{2n^2\phi^{-2}}(S^2 - n\phi^{-1})^2 + \\ &\quad \frac{1}{3n^3\phi^{-3}}(S^2 - n\phi^{-1})^3 - \frac{1}{4n^4\phi^{-4}}(S^2 - n\phi^{-1})^4 + \dots \end{aligned}$$

Utilizando a função geradora de momentos da distribuição qui-quadrado, obtemos que

$$\begin{aligned} \phi^2 E\{(S^2)^2\} &= 2(n - q) + (n - q)^2, \\ \phi^3 E\{(S^2)^3\} &= 8(n - q) + 6(n - q)^2 + (n - q)^3, \\ \phi^4 E\{(S^2)^4\} &= 48(n - q) + 44(n - q)^2 + 12(n - q)^2 + (n - q)^4 \end{aligned}$$

e

$$\begin{aligned} \phi E\{S^2 - n\phi^{-1}\} &= -q, \\ \phi^2 E\{(S^2 - n\phi^{-1})^2\} &= 2(n - q) + q^2, \\ \phi^3 E\{(S^2 - n\phi^{-1})^3\} &= 8n - 6nq, \\ \phi^4 E\{(S^2 - n\phi^{-1})^4\} &= 12n^2 + O(n^{-3}). \end{aligned}$$

Assim,

$$E\{\log S^2\} = \log n - \log \phi - \frac{q}{n} - \frac{n - q}{n^2} - \frac{q^2}{2n^2} + \frac{8 - 6q}{3n^2} - \frac{3}{n^2} + O(n^{-3}),$$

e substituindo em (2.18) obtemos

$$\begin{aligned} E\{\mathcal{W}(\phi)\} &= n \log n + n \log \phi - n \log n + q + \frac{n - q}{n} + \frac{q^2}{2n} - \frac{8 - 6q}{3n} \\ &\quad + \frac{3}{n} - n - n \log \phi + n - q + O(n^{-2}), \end{aligned}$$

que se reduz a

$$E\{\mathcal{W}(\phi)\} = 1 + \frac{3q^2 + 6q + 2}{6n} + O(n^{-2}).$$

Este resultado permite-nos definir a estatística da razão de verossimilhanças corrigida como

$$\mathcal{W}^*(\phi) = \left(1 + \frac{3q^2 + 6q + 2}{6n}\right)^{-1} \mathcal{W}(\phi),$$

de onde vem

$$E\{\mathcal{W}^*(\phi)\} = 1 + O(n^{-2}).$$

Desta forma encontramos

$$\xi(\phi) = \frac{3q^2 + 6q + 2}{6n}. \quad (2.19)$$

Procederemos agora à obtenção da expressão de ξ^\dagger neste modelo. Da função de verossimilhança perfilada ajustada de Stern $\bar{L}_P(\phi)$ dada na Seção 2.1.3 temos

$$\bar{L}_P(\bar{\phi}) = -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \bar{\phi} - \frac{\bar{\phi}}{2} S^2 + \frac{q}{2} \left(1 - \frac{\bar{\phi}}{\hat{\phi}}\right) + \frac{q}{4} \left(1 - \frac{\bar{\phi}}{\hat{\phi}}\right)^2.$$

e

$$\begin{aligned} \bar{\mathcal{W}}(\phi) = 2 \left\{ -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log \bar{\phi} - \frac{\bar{\phi}}{2} S^2 + \frac{q}{2} \left(1 - \frac{\bar{\phi}}{\hat{\phi}}\right) + \frac{q}{4} \left(1 - \frac{\bar{\phi}}{\hat{\phi}}\right)^2 \right. \\ \left. + \frac{n}{2} \log(2\pi) - \frac{n}{2} \log \phi + \frac{\phi}{2} S^2 - \frac{q}{2} \left(1 - \frac{\phi}{\hat{\phi}}\right) - \frac{q}{4} \left(1 - \frac{\phi}{\hat{\phi}}\right)^2 \right\}. \end{aligned} \quad (2.20)$$

Lembrando que

$$\bar{\phi} = \hat{\phi} \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}}\right),$$

obtemos

$$\log \bar{\phi} = \log \hat{\phi} + \log \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}}\right),$$

e dado que $\hat{\phi} = n/S^2$ temos $\log \hat{\phi} = \log n - \log S^2$. Também

$$\begin{aligned} \frac{\bar{\phi}}{2} S^2 &= \frac{\hat{\phi}}{2} \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}}\right) S^2 = \frac{n}{2S^2} \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}}\right) S^2 \\ &= \frac{n}{2} + \frac{n^2}{4q} - \frac{n}{2} \sqrt{1 + \frac{n^2}{4q^2}}, \end{aligned}$$

$$1 - \frac{\bar{\phi}}{\hat{\phi}} = -\frac{n}{2q} + \sqrt{1 + \frac{n^2}{4q^2}} \quad \text{e} \quad 1 - \frac{\phi}{\hat{\phi}} = 1 - \frac{\phi S^2}{n} = (n\phi^{-1} - S^2) \frac{\phi}{n}.$$

Substituindo estes resultados em (2.20) vemos que

$$\begin{aligned} \overline{\mathcal{W}}(\phi) = & 2 \left\{ \frac{n}{2} \log n - \frac{n}{2} \log S^2 + \frac{n}{2} \log \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}} \right) - \frac{n}{2} \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}} \right) \right. \\ & + \frac{q}{2} \left(-\frac{n}{2q} + \sqrt{1 + \frac{n^2}{4q^2}} \right) + \frac{q}{4} \left(-\frac{n}{2q} + \sqrt{1 + \frac{n^2}{4q^2}} \right)^2 - \frac{n}{2} \log \phi + \frac{\phi}{2} S^2 \\ & \left. - \frac{q\phi}{2n} (n\phi^{-1} - S^2) - \frac{q\phi^2}{4n^2} (n\phi^{-1} - S^2)^2 \right\}. \end{aligned} \quad (2.21)$$

Para escrever de maneira mais simples a esperança da estatística da razão de verossimilhanças ajustada de Stern acima utilizaremos duas expansões em série de Taylor, a primeira delas é

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \frac{5x^4}{128} + \frac{7x^5}{256} - \frac{21x^6}{1024} + O(x^7),$$

que será aplicada na função $\sqrt{1+n^2/4q^2}$. Como

$$\sqrt{1 + \frac{n^2}{4q^2}} = \sqrt{\left(\frac{n}{2q}\right)^2 \left(1 + \frac{4q^2}{n^2}\right)} = \frac{n}{2q} \sqrt{1 + \frac{4q^2}{n^2}},$$

obtemos

$$\begin{aligned} \sqrt{1 + \frac{n^2}{4q^2}} &= \frac{n}{2q} \left(1 + \frac{2q^2}{n^2} - \frac{2q^4}{n^4} + \frac{4q^6}{n^6} + O(n^{-8}) \right) \\ &= \frac{n}{2q} + \frac{q}{n} - \frac{q^3}{n^3} + O(n^{-5}). \end{aligned}$$

A outra expansão em série de Taylor a ser utilizada é

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Observando que

$$\log \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}} \right) = \log \left(1 + \frac{n}{2q} \left(1 - \sqrt{1 + \frac{4q^2}{n^2}} \right) \right),$$

obtemos

$$\begin{aligned} \log \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}} \right) &= \frac{n}{2q} \left(1 - \sqrt{1 + \frac{4q^2}{n^2}} \right) - \frac{n^2}{8q^2} \left(1 - \sqrt{1 + \frac{4q^2}{n^2}} \right)^2 + \dots \\ &= -\frac{q}{n} - \frac{q^2}{2n^2} + \frac{q^3}{n^3} + O(n^{-4}), \end{aligned}$$

e, portanto,

$$\frac{n}{2} \log \left(1 + \frac{n}{2q} - \sqrt{1 + \frac{n^2}{4q^2}} \right) = -\frac{q}{2} - \frac{q^2}{4n} + \frac{q^3}{2n^2} + O(n^{-3}).$$

Também,

$$\begin{aligned} \frac{n}{2} \sqrt{1 + \frac{n^2}{4q^2}} &= \frac{n^2}{4q} + \frac{q}{2} - \frac{q^3}{2n^2} + O(n^{-4}), \\ \frac{q}{2} \sqrt{1 + \frac{n^2}{4q^2}} &= \frac{n}{4} + \frac{q^2}{2n} + O(n^{-3}) \end{aligned}$$

e

$$\frac{q}{4} \left(-\frac{n}{2q} + \sqrt{1 + \frac{n^2}{4q^2}} \right)^2 = \frac{q^3}{4n^2} + O(n^{-4}).$$

Substituindo as expressões já obtidas de $E\{n\phi^{-1} - S^2\}$, $E\{(n\phi^{-1} - S^2)^2\}$ e $E\{\log S^2\}$ em (2.21), obtemos

$$E\{\overline{\mathcal{W}}(\phi)\} = 1 + \frac{1}{3n} + O(n^{-2}).$$

Podemos definir então a correção de Bartlett para a estatística da razão de verossimilhanças ajustada de Stern como

$$\overline{\xi}(\phi) = \frac{1}{3n}, \tag{2.22}$$

e, lembrando que $\overline{\xi}(\phi) = \xi(\phi) + \xi^\dagger(\phi)$ com $\xi(\phi)$ obtido em (2.19), temos

$$\xi^\dagger(\phi) = -\frac{q^2}{2n} - \frac{q}{n}$$

e concluimos que a estatística $\overline{\mathcal{W}}(\phi)$ corrigida é da forma

$$\overline{\mathcal{W}}^*(\phi) = \left(1 + \frac{1}{3n} \right)^{-1} \overline{\mathcal{W}}(\phi).$$

Posteriormente, na Seção 3.2.1, observaremos que a expressão (2.22) é obtida também através de uma situação particular nos modelos lineares generalizados.

Capítulo 3

Ajuste para a verossimilhança perfilada em modelos lineares generalizados

Desenvolvemos no capítulo anterior resultados gerais acerca do ajuste de Stern para a verossimilhança perfilada e da correção de Bartlett para o teste da razão de verossimilhanças obtido a partir de tal verossimilhança ajustada. Neste capítulo, aplicaremos os resultados aos modelos lineares generalizados.

Duas situações serão consideradas: na primeira o parâmetro de interesse é o de dispersão que, nesse caso, é ortogonal ao vetor de parâmetros de perturbação. Na segunda situação, o vetor de parâmetros de interesse é parte do vetor de parâmetros de regressão e a ortogonalidade não ocorre. Obteremos, em cada caso, a verossimilhança perfilada ajustada e, posteriormente, a correção de Bartlett para o teste de interesse.

Nossos resultados são avaliados através de simulações de Monte Carlo. No primeiro estudo de simulação, consideraremos de interesse o parâmetro de dispersão na regressão linear normal múltipla; a seguir, consideraremos os parâmetros de regressão como de interesse e ϕ como o parâmetro de perturbação. O modelo de regressão gama será considerado posteriormente.

Interessa-nos, nessa situação, fazer inferência sobre parte do vetor β .

Em todos os exemplos de simulação comparamos o comportamento do teste da razão de verossimilhanças, do teste da razão de verossimilhanças ajustado e de suas correspondentes versões corrigidas por correções de Bartlett.

3.1 Função de verossimilhança perfilada ajustada nos modelos lineares generalizados

Os modelos lineares generalizados foram introduzidos na Seção 1.3 e, tanto aqui quanto na Seção 3.2, não faremos restrição quanto à função de ligação. Já nos exemplos das Seções 3.2.3 e 3.2.4 consideraremos somente a ligação canônica.

3.1.1 Dispersão como parâmetro de interesse

Se nosso interesse é realizar inferência acerca do parâmetro de dispersão, podemos utilizar a verossimilhança perfilada com este objetivo e para isso consideraremos $\psi = \phi$ como o parâmetro de interesse e $\zeta = \beta$, o vetor de parâmetros de perturbação, de dimensão q . Estes parâmetros são ortogonais no sentido de Cox & Reid (1987), ou seja, $E\{-\partial^2 L / \partial \phi \partial \beta_i\} = 0$, para $i = 1, \dots, q$.

Desenvolveremos a seguir as expressões das funções $\Lambda(\vartheta)$ e $\Theta(\vartheta)$, dadas nos Teoremas 2.1.1 e 2.1.2, correspondentes aos termos do ajuste de Stern para a verossimilhança perfilada nesta situação. Os cumulantes de derivadas do logaritmo da função de verossimilhança necessários à obtenção de tais quantidades são dados no Apêndice B.

Do Teorema 2.1.1 e dado que $a = \phi$, temos

$$\Lambda(\vartheta) = \lambda^{\phi\phi} \lambda^{ij} \left(\lambda_{\phi i/j} - \frac{1}{2} \lambda_{\phi ij} \right) L_{\phi}, \quad (3.1)$$

onde L_{ϕ} é a função escore para o parâmetro ϕ , ou seja,

$$L_{\phi} = \sum_{l=1}^n \{y_l \theta^l - b(\theta^l) + c(y_l) + d'_1(\phi)\}.$$

Da ortogonalidade dos vetores ϕ e β obtemos que $\lambda_{\phi i/j} = 0$, já que $\lambda_{\phi i} = 0$, e utilizando a matriz de pesos W , definida na Seção 1.3.2, podemos escrever

$$\begin{aligned}\lambda^{ij}\lambda_{\phi ij} &= \sum_{i,j=1}^q \lambda^{ij} \sum_{m=1}^n -w_m x_{mi} x_{mj} = - \sum_{m=1}^n w_m \left(\sum_{i,j=1}^q \lambda^{ij} x_{mi} x_{mj} \right) \\ &= \phi^{-1} \sum_{m=1}^n w_m z_{mm},\end{aligned}$$

onde $z_{mm} = -\phi \sum_{i,j=1}^q \lambda^{ij} x_{mi} x_{mj}$ é o m -ésimo elemento da diagonal principal da matriz Z , definida como $Z = X(X^\top W X)^{-1} X^\top$. Então,

$$\begin{aligned}\lambda^{ij}\lambda_{\phi ij} &= \phi^{-1} \text{tr}\{WZ\} = \phi^{-1} \text{tr}\{WX(X^\top W X)^{-1} X^\top\} \\ &= \phi^{-1} \text{tr}\{X^\top W X (X^\top W X)^{-1}\} = \frac{q}{\phi}.\end{aligned}$$

Como $\lambda^{\phi\phi} = 1/nd_1''(\phi)$, obtemos de (3.1) que

$$\tilde{\Lambda}(\phi) = -\frac{q}{2n\phi d_1''(\phi)} \tilde{L}_\phi, \quad (3.2)$$

onde \tilde{L}_ϕ iguala-se a L_ϕ avaliada em $\hat{\beta}$, o estimador de máxima verossimilhança de β .

Obtivemos assim a expressão para o termo de correção do vício da função escore perfilada. Desta forma, a função de verossimilhança perfilada ajustada $L_P(\phi) + \tilde{\Lambda}(\phi)$ possui vício de função escore de ordem $O(n^{-1})$, onde

$$L_P(\phi) = \phi \sum_{l=1}^n \left\{ y_l \hat{\theta}^l - b(\hat{\theta}^l) + c(y_l) \right\} + nd_1(\phi) + \sum_{l=1}^n d_2(y_l),$$

é a função de verossimilhança perfilada para o parâmetro de dispersão nos modelos lineares generalizados, na qual $\hat{\theta} = \theta(\hat{\beta})$.

Desenvolveremos agora o termo adicional para a correção do vício da variância da função escore perfilada, a qual constitui a principal proposta no ajuste de Stern. Trabalharemos aqui, e nos seguintes desenvolvimentos, com a expressão da função B^{ab} obtida por nós (ver Teorema 2.1.2 e equação (2.6)).

Avaliando a expressão (2.6) quando $a = b = \phi$, vemos que

$$\begin{aligned}
B^{\phi\phi} = & \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ik} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi ij} \lambda_{\phi kl} - \lambda_{\phi ij} \lambda_{\phi k/l} - \frac{1}{2} \lambda_{\phi ij} \lambda_{kl/\phi} - \lambda_{\phi i/j} \lambda_{\phi kl} + \lambda_{\phi i/j} \lambda_{kl/\phi} \right. \\
& \left. + \lambda_{\phi i/j} \lambda_{\phi l/k} + \lambda_{ij/\phi} \lambda_{\phi k/l} - \frac{1}{2} \lambda_{ij/\phi} \lambda_{\phi kl} \right) \\
& + \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi\phi\phi} \lambda_{\phi jl} - \lambda_{\phi\phi\phi} \lambda_{\phi j/l} + \lambda_{\phi\phi/\phi} \lambda_{\phi j/l} + \frac{1}{2} \lambda_{\phi\phi/\phi} \lambda_{\phi jl} \right. \\
& \left. + \lambda_{\phi\phi/\phi} \lambda_{\phi j/l} - \frac{1}{2} \lambda_{\phi\phi/\phi} \lambda_{\phi jl} \right) \\
& + \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ik} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi\phi i} \lambda_{jkl} - \lambda_{\phi\phi i} \lambda_{jk/l} + \lambda_{\phi i/\phi} \lambda_{jk/l} - \frac{1}{2} \lambda_{\phi i/\phi} \lambda_{jkl} \right. \\
& \left. + \lambda_{\phi i/\phi} \lambda_{jk/l} - \frac{1}{2} \lambda_{\phi i/\phi} \lambda_{jkl} \right) \\
& + \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ij} \left(\frac{1}{2} \lambda_{\phi\phi ij} - \lambda_{\phi\phi i/j} - \frac{1}{2} \lambda_{\phi ij/\phi} - \frac{1}{2} \lambda_{\phi ij/\phi} + \lambda_{\phi i/\phi j} + \lambda_{\phi i/\phi j} \right).
\end{aligned}$$

Os cumulantes $\lambda_{\phi k/l}$, $\lambda_{\phi i/j}$, $\lambda_{\phi l/k}$, $\lambda_{\phi k/l}$, $\lambda_{\phi j/l}$, $\lambda_{\phi\phi i}$, $\lambda_{\phi i/\phi}$, $\lambda_{\phi\phi ij}$, $\lambda_{\phi\phi i/j}$, $\lambda_{\phi ij/\phi}$ e $\lambda_{\phi i/\phi j}$ são nulos pela ortogonalidade entre os parâmetros ϕ e β . Desta forma, a expressão anterior se reduz a

$$\begin{aligned}
B^{\phi\phi} = & \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ik} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi ij} \lambda_{\phi kl} - \frac{1}{2} \lambda_{\phi ij} \lambda_{kl/\phi} - \frac{1}{2} \lambda_{ij/\phi} \lambda_{\phi kl} \right) \\
& + \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi\phi\phi} \lambda_{\phi jl} + \frac{1}{2} \lambda_{\phi\phi/\phi} \lambda_{\phi jl} - \frac{1}{2} \lambda_{\phi\phi/\phi} \lambda_{\phi jl} \right).
\end{aligned}$$

Uma vez que $\lambda_{kl/\phi} = \lambda_{\phi kl}$, $\lambda_{ij/\phi} = \lambda_{\phi ij}$ e $\lambda_{\phi\phi/\phi} = \lambda_{\phi\phi\phi}$, obtemos

$$\begin{aligned}
B^{\phi\phi} = & \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ik} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi ij} \lambda_{\phi kl} - \frac{1}{2} \lambda_{\phi ij} \lambda_{\phi kl} - \frac{1}{2} \lambda_{\phi ij} \lambda_{\phi kl} \right) \\
& + \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{jl} \left(\frac{1}{2} \lambda_{\phi\phi\phi} \lambda_{\phi jl} + \frac{1}{2} \lambda_{\phi\phi\phi} \lambda_{\phi jl} - \frac{1}{2} \lambda_{\phi\phi\phi} \lambda_{\phi jl} \right),
\end{aligned}$$

que se reduz a

$$B^{\phi\phi} = -\frac{1}{2} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{jl} \lambda_{\phi\phi\phi} \lambda_{\phi jl} - \frac{1}{2} \lambda^{\phi\phi} \lambda^{\phi\phi} \lambda^{ik} \lambda^{jl} \lambda_{\phi ij} \lambda_{\phi kl}.$$

Como $\lambda_{\phi\phi\phi} = nd_1'''(\phi)$ e $\lambda^{jl}\lambda_{\phi jl} = q/\phi$, temos

$$\begin{aligned}\lambda^{ik}\lambda^{jl}\lambda_{\phi ij}\lambda_{\phi kl} &= \sum_{i,k,j,l=1}^q \lambda^{ik}\lambda^{jl} \sum_{m_1=1}^n -w_{m_1}x_{m_1i}x_{m_1j} \sum_{m_2=1}^n -w_{m_2}x_{m_2k}x_{m_2l} \\ &= \sum_{m_1,m_2=1}^n w_{m_1} \sum_{i,k=1}^q \lambda^{ik}x_{m_1i}x_{m_2k} \sum_{j,l=1}^q \lambda^{jl}x_{m_1j}x_{m_2l}w_{m_2} \\ &= \phi^{-2}1^\top W(Z \otimes Z)W1,\end{aligned}$$

já que tanto $-\phi \sum_{i,k=1}^q \lambda^{ik}x_{m_1i}x_{m_2k}$ quanto $-\phi \sum_{j,l=1}^q \lambda^{jl}x_{m_1j}x_{m_2l}$ são iguais a $z_{m_1m_2}$, o elemento (m_1, m_2) da matriz Z . Na expressão acima foi utilizado o produto direto de matrizes, simbolizado por \otimes , também conhecido como produto de Hadamard (Rao, 1973).

Utilizando o resultado (C.2) do Anexo C e observando que a dimensão da matriz $X^\top WX$ é q , concluímos que $\lambda^{ik}\lambda^{jl}\lambda_{\phi ij}\lambda_{\phi kl} = q/\phi^2$. Obtemos, portanto,

$$B^{\phi\phi} = -\frac{1}{2} \frac{nd_1'''(\phi)}{(nd_1''(\phi))^3} \frac{q}{\phi} - \frac{1}{2} \frac{1}{(nd_1''(\phi))^2} \frac{q}{\phi^2},$$

que pode ser escrito como

$$B^{\phi\phi} = -\frac{q}{2\phi} \left(\frac{1}{nd_1''(\phi)} \right)^2 \left(\frac{d_1'''(\phi)}{d_1''(\phi)} + \frac{1}{\phi} \right).$$

Conclui-se, da definição de $\Theta(\vartheta)$ dada no Teorema 2.1.2, que

$$\tilde{\Theta}(\phi) = -\frac{q}{4\phi} \left(\frac{1}{nd_1''(\phi)} \right)^2 \left(\frac{d_1'''(\phi)}{d_1''(\phi)} + \frac{1}{\phi} \right) \tilde{L}_\phi^2. \quad (3.3)$$

Assim, o logaritmo da função de verossimilhança perfilada ajustada para o parâmetro de dispersão é obtido a partir das expressões (3.2) e (3.3) e é da forma

$$\bar{L}_P(\phi) = L_P(\phi) - \frac{q}{2n\phi d_1''(\phi)} \tilde{L}_\phi - \frac{1}{4} \frac{q}{\phi} \left(\frac{1}{nd_1''(\phi)} \right)^2 \left(\frac{d_1'''(\phi)}{d_1''(\phi)} + \frac{1}{\phi} \right) \tilde{L}_\phi^2. \quad (3.4)$$

3.1.2 Exemplo: modelo de regressão normal linear

Este exemplo já foi estudado na Seção 2.1.3 e aqui o consideraremos como um caso particular dos modelos lineares generalizados. Nosso objetivo é fazer uma verificação parcial de que o resultado obtido por nós para a função B^{ab} em (3.3) está correto.

A forma do logaritmo da função de verossimilhança foi obtida na seção referida, onde chegamos à função escore para ϕ como sendo

$$L_\phi = \frac{n}{2\phi} \left(1 - \frac{\phi}{\hat{\phi}} \right).$$

Nesta situação, $d_1(\phi) = \log(\phi)/2$ e $d_1''(\phi) = -1/2\phi^2$. Com estes resultados podemos escrever (3.2) como

$$\tilde{\Lambda}(\phi) = -\frac{q}{2n\phi \left(-\frac{1}{2\phi^2} \right)} \tilde{L}_\phi = \frac{q\phi}{n} \tilde{L}_\phi,$$

e substituindo a expressão da função escore para ϕ , obtemos

$$\tilde{\Lambda}(\phi) = \frac{q}{2} \left(1 - \frac{\phi}{\hat{\phi}} \right).$$

Agora, observando que $d_1''(\phi) = -1/2\phi^2$, $d_1'''(\phi) = 1/\phi^3$ e que $\tilde{L}_\phi = L_\phi$ temos de (3.3) que

$$\tilde{\Theta}(\phi) = \frac{q}{n^2} \tilde{L}_\phi^2 = \frac{q}{4} \left(1 - \frac{\phi}{\hat{\phi}} \right)^2,$$

e, então, o logaritmo da função de verossimilhança perfilada ajustada é

$$\bar{L}_P(\phi) = -\frac{n}{2} \log 2\pi + \frac{n}{2} \log \phi - \frac{n\phi}{2\hat{\phi}} + \frac{q}{2} \left(1 - \frac{\phi}{\hat{\phi}} \right) + \frac{q}{4} \left(1 - \frac{\phi}{\hat{\phi}} \right)^2.$$

Na Seção 2.1.3 desenvolvemos este mesmo exemplo só que, naquele momento, não fizemos uso das expressões das funções de ajuste $\Lambda(\vartheta)$ e $\Theta(\vartheta)$. Se compararmos a expressão do logaritmo da função de verossimilhança perfilada ajustada obtida em (2.14) e a expressão acima, concluiremos que elas coincidem. Esta é, portanto, uma verificação parcial de que nossos resultados estão corretos.

Para ilustrar o comportamento das funções de verossimilhança perfilada e perfilada ajustada simulamos cem amostras, com tamanho de amostra $n = 10$, do modelo de regressão $N(\mathbf{X}\beta, \phi^{-1}\mathbf{I})$, para o qual escolhemos $\beta = (1, 1, 1)^\top$, $\phi = 1$ e a matriz das covariáveis \mathbf{X} foi obtida através de amostras independentes da distribuição uniforme no intervalo $(0, 1)$. O resultado é apresentado na Figura 3.1.

Observamos desta figura que o ponto de máximo da função $\bar{L}_P(\phi)$ é mais próximo de um do que o ponto de máximo da função $L_P(\phi)$, e portanto, o estimador obtido da verossimilhança

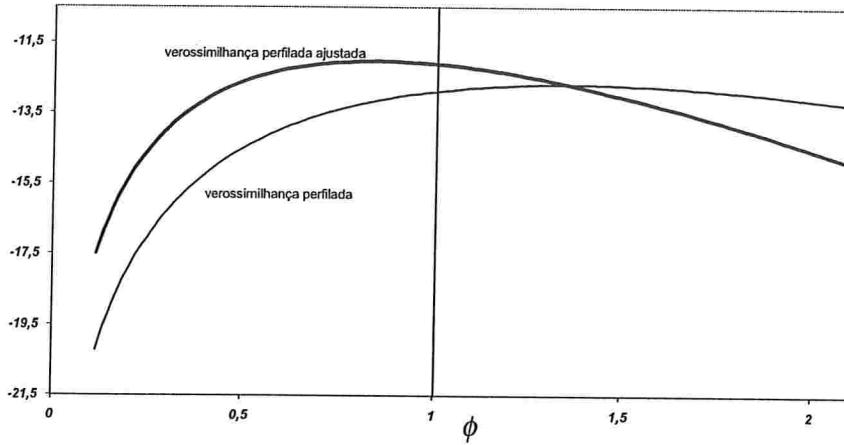


Figura 3.1: Comportamento médio, em cem amostras, do logaritmo das funções de verossimilhança perfilada e perfilada ajustada. Parâmetro de interesse $\phi = 1$, $n = 10$ e vetor de parâmetros de regressão $\beta = (1, 1, 1)^\top$.

perfilada ajustada tende a ser mais próximo do valor do parâmetro. Isto nos mostra uma das boas conseqüências de ajustar as verossimilhanças perfiladas e constitui um indicativo de que os nossos resultados estão corretos.

Provaremos a seguir que a função escore perfilada ajustada satisfaz as propriedades dela exigidas. Primeiro provaremos que

$$E \left\{ \frac{d}{d\phi} \bar{L}_P(\phi) \right\} = O(n^{-1}).$$

Neste sentido, encontremos agora as funções $d\bar{L}_P(\phi)/d\phi$ e $d^2\bar{L}_P(\phi)/d\phi^2$ e as esperanças correspondentes. Assim,

$$\frac{d}{d\phi} \bar{L}_P(\phi) = \frac{n}{2} \left(\frac{1}{\phi} - \frac{1}{\hat{\phi}} \right) - \frac{q}{2\hat{\phi}} - \frac{q}{2\hat{\phi}} \left(1 - \frac{\phi}{\hat{\phi}} \right)$$

e

$$E \left\{ \frac{d}{d\phi} \bar{L}_P(\phi) \right\} = \frac{n}{2\phi} - \frac{n}{2} E \left\{ \frac{1}{\hat{\phi}} \right\} - \frac{q}{2} E \left\{ \frac{1}{\hat{\phi}} \right\} - \frac{q}{2} E \left\{ \frac{1}{\hat{\phi}} \right\} + \frac{q\phi}{2} E \left\{ \frac{1}{\hat{\phi}^2} \right\}.$$

Como $\phi S^2 \sim \chi^2(n - q)$, temos $E\{S^2\} = (n - q)/\phi$ e, portanto, $E\{1/\hat{\phi}\} = E\{S^2/n\} = (n - q)/n\phi$. Lembrando que $E\{S^2\}^2 = \text{var}\{S^2\} + E^2\{S^2\}$, e que $\text{var}\{S^2\} = 2(n - q)/\phi^2$,

obtemos

$$\mathbb{E} \left\{ \frac{1}{\widehat{\phi}^2} \right\} = \mathbb{E} \left\{ \left(\frac{S^2}{n} \right)^2 \right\} = \frac{2(n-q)}{n^2\phi^2} + \frac{(n-q)^2}{n^2\phi^2}.$$

Desta forma, podemos escrever

$$\begin{aligned} \mathbb{E} \left\{ \frac{d}{d\phi} \bar{L}_P(\phi) \right\} &= \frac{n}{2\phi} - \frac{n-q}{2\phi} - \frac{q(n-q)}{2n\phi} - \frac{q(n-q)}{2\phi} + \frac{q\phi}{2} \left(\frac{2(n-q)}{n^2\phi^2} + \frac{(n-q)^2}{n^2\phi^2} \right) \\ &= \frac{q}{n\phi} - \frac{q^2}{n^2\phi} + \frac{q^3}{2n^2\phi}, \end{aligned}$$

que é uma função de ordem $O(n^{-1})$, como esperado.

Verifiquemos agora se também

$$\mathbb{E} \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} + \mathbb{E} \left\{ \frac{d^2}{d\phi^2} \bar{L}_P(\phi) \right\} = O(n^{-1}).$$

Com este objetivo, obtemos

$$\frac{d^2}{d\phi^2} \bar{L}_P(\phi) = \frac{d}{d\phi} \left(\frac{n}{2\phi} - \frac{n}{2\widehat{\phi}} - \frac{q}{\widehat{\phi}} + \frac{q\phi}{2\widehat{\phi}^2} \right) = -\frac{n}{2\phi^2} + \frac{q}{2\widehat{\phi}^2}$$

e

$$\begin{aligned} \mathbb{E} \left\{ \frac{d^2}{d\phi^2} \bar{L}_P(\phi) \right\} &= -\frac{n}{2\phi^2} + \mathbb{E} \left\{ \frac{q}{2\widehat{\phi}^2} \right\} = -\frac{n}{2\phi^2} + \frac{q}{2} \left(\frac{2(n-q)}{n^2\phi^2} + \frac{(n-q)^2}{n^2\phi^2} \right) \\ &= -\frac{n}{2\phi^2} + \frac{q}{n\phi^2} - \frac{q^2}{n^2\phi^2} + \frac{q}{2\phi^2} - \frac{q^2}{n\phi^2} + \frac{q^3}{2n^2\phi^2} \\ &= -\frac{n}{2\phi^2} + \frac{q}{2\phi^2} + O(n^{-1}). \end{aligned}$$

Também,

$$\begin{aligned} \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 &= \frac{n^2}{4} \left(\frac{1}{\phi} - \frac{1}{\widehat{\phi}} \right)^2 - \frac{nq}{\widehat{\phi}} \left(\frac{1}{\phi} - \frac{1}{\widehat{\phi}} \right) \left(1 - \frac{\phi}{2\widehat{\phi}} \right) + \frac{q^2}{\widehat{\phi}^2} \left(1 - \frac{\phi}{2\widehat{\phi}} \right)^2 \\ &= \frac{n^2}{4} \left(\frac{1}{\phi^2} - \frac{2}{\phi\widehat{\phi}} + \frac{1}{\widehat{\phi}^2} \right) - \frac{nq}{\widehat{\phi}} \left(\frac{1}{\phi} - \frac{1}{\widehat{\phi}} \right) + \frac{nq\phi}{2\widehat{\phi}^2} \left(\frac{1}{\phi} - \frac{1}{\widehat{\phi}} \right) + \frac{q^2}{\widehat{\phi}^2} \left(1 - \frac{\phi}{2\widehat{\phi}} \right)^2 \\ &= \frac{n^2}{4\phi^2} + \frac{n^2}{4\widehat{\phi}^2} - \frac{n^2}{2\phi\widehat{\phi}} - \frac{nq}{\phi\widehat{\phi}} + \frac{3nq}{2\widehat{\phi}^2} + \frac{q^2}{\widehat{\phi}^2} - \frac{nq\phi}{2\widehat{\phi}^3} - \frac{q^2\phi}{\widehat{\phi}^3} + \frac{q^2\phi^2}{4\widehat{\phi}^4}. \end{aligned}$$

Tomando esperança, temos que

$$\begin{aligned} \mathbb{E} \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} &= \frac{n^2}{4\phi^2} + \frac{n^2}{4} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^2} \right\} - \frac{n^2}{2\phi} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}} \right\} - \frac{nq}{\phi} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}} \right\} + \frac{3nq}{2} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^2} \right\} \\ &\quad + q^2 \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^2} \right\} - \frac{nq\phi}{2} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^3} \right\} - q^2\phi \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^3} \right\} + \frac{q^2\phi^2}{4} \mathbb{E} \left\{ \frac{1}{\widehat{\phi}^4} \right\}. \end{aligned} \tag{3.5}$$

Como

$$\begin{aligned} \mathbb{E} \left\{ \frac{1}{\widehat{\phi^3}} \right\} &= \mathbb{E} \left\{ \frac{S^6}{n^3} \right\} = \frac{1}{(n\phi)^3} (8(n-q) + 6(n-q)^2 + (n-q)^3), \\ \mathbb{E} \left\{ \frac{1}{\widehat{\phi^4}} \right\} &= \mathbb{E} \left\{ \frac{S^8}{n^4} \right\} \\ &= \frac{1}{(n\phi)^4} (48(n-q) + 44(n-q)^2 + 12(n-q)^3 + (n-q)^4), \end{aligned}$$

obtemos

$$\begin{aligned} \mathbb{E} \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} &= \frac{n^2}{4\phi^2} + \frac{n-q}{2\phi^2} + \frac{(n-q)^2}{4\phi^4} - \frac{n(n-q)}{2\phi^2} \\ &\quad - \frac{q(n-q)}{\phi^2} + \frac{3q(n-q)}{n\phi^2} + \frac{3q(n-q)^2}{2n\phi^2} + \frac{2q^2(n-q)}{n^2\phi^2} \\ &\quad + \frac{q^2(n-q)^2}{n^2\phi^2} - \frac{4q(n-q)}{n^2\phi^2} - \frac{3q(n-q)^2}{n^2\phi^2} - \frac{q(n-q)^3}{2n^2\phi^2} \\ &\quad - \frac{8q^2(n-q)}{n^3\phi^2} - \frac{6q^2(n-q)^2}{n^3\phi^2} - \frac{q^2(n-q)^3}{n^3\phi^2} + \frac{12q^2(n-q)}{n^4\phi^2} \\ &\quad + \frac{11q^2(n-q)^2}{n^4\phi^2} + \frac{3q^2(n-q)^3}{n^4\phi^2} + \frac{q^2(n-q)^4}{4n^4\phi^2}. \end{aligned}$$

Desenvolvendo os termos da expressão anterior, temos

$$\begin{aligned} \mathbb{E} \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} &= \frac{n^2}{4\phi^2} + \frac{n}{2\phi^2} - \frac{q}{2\phi^2} + \frac{n^2}{4\phi^2} - \frac{nq}{2\phi^2} + \frac{q^2}{4\phi^2} - \frac{n^2}{2\phi^2} + \frac{nq}{2\phi^2} - \frac{nq}{\phi^2} \\ &\quad + \frac{q^2}{\phi^2} - \frac{3q^2}{n\phi^2} + \frac{3nq}{2\phi^2} - \frac{3q^2}{\phi^2} + \frac{3q^3}{2n\phi^2} + \frac{2q^2}{n\phi^2} - \frac{2q^3}{n^2\phi^2} \\ &\quad + \frac{q^2}{\phi^2} - \frac{2q^3}{n\phi^2} + \frac{q^4}{n^2\phi^2} - \frac{4q}{n\phi^2} + \frac{4q^2}{n^2\phi^2} + \frac{6q^2}{n\phi^2} - \frac{3q^3}{n^2\phi^2} \\ &\quad - \frac{nq}{2\phi^2} + \frac{3q^2}{2\phi^2} - \frac{3q^3}{2n\phi^2} + \frac{q^4}{2n^2\phi^2} - \frac{8q^2}{n^2\phi^2} + \frac{8q^3}{n^3\phi^2} - \frac{6q^2}{n\phi^2} \\ &\quad + \frac{12q^3}{n^2\phi^2} - \frac{6q^4}{n^2\phi^2} - \frac{q^2}{\phi^2} - \frac{3q^3}{n\phi^2} - \frac{3q^4}{n^2\phi^2} + \frac{q^4}{n^3\phi^2} + \frac{12q^2}{n^3\phi^2} \\ &\quad - \frac{12q^3}{n^4\phi^2} + \frac{11q^2}{n^2\phi^2} - \frac{22q^3}{n^3\phi^2} + \frac{11q^3}{n^4\phi^2} + \frac{3q^2}{n\phi^2} - \frac{9q^3}{n^2\phi^2} + \frac{9q^4}{n^3\phi^2} \\ &\quad - \frac{3q^5}{n^4\phi^2} + \frac{q^2}{4\phi^2} - \frac{q^3}{n\phi^2} + \frac{3q^4}{2n^2\phi^2} - \frac{q^5}{n^3\phi^2} + \frac{q^6}{4n^4\phi^2}. \end{aligned}$$

Vários termos são de ordem $O(n^{-1})$ e, então,

$$\begin{aligned} E \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} &= \frac{n^2}{4\phi^2} + \frac{n}{2\phi^2} - \frac{q}{2\phi^2} + \frac{n^2}{4\phi^2} - \frac{nq}{2\phi^2} + \frac{q^2}{4\phi^2} - \frac{n^2}{2\phi^2} + \frac{nq}{2\phi^2} - \frac{nq}{\phi^2} \\ &\quad + \frac{q^2}{\phi^2} + \frac{3nq}{2\phi^2} - \frac{3q^2}{\phi^2} + \frac{q^2}{\phi^2} - \frac{nq}{2\phi^2} + \frac{3q^2}{2\phi^2} - \frac{q^2}{\phi^2} + \frac{q^2}{4\phi^2} + O(n^{-1}) \\ &= \frac{n}{2\phi^2} - \frac{q}{2\phi^2} - \frac{nq}{\phi^2} \left(\frac{1}{2} - \frac{1}{2} + 1 - \frac{3}{2} + \frac{1}{2} \right) \\ &\quad + \frac{q^2}{\phi^2} \left(\frac{1}{4} - 3 + 1 + \frac{3}{2} + \frac{1}{4} \right) + O(n^{-1}). \end{aligned}$$

Podemos concluir que

$$E \left\{ \left(\frac{d}{d\phi} \bar{L}_P(\phi) \right)^2 \right\} + E \left\{ \frac{d^2}{d\phi^2} \bar{L}_P(\phi) \right\} = + \frac{n}{2\phi^2} - \frac{q}{2\phi^2} - \frac{n}{2\phi^2} + \frac{q}{2\phi^2} + O(n^{-1}),$$

ou seja,

$$E \left\{ \frac{d}{d\phi} \bar{L}_P(\phi) \right\}^2 + E \left\{ \frac{d^2}{d\phi^2} \bar{L}_P(\phi) \right\} = O(n^{-1}).$$

Provamos então que a função escore ajustada satisfaz as propriedades dela exigidas neste exemplo.

3.1.3 Parte do vetor de parâmetros de regressão como de interesse e parâmetro de dispersão conhecido

Consideremos a situação em que o parâmetro de dispersão é conhecido, como ocorre na regressão logística, nos modelos de regressão de Poisson ou no modelo de regressão exponencial. Aqui o vetor de interesse é $\psi = (\beta^1, \dots, \beta^p)$ e o vetor de perturbação é $\zeta = (\beta^{p+1}, \dots, \beta^{p+q})$, que induz uma partição em X , a matriz do modelo, da forma

$$X = (X_1, X_2),$$

onde X_1 é uma matriz $n \times p$ e X_2 uma matriz $n \times q$.

Definidos os vetores de parâmetros, observamos que

$$\begin{aligned} \frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} + \lambda_{rt/u} \lambda_{sw/v} + \lambda_{rt/u} \lambda_{vw/s} - \frac{1}{2} \lambda_{rtu} \lambda_{vw/s} = \\ \frac{1}{2} \lambda_{rtu} \lambda_{svw} - \frac{5}{2} \lambda_{rtu} \lambda_{sv/w} + 2 \lambda_{rt/u} \lambda_{sv/w}, \end{aligned}$$

pela forma particular que assumem os cumulantes λ_{rtu} e $\lambda_{rt/u}$, obtidos em (B.4) e (B.5), respectivamente. Dado que $\nu^{st} = 0$, se algum dos índices s ou t assumir valores no conjunto $\{1, \dots, p\}$ então, do Teorema 2.1.1 temos

$$\Lambda(\vartheta) = \sum_{a,j,k=1}^p \sum_{r=1}^{p+q} \lambda^{ar} \nu^{jk} \left(\lambda_{rj/k} - \frac{1}{2} \lambda_{rjk} \right) L_a. \quad (3.6)$$

Por outro lado, vemos que

$$\begin{aligned} \lambda_{rj/k} - \frac{1}{2} \lambda_{rjk} &= -\phi \sum_{l=1}^n (f+g)_l x_{lr} x_{lj} x_{lk} + \frac{\phi}{2} \sum_{l=1}^n (f+2g)_l x_{lr} x_{lj} x_{lk} \\ &= -\frac{\phi}{2} \sum_{l=1}^n f_l x_{lr} x_{lj} x_{lk}, \end{aligned}$$

sendo f e g definidos em (B.1) e (B.2), respectivamente. Substituindo a expressão anterior em (3.6), obtemos

$$\Lambda(\vartheta) = -\frac{\phi}{2} \sum_{a=1}^p \sum_{l=1}^n f_l \sum_{r=1}^n \lambda^{ar} x_{lr} \sum_{j,k=1}^n \nu^{jk} x_{lj} x_{lk} L_a.$$

Mas $\sum_{r=1}^{p+q} \lambda^{ar} x_{lr} = -d_{al}$ e $\sum_{j,k=1}^n \nu^{jk} x_{lj} x_{lk} = -\phi^{-1} z_{2l}$, onde d_{al} é o elemento (a, l) da matriz $D = K^{11} X_1^\top + K^{12} X_2^\top$ e z_{2l} é o l -ésimo elemento da diagonal da matriz $Z_2 = X_2 (X_2^\top W X_2)^{-1} X_2^\top$. As matrizes K^{11} e K^{12} são os blocos correspondentes da partição na inversa da matriz de informação de Fisher, para o vetor β , induzida pela partição em X . A matriz de informação para o vetor β foi obtida em (1.12) como sendo $\phi(X^\top W X)$, de inversa $\phi^{-1}(X^\top W X)^{-1}$.

Desta forma,

$$\Lambda(\vartheta) = -\frac{1}{2} \mathbf{1}^\top F Z_{2d} D^\top X_1^\top V^{-1} T (y - \mu), \quad (3.7)$$

onde a matriz Z_{2d} corresponde à diagonalização da matriz Z_2 , isto é, todos os elementos fora da diagonal principal em Z_{2d} são zero e nela coincidem com os elementos de Z_2 , $F = \text{diag}\{f_1, \dots, f_n\}$, $V = \text{diag}\{V_1, \dots, V_n\}$ e $T = \text{diag}\{T_1, \dots, T_n\}$, com $T_l = d\mu^l/d\eta^l$, já definida na Seção 1.3.2.

Pela sua importância, encontramos expressões para as matrizes K^{11} e K^{12} , e provou-se que estas podem ser escritas como $K^{11} = \phi^{-1}(R^T W R)^{-1}$ e $K^{12} = -\phi^{-1}(R^T W R)^{-1} C^T$ (Ferrari, 1991), onde $R = X_1 - X_2 C$ e $C = (X_2^T W X_2)^{-1} X_2^T W X_1$. Temos então

$$\begin{aligned} D &= \phi^{-1} \{ (R^T W R)^{-1} X_1^T - (R^T W R)^{-1} C^T X_2^T \} \\ &= \phi^{-1} (R^T W R)^{-1} \{ X_1^T - C^T X_2^T \} \end{aligned}$$

e, desta forma, podemos escrever

$$D = \phi^{-1} (R^T W R)^{-1} R^T.$$

Desenvolveremos a seguir a expressão da função de ajuste $\Theta(\vartheta)$. Com este objetivo separamos esta função em três termos. O primeiro deles é

$$\begin{aligned} \sum \lambda^{ar} \lambda^{bs} \lambda^{tv} \lambda^{uw} & \left(\frac{1}{2} \lambda_{rst} \lambda_{uvw} - \lambda_{rst} \lambda_{uv/w} + \lambda_{rt/s} \lambda_{uv/w} + \lambda_{st/r} \lambda_{uv/w} \right. \\ & \left. - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} - \frac{1}{2} \lambda_{st/r} \lambda_{uvw} \right) L_a L_b. \end{aligned} \quad (3.8)$$

Notemos que

$$\begin{aligned} -\lambda_{rst} \lambda_{uv/w} &= -\phi^2 \sum_{l=1}^n (f+2g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+g)_m x_{mu} x_{mv} x_{mw}, \\ -\frac{1}{2} \lambda_{rs/t} \lambda_{uvw} &= -\frac{1}{2} \phi^2 \sum_{l=1}^n (f+g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+2g)_m x_{mu} x_{mv} x_{mw}, \\ -\frac{1}{2} \lambda_{st/r} \lambda_{uvw} &= -\frac{1}{2} \phi^2 \sum_{l=1}^n (f+g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+2g)_m x_{mu} x_{mv} x_{mw}, \\ \frac{1}{2} \lambda_{rst} \lambda_{uvw} &= \frac{1}{2} \phi^2 \sum_{l=1}^n (f+2g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+2g)_m x_{mu} x_{mv} x_{mw}, \\ \lambda_{rt/s} \lambda_{uv/w} &= \phi^2 \sum_{l=1}^n (f+g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+g)_m x_{mu} x_{mv} x_{mw} \end{aligned}$$

e que

$$\lambda_{st/r} \lambda_{uv/w} = \phi^2 \sum_{l=1}^n (f+g)_l x_{lr} x_{ls} x_{lt} \sum_{m=1}^n (f+g)_m x_{mu} x_{mv} x_{mw}.$$

Então, o termo dado em (3.8) é igual a

$$\begin{aligned}
& \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+2g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& - \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+2g)_i (f+g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& + \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& + \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& - \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_i (f+2g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& - \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_i (f+2g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw}.
\end{aligned}$$

Definamos a matriz $H = (h_{lm})$, onde cada elemento é definido como

$$h_{lm} = \sum_{a,b} d_{al} L_a d_{bm} L_b.$$

Desta forma, podemos escrever

$$\sum_{a,b} L_a L_b \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} = h_{ll}.$$

Então, matricialmente, (3.8) pode ser escrito como

$$1^\top Z Z_{2d} \left\{ \frac{1}{2}(F+2G)(F+2G) - 2(F+2G)(F+G) + 2(F+G)(F+G) \right\} H_d 1, \quad (3.9)$$

onde H_d é a diagonalização da matriz H e $G = \text{diag}\{g_1, \dots, g_n\}$.

O segundo termo de $\Theta(\vartheta)$,

$$\begin{aligned}
& \sum \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} + \lambda_{rt/u} \lambda_{vw/s} + \lambda_{rt/u} \lambda_{sw/v} \right. \\
& \left. + \lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{rtu} \lambda_{vw/s} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) L_a L_b,
\end{aligned}$$

é igual a

$$\begin{aligned}
& \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+2g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& - \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+2g)_l (f+g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& - \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_l (f+2g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& + \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& + \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& + \phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_{l,m} \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& - \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+2g)_l (f+g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& - \frac{1}{2}\phi^2 \sum_{a,b} L_a L_b \sum_{l,m} (f+g)_l (f+2g)_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw}.
\end{aligned}$$

Matricialmente, este termo pode ser expresso como

$$1^\top (H \otimes Z_2 \otimes Z_2) \left\{ \frac{1}{2}(F+2G)(F+2G) - 3(F+2G)(F+G) + 3(F+G)(F+G) \right\} 1. \quad (3.10)$$

O terceiro termo de $\Theta(\vartheta)$ é da forma

$$- \sum \lambda^{ar} \lambda^{bs} \nu^{tu} \left(\frac{1}{2} \lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su} + \lambda_{st/ru} - \frac{1}{2} \lambda_{rtu/s} - \frac{1}{2} \lambda_{stu/r} \right) L_a L_b, \quad (3.11)$$

e nele temos que (ver B.1)

$$-\frac{1}{2} \lambda_{rtu/s} - \frac{1}{2} \lambda_{stu/r} - \lambda_{rst/u} = 2\phi \sum_l \frac{d}{d\eta_l} (f+2g)_l x_{lr} x_{ls} x_{lt} x_{lu}$$

e

$$\lambda_{rt/su} + \lambda_{st/ru} = -2\phi \sum_l \frac{d}{d\eta_l} (f+g)_l x_{lr} x_{ls} x_{lt} x_{lu}.$$

Desta forma, (3.11) é

$$-\sum \lambda^{ar} \lambda^{bs} \nu^{tu} \left(-\frac{\phi}{2} \sum_l e_l x_{lr} x_{ls} x_{lt} x_{lu} - 2\phi \sum_l \frac{d}{d\eta_l} (f+g)_l x_{lr} x_{ls} x_{lt} x_{lu} + 2\phi \sum_l \frac{d}{d\eta_l} (f+2g)_l x_{lr} x_{ls} x_{lt} x_{lu} \right) L_a L_b,$$

onde e_l é definido em (B.6). Trabalhando um pouco mais a expressão acima obtemos que (3.11) é igual a

$$\begin{aligned} & \frac{1}{2} \phi \sum_{a,b,l} L_a(\vartheta) L_b(\vartheta) (\lambda^{ar} x_{lr}) (\lambda^{bs} x_{ls}) (\nu^{tu} x_{lt} x_{lu}) e_l \\ & - 2\phi \sum_{a,b} L_a(\vartheta) L_b(\vartheta) \sum_l \frac{d}{d\eta_l} (f+2g)_l (\lambda^{ar} x_{lr}) (\lambda^{bs} x_{ls}) (\nu^{tu} x_{lt} x_{lu}) \\ & + 2\phi \sum_{a,b} L_a(\vartheta) L_b(\vartheta) \sum_l \frac{d}{d\eta_l} (f+g)_l (\lambda^{ar} x_{lr}) (\lambda^{bs} x_{ls}) (\nu^{tu} x_{lt} x_{lu}). \end{aligned}$$

Matricialmente podemos escrever este termo como

$$-\frac{1}{4} \mathbf{1}^\top Z_{2d} (E - 4J) H_d \mathbf{1},$$

onde as matrizes E e J são matrizes de dimensão n , definidas como $E = \text{diag}\{e_1, \dots, e_n\}$ e $J = \text{diag}\{j_1, \dots, j_n\}$, sendo que $j_l = dg_l/d\eta_l$, com

$$\begin{aligned} \frac{dg}{d\eta} = & -\frac{4}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} + \frac{1}{V} \left\{ \left(\frac{d^2\mu}{d\eta^2} \right)^2 + \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3} \right\} \\ & + \frac{1}{V^2} \left(\frac{d\mu}{d\eta} \right)^4 \left\{ \frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - \frac{d^2V}{d\mu^2} \right\}. \end{aligned}$$

Esta expressão é obtida no Apêndice B.

As matrizes que aparecem na expressão acima são todas diagonais, por esta razão a soma de todos os elementos é igual ao traço do produto de matrizes envolvido. Assim, este termo é igual a

$$-\frac{1}{4} \text{tr}(Z_{2d}\{E - 4J\}H_d). \quad (3.12)$$

Desta forma, chegamos à expressão de $\Theta(\vartheta)$, o termo do ajuste que corrige o vício da informação da função escore perfilada quando o parâmetro de dispersão é conhecido. Este é

$$\begin{aligned} \Theta(\vartheta) = & \frac{1}{2} \mathbf{1}^\top Z Z_{2d} \left\{ \frac{1}{2} (F + 2G)(F + 2G) - 2(F + 2G)(F + G) \right. \\ & \left. + 2(F + G)(F + G) \right\} H_d \mathbf{1} + \frac{1}{2} \mathbf{1}^\top (H \otimes Z_2 \otimes Z_2) \left\{ \frac{1}{2} (F + 2G)(F + 2G) \right. \\ & \left. - 3(F + 2G)(F + G) + 3(F + G)(F + G) \right\} \mathbf{1} - \frac{1}{4} \text{tr}\{Z_{2d}(E - 4J)H_d\}, \end{aligned} \quad (3.13)$$

resultado da soma dos termos em (3.9), (3.10) e (3.12).

Podemos escrever de maneira diferente o resultado anterior. Temos que

$$\begin{aligned} & \frac{1}{4} \mathbf{1}^\top Z Z_{2d} \{(F + 2G)(F + 2G)\} H_d \mathbf{1} + \frac{1}{4} \mathbf{1}^\top (H \otimes Z_2 \otimes Z_2) (F + 2G)(F + 2G) \mathbf{1} = \\ & = \frac{1}{4} \mathbf{1}^\top (F + 2G) \{Z Z_{2d} H_d + H \otimes Z_2 \otimes Z_2\} (F + 2G) \mathbf{1}. \end{aligned}$$

Por outro lado,

$$\begin{aligned} & \frac{1}{2} \mathbf{1}^\top Z Z_{2d} \{-2(F + 2G)(F + G) + 2(F + G)(F + G)\} H_d \mathbf{1} = \\ & = -\frac{1}{2} \mathbf{1}^\top G (2Z Z_{2d} H_d) (F + G) \mathbf{1} \end{aligned}$$

e

$$\begin{aligned} & \frac{1}{2} \mathbf{1}^\top (H \otimes Z_2 \otimes Z_2) \{-3(F + 2G)(F + G) + 3(F + G)(F + G)\} \mathbf{1} = \\ & = -\frac{1}{2} G \{2Z Z_{2d} H_d + 3H \otimes Z_2 \otimes Z_2\} (F + G) \mathbf{1}. \end{aligned}$$

Somando estas duas expressões obtemos

$$-\frac{1}{2} G \{2Z Z_{2d} H_d + 3H \otimes Z_2 \otimes Z_2\} (F + G) \mathbf{1}$$

e, portanto, a expressão em (3.13) pode ser escrita como

$$\begin{aligned} \Theta(\vartheta) = & \frac{1}{4} \mathbf{1}^\top (F + 2G) \{Z Z_{2d} H_d + H \otimes Z_2 \otimes Z_2\} (F + 2G) \mathbf{1} \\ & - \frac{1}{2} G \{2Z Z_{2d} H_d + 3H \otimes Z_2 \otimes Z_2\} (F + G) \mathbf{1} - \frac{1}{4} \text{tr}\{Z_{2d}(E - 4J)H_d\}. \end{aligned} \quad (3.14)$$

3.1.4 Parte do vetor de parâmetros de regressão como de interesse e parâmetro de dispersão desconhecido

Decidir acerca das variáveis que influenciam a resposta constitui um problema típico em modelos de regressão. Para tomar decisões neste sentido devemos considerar parte do vetor de parâmetros de regressão como de interesse, que chamaremos de β_1 , constituindo a outra parte de vetor de parâmetros de regressão, que chamaremos de β_2 , junto com o parâmetro de dispersão, os parâmetros de perturbação. Neste caso é fácil mostrar que a forma da função de ajuste $\Lambda(\vartheta)$ permanece igual à obtida em (3.7).

Portanto, nos dedicaremos a encontrar a forma do acréscimo na expressão de $\Theta(\vartheta)$, em (3.14), provocado pelo desconhecimento do parâmetro de dispersão ϕ . Em outras palavras, vamos mostrar que $\Theta(\vartheta)$, nesta situação, será como em (3.14) acrescida de termos que surgem quando algum sub-índice atinge o valor correspondente a ϕ . Interessa-nos, portanto, considerar as diferentes situações nas quais algum sub-índice atinge o índice do parâmetro ϕ .

Com este objetivo vamos encontrar estes novos termos nas 16 situações em que algum dos sub-índices t, u, v e w representa o parâmetro ϕ , isto porque se algum dos sub-índices r ou s assumir o valor correspondente a ϕ , a expressão toda se anula. Observando a função B^{ab} , decidimos separá-la em duas partes. Isto induz uma separação equivalente na função de ajuste $\Theta(\vartheta)$. Primeiramente consideremos o seguinte termo de $\Theta(\vartheta)$

$$\begin{aligned} & \frac{1}{2} \left\{ \lambda^{ar} \lambda^{bs} \lambda^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rst} \lambda_{uvw} - \lambda_{rst} \lambda_{uv/w} + \lambda_{rt/s} \lambda_{uv/w} + \lambda_{st/r} \lambda_{uv/w} \right. \right. \\ & \quad \left. \left. - \frac{1}{2} \lambda_{rt/s} \lambda_{uvw} - \frac{1}{2} \lambda_{st/r} \lambda_{uvw} \right) \right. \\ & \quad + \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{uw} \left(\frac{1}{2} \lambda_{rtu} \lambda_{svw} - \lambda_{rtu} \lambda_{sv/w} - \lambda_{rt/u} \lambda_{svw} + \lambda_{rt/u} \lambda_{sv/w} \right. \\ & \quad \left. \left. + \lambda_{rt/u} \lambda_{uw/s} + \lambda_{tu/r} \lambda_{sv/w} - \frac{1}{2} \lambda_{rtu} \lambda_{uw/s} - \frac{1}{2} \lambda_{tu/r} \lambda_{svw} \right) \right\} L_a L_b, \end{aligned} \tag{3.15}$$

e nos ítems a seguir, estudaremos a forma que assume a expressão anterior em cada uma das diferentes situações em que algum dos sub-índices t, u, v e w representa o parâmetro ϕ .

- (i) Sejam $t, u, v, w = \phi$, ou seja, todos os sub-índices assumem o valor $p + q + 1$, correspon-

dente ao parâmetro ϕ . Vemos que a expressão em (3.15) pode ser escrita como

$$\begin{aligned} & \frac{1}{2} \left\{ \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rs\phi} \lambda_{\phi\phi\phi} - \lambda_{rs\phi} \lambda_{\phi\phi/\phi} + \lambda_{r\phi/s} \lambda_{\phi\phi/\phi} + \lambda_{s\phi/r} \lambda_{\phi\phi/\phi} \right. \right. \\ & \quad \left. \left. - \frac{1}{2} \lambda_{r\phi/s} \lambda_{\phi\phi\phi} - \frac{1}{2} \lambda_{s\phi/r} \lambda_{\phi\phi\phi} \right) \right. \\ & \quad + \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{r\phi\phi} \lambda_{s\phi\phi} - \lambda_{r\phi\phi} \lambda_{s\phi/\phi} - \lambda_{r\phi/\phi} \lambda_{s\phi\phi} + \lambda_{r\phi/\phi} \lambda_{s\phi/\phi} \right. \\ & \quad \left. + \lambda_{r\phi/\phi} \lambda_{\phi\phi/s} + \lambda_{\phi\phi/r} \lambda_{s\phi/\phi} - \frac{1}{2} \lambda_{r\phi\phi} \lambda_{\phi\phi/s} - \frac{1}{2} \lambda_{\phi\phi/r} \lambda_{s\phi\phi} \right) \left. \right\} L_a L_b = \\ & \quad \frac{1}{2} \left\{ \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \lambda^{\phi\phi} \left(\frac{1}{2} \lambda_{rs\phi} \lambda_{\phi\phi\phi} - \lambda_{rs\phi} \lambda_{\phi\phi/\phi} \right) \right\} L_a L_b, \end{aligned}$$

já que os cumulantes $\lambda_{r\phi/\phi}$, $\lambda_{s\phi/\phi}$, $\lambda_{r\phi/s}$, $\lambda_{r\phi\phi}$, $\lambda_{s\phi/\phi}$, $\lambda_{s\phi\phi}$, $\lambda_{\phi\phi/s}$ e $\lambda_{\phi\phi/r}$ se anulam.

Por outro lado, dado que $\lambda_{\phi\phi/\phi} = \lambda_{\phi\phi\phi}$, temos que a expressão anterior se reduz a $-\frac{1}{4} \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{\phi\phi} \lambda_{rs\phi} \lambda_{\phi\phi\phi} L_a L_b$ e, portanto,

$$\begin{aligned} & -\frac{1}{4} \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{\phi\phi} \lambda_{rs\phi} \lambda_{\phi\phi\phi} L_a L_b = \\ & = -\frac{nd_1'''(\phi)}{4n^2 d_1''(\phi)^2} \sum_{a,b,r,s} \lambda^{ar} \lambda^{bs} \sum_l -w_l x_{lr} x_{ls} L_a L_b \\ & = \frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \sum_{a,b} \sum_l w_l \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} L_a L_b \\ & = \frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \sum_l w_l \sum_{a,b} \sum_r \lambda^{ar} x_{lr} L_a \sum_s \lambda^{bs} x_{ls} L_b \\ & = \frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \sum_l w_l \sum_{a,b} d_{al} L_a d_{bl} L_b \\ & = \frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \sum_l w_l h_{ll} = \frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \text{tr}\{WH\}, \end{aligned} \tag{3.16}$$

este sendo o primeiro termo a adicionar para encontrar a forma de $\Theta(\vartheta)$.

(ii) Sejam $u, w = \phi$ e $t, v \in \{1, 2, \dots, p + q\}$. Temos que (3.15) é igual a

$$\begin{aligned} & \frac{1}{2} \left\{ \lambda^{ar} \lambda^{bs} \lambda^{tv} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rst} \lambda_{\phi v \phi} - \lambda_{rst} \lambda_{\phi v / \phi} + \lambda_{rt/s} \lambda_{\phi v / \phi} + \lambda_{st/r} \lambda_{\phi v / \phi} \right. \right. \\ & \quad \left. \left. - \frac{1}{2} \lambda_{rt/s} \lambda_{\phi v \phi} - \frac{1}{2} \lambda_{st/r} \lambda_{\phi v \phi} \right) \right. \\ & \quad + \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rt\phi} \lambda_{sv\phi} - \lambda_{rt\phi} \lambda_{sv/\phi} - \lambda_{rt/\phi} \lambda_{sv\phi} + \lambda_{rt/\phi} \lambda_{s\phi/v} \right. \\ & \quad \left. \left. + \lambda_{rt/\phi} \lambda_{v\phi/s} + \lambda_{t\phi/r} \lambda_{sv/\phi} - \frac{1}{2} \lambda_{rt\phi} \lambda_{v\phi/s} - \frac{1}{2} \lambda_{t\phi/r} \lambda_{sv\phi} \right) \right\} L_a L_b. \end{aligned}$$

Dado que os cumulantes $\lambda_{\phi v \phi}$, $\lambda_{\phi v / \phi}$, $\lambda_{s\phi/v}$, $\lambda_{v\phi/s}$ e $\lambda_{t\phi/r}$ são nulos, este termo se reduz a

$$\frac{1}{2} \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rt\phi} \lambda_{sv\phi} - \lambda_{rt\phi} \lambda_{sv/\phi} - \lambda_{rt/\phi} \lambda_{sv\phi} \right) L_a L_b. \quad (3.17)$$

Também temos que $\lambda_{sv/\phi} = \lambda_{sv\phi}$ e $\lambda_{rt/\phi} = \lambda_{rt\phi}$ e, portanto, a expressão acima se reduz a $-\frac{3}{4} \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{\phi\phi} \lambda_{rt\phi} \lambda_{sv\phi} L_a L_b$. Substituindo aqui as expressões dos cumulantes correspondentes, temos que

$$\begin{aligned} & -\frac{3}{4} \sum_{a,b,r,s,t,v} \lambda^{ar} \lambda^{bs} \nu^{tv} \nu^{\phi\phi} \lambda_{rt\phi} \lambda_{sv\phi} L_a L_b = \\ & = -\frac{3}{4nd_1''(\phi)} \sum_{a,b,r,s,t,v} \lambda^{ar} \lambda^{bs} \nu^{tv} \sum_l -w_l x_{lr} x_{lt} \sum_m -w_m x_{ms} x_{mv} L_a L_b \\ & = -\frac{3}{4nd_1''(\phi)} \sum_{a,b,l,m} w_l \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} w_m L_a L_b \\ & = \frac{3}{4nd_1''(\phi)} \sum_{a,b,l,m} \phi^{-1} w_l d_{al} d_{bm} z_{2_{lm}} w_m L_a L_b \\ & = \frac{3}{4n\phi d_1''(\phi)} \sum_{l,m} w_l z_{2_{lm}} \left(\sum_{a,b} d_{al} L_a d_{bm} L_b \right) w_m \\ & = \frac{3}{4n\phi d_1''(\phi)} \sum_{l,m} w_l z_{2_{lm}} h_{lm} w_m = \frac{3}{4n\phi d_1''(\phi)} 1^\top W (Z_2 \otimes H) W 1. \end{aligned}$$

Mas $1^\top W (Z_2 \otimes H) W 1 = \text{tr}\{W Z_2 H W\}$ (ver Apêndice C) e, então, obtemos que a expressão em (3.17) é igual a

$$\frac{3}{4n\phi d_1''(\phi)} \text{tr}\{W Z_2 H W\}, \quad (3.18)$$

constituindo esta mais uma parcela a ser adicionada à função de ajuste.

- (iii) Agora consideremos a situação em que $u, w \in \{1, 2, \dots, p + q\}$ e $t, v = \phi$. Neste caso (3.15) é da forma

$$\begin{aligned} & \frac{1}{2} \sum_{a,b,r,s,u,w} \left\{ \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{uw} \left(\frac{1}{2} \lambda_{rs\phi} \lambda_{u\phi w} - \lambda_{rs\phi} \lambda_{u\phi/w} + \lambda_{r\phi/s} \lambda_{u\phi/w} + \lambda_{s\phi/r} \lambda_{u\phi/w} \right. \right. \\ & \quad \left. \left. - \frac{1}{2} \lambda_{r\phi/s} \lambda_{u\phi w} - \frac{1}{2} \lambda_{s\phi/r} \lambda_{u\phi w} \right) \right. \\ & \quad \left. + \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{uw} \left(\frac{1}{2} \lambda_{r\phi u} \lambda_{s\phi w} - \lambda_{r\phi u} \lambda_{s\phi/w} - \lambda_{r\phi/u} \lambda_{s\phi w} + \lambda_{r\phi/u} \lambda_{sw/\phi} \right. \right. \\ & \quad \left. \left. + \lambda_{r\phi/u} \lambda_{\phi w/s} + \lambda_{\phi u/r} \lambda_{s\phi/w} - \frac{1}{2} \lambda_{r\phi u} \lambda_{\phi w/s} - \frac{1}{2} \lambda_{\phi u/r} \lambda_{s\phi w} \right) \right\} L_a L_b. \end{aligned}$$

Lembrando que os cumulantes $\lambda_{u\phi/w}$, $\lambda_{r\phi/s}$, $\lambda_{s\phi/r}$, $\lambda_{s\phi/w}$, $\lambda_{r\phi/u}$, $\lambda_{\phi w/s}$ e $\lambda_{\phi u/r}$ se anulam, a expressão anterior se reduz a

$$\frac{1}{2} \sum_{a,b,r,s,u,w} \left(\frac{1}{2} \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{uw} \lambda_{rs\phi} \lambda_{u\phi w} + \frac{1}{2} \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{uw} \lambda_{r\phi u} \lambda_{s\phi w} \right) L_a L_b.$$

Desenvolveremos agora os termos que aparecem na expressão anterior. O primeiro destes é

$$\begin{aligned} & \frac{1}{4} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{uw} \lambda_{rs\phi} \lambda_{u\phi w} L_a L_b = \\ & = \frac{1}{4n d_1''(\phi)} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \nu^{uw} \lambda_{rs\phi} \lambda_{uw\phi} L_a L_b \\ & = \frac{1}{4n d_1''(\phi)} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \nu^{uw} \sum_l -w_l x_{lr} x_{ls} \sum_m -w_m x_{mu} x_{mw} L_a L_b \\ & = \frac{1}{4n d_1''(\phi)} \sum_{a,b} \sum_{l,m} w_l w_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ls} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} L_a L_b \\ & = -\frac{\phi^{-1}}{4n \phi d_1''(\phi)} \sum_{l,m} w_l w_m z_{2mm} \sum_{a,b} d_{al} L_a d_{bl} L_b \\ & = -\frac{1}{4n \phi d_1''(\phi)} \sum_{l,m} w_l w_m z_{2mm} h_{ll} = -\frac{1}{4n \phi d_1''(\phi)} \sum_m w_m z_{2mm} \sum_l w_l h_{ll} \\ & = -\frac{1}{4n \phi d_1''(\phi)} \text{tr}\{W Z_2\} \text{tr}\{W H\}. \end{aligned}$$

Por outro lado,

$$\begin{aligned}\text{tr}\{WZ_2\} &= \text{tr}\{WX_2(X_2^\top WX_2)^{-1}X_2^\top\} = \text{tr}\{X_2^\top WX_2(X_2^\top WX_2)^{-1}\} \\ &= \text{tr}\{I_q\} = q,\end{aligned}$$

o número de parâmetros de perturbação no vetor de parâmetros de regressão. Desta forma, obtemos que

$$\frac{1}{4} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \lambda^{\phi\phi} \nu^{uw} \lambda_{rs\phi} \lambda_{u\phi w} L_a(\vartheta) L_b(\vartheta) = -\frac{q}{4n\phi d_1''(\phi)} \text{tr}\{WH\}. \quad (3.19)$$

O segundo termo,

$$\frac{1}{4} \sum \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{uw} \lambda_{r\phi u} \lambda_{s\phi w} L_a L_b,$$

é igual a

$$\begin{aligned}\frac{1}{4nd_1''(\phi)} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \nu^{uw} \lambda_{r\phi u} \lambda_{s\phi w} L_a L_b &= \frac{1}{4nd_1''(\phi)} \sum_{a,b,r,s,u,w} \lambda^{ar} \lambda^{bs} \nu^{uw} \sum_l -w_l x_{lr} x_{lu} \sum_m -w_m x_{ms} x_{mw} L_a L_b \\ &= \frac{1}{4nd_1''(\phi)} \sum_{l,m} w_l w_m \sum_r \lambda^{ar} x_{lr} \sum_s \lambda^{bs} x_{ms} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} L_a L_b \\ &= -\frac{\phi^{-1}}{4n\phi d_1''(\phi)} \sum_{l,m} w_l z_{2lm} w_m \sum_{a,b} d_{al} L_a d_{bm} L_b \\ &= -\frac{1}{4n\phi d_1''(\phi)} \sum_{l,m} w_l z_{2lm} h_{lm} w_m \\ &= -\frac{1}{4n\phi d_1''(\phi)} \mathbf{1}^\top W (Z_2 \otimes H) W \mathbf{1} \\ &= -\frac{1}{4n\phi d_1''(\phi)} \text{tr}\{WZ_2WH\}.\end{aligned}$$

Portanto, podemos afirmar que o segundo termo é da forma

$$\frac{1}{4} \sum \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{uw} \lambda_{r\phi u} \lambda_{s\phi w} L_a L_b = -\frac{1}{4n\phi d_1''(\phi)} \text{tr}\{WZ_2WH\}.$$

Então, o termo a ser acrescido na função de ajuste é soma da expressão anterior e (3.19), ou seja,

$$-\frac{q}{4n\phi d_1''(\phi)} \text{tr}\{WH\} - \frac{1}{4n\phi d_1''(\phi)} 1^\top \text{tr}\{WZ_2WH\}. \quad (3.20)$$

- (iv) O caso em que $u, w, t, v \in \{1, 2, \dots, p+q\}$ reduz-se à situação em que ϕ é conhecido. É por esta razão que somamos termos à expressão em (3.14) para encontrar a função de ajuste quando o parâmetro ϕ é desconhecido.
- (vi) Todas as outras situações possíveis de serem consideradas implicam $\nu^{uw} = \nu^{\phi w} = 0$ ou $\nu^{uw} = \nu^{u\phi} = 0$ ou $\nu^{tv} = \nu^{t\phi} = 0$ ou $\nu^{tv} = \nu^{\phi v} = 0$, anulando-se todos os termos em (3.14). Portanto, não temos mais termos a acrescentar em (3.14) além daqueles obtidos até o item (iii).

A segunda parte de $\Theta(\vartheta)$ é da forma

$$-\frac{1}{2} \sum \lambda^{ar} \lambda^{bs} \nu^{tu} \left(\frac{1}{2} \lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su} + \lambda_{st/ru} - \frac{1}{2} \lambda_{rtu/s} - \frac{1}{2} \lambda_{stu/r} \right) L_a L_b, \quad (3.21)$$

e vamos estudá-la quando os sub-índices t ou u representam o parâmetro ϕ . Observemos que desta forma completaremos a expressão em (2.6).

- (i) Se $t = \phi$ e $u \neq \phi$, então $\nu^{tu} = \nu^{\phi u} = 0$ e, portanto, a expressão em (3.21) é nula.
- (ii) Quando $t \neq \phi$ e $u = \phi$, estamos em uma situação similar à anterior, ou seja, aqui também $\nu^{tu} = \nu^{t\phi} = 0$ e, portanto, (3.21) também se anula.
- (iii) Se $t, u = \phi$, então (3.21) assume a forma

$$-\frac{1}{2} \sum \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rs\phi\phi} - \lambda_{rs\phi/\phi} + \lambda_{r\phi/s\phi} + \lambda_{s\phi/r\phi} - \frac{1}{2} \lambda_{r\phi\phi/s} - \frac{1}{2} \lambda_{s\phi\phi/r} \right) L_a L_b,$$

mas os cumulantes $\lambda_{rs\phi\phi}$, $\lambda_{rs\phi/\phi}$, $\lambda_{r\phi/s\phi}$, $\lambda_{s\phi/r\phi}$, $\lambda_{r\phi\phi/s}$ e $\lambda_{s\phi\phi/r}$ se anulam e, portanto, não temos termos a adicionar.

Assim, o acréscimo na expressão da função de ajuste é a soma de (3.16), (3.18) e (3.20), obtidas anteriormente, ou seja,

$$\frac{d_1'''(\phi)}{4nd_1''(\phi)^2} \text{tr}\{WH\} + \frac{1}{2n\phi d_1''(\phi)} \text{tr}\{WZ_2WH\} - \frac{q}{4n\phi d_1''(\phi)} \text{tr}\{WH\}.$$

Substituindo a expressão de $\Theta(\vartheta)$ para ϕ conhecido em (3.14), temos que, se ϕ for desconhecido,

$$\begin{aligned} \Theta(\vartheta) = & \frac{1}{4} \mathbf{1}^\top (F + 2G)(ZZ_{2d}H_d + Z_2 \otimes Z_2 \otimes H)(F + 2G)\mathbf{1} \\ & - \frac{1}{2} \mathbf{1}^\top G(2ZZ_{2d}H_d + 3Z_2 \otimes Z_2 \otimes H)(F + G)\mathbf{1} \\ & - \frac{1}{4} \text{tr}\{Z_{2d}(E - 4J)H_d\} + \frac{1}{4n} \left\{ \frac{d_1'''(\phi)}{d_1''(\phi)^2} - \frac{q}{\phi d_1''(\phi)} \right\} \text{tr}\{WH\} \\ & + \frac{1}{2n\phi d_1''(\phi)} \text{tr}\{WZ_2WH\}. \end{aligned} \quad (3.22)$$

3.1.5 Exemplo: todos os parâmetros de regressão de interesse

Desenvolvemos a seguir a forma das funções de ajuste $\Lambda(\vartheta)$ e $\Theta(\vartheta)$ com o objetivo de conferir os resultados da seção anterior. Considerar todos os parâmetros de regressão como de interesse significa escolher $\psi = \beta$ e, portanto, de dimensão p e o vetor de perturbação $\zeta = \phi$, unidimensional. É interessante observar que, desta escolha, temos ortogonalidade entre os vetores ψ e ζ , já que $E\{L_{\psi\zeta}\} = 0$.

Observemos primeiramente que

$$\nu = \begin{pmatrix} 0 & 0 \\ 0 & \lambda^{\phi\phi} \end{pmatrix},$$

que é uma matriz $p + 1 \times p + 1$. Vemos então que

$$\Lambda(\vartheta) = \sum_{a,r} \lambda^{ar} \lambda^{\phi\phi} \left(\lambda_{r\phi/\phi} - \frac{1}{2} \lambda_{r\phi\phi} \right) L_a.$$

Acompanhando os diferentes valores do índice r , esta função pode ser escrita como

$$\Lambda(\vartheta) = \sum_{a,b} \lambda^{ab} \lambda^{\phi\phi} \left(\lambda_{b\phi/\phi} - \frac{1}{2} \lambda_{b\phi\phi} \right) L_a + \sum_a \lambda^{a\phi} \lambda^{\phi\phi} \left(\lambda_{\phi\phi/\phi} - \frac{1}{2} \lambda_{\phi\phi\phi} \right) L_a,$$

e dado que $\lambda_{b\phi/\phi} = 0$, $\lambda_{b\phi\phi} = 0$ e $\lambda^{a\phi} = 0$, obtemos que $\Lambda(\vartheta)$ acima é nula.

Nossa atenção está agora na função de ajuste $\Theta(\vartheta)$. Lembrando que ϕ é o parâmetro de perturbação, temos

$$\begin{aligned} B^{ab} = & \lambda^{ar} \lambda^{bs} \lambda^{tv} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rst} \lambda_{\phi v\phi} - \lambda_{rst} \lambda_{\phi v/\phi} + \lambda_{rt/s} \lambda_{\phi v/\phi} + \lambda_{st/r} \lambda_{\phi v/\phi} \right. \\ & \left. - \frac{1}{2} \lambda_{rt/s} \lambda_{\phi v\phi} - \frac{1}{2} \lambda_{st/r} \lambda_{\phi v\phi} \right) \\ & + \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{r\phi\phi} \lambda_{s\phi\phi} - \lambda_{r\phi\phi} \lambda_{s\phi/\phi} - \lambda_{r\phi/\phi} \lambda_{s\phi\phi} + \lambda_{r\phi/\phi} \lambda_{s\phi/\phi} \right. \\ & \left. + \lambda_{r\phi/\phi} \lambda_{\phi\phi/s} + \lambda_{\phi\phi/r} \lambda_{s\phi/\phi} - \frac{1}{2} \lambda_{r\phi\phi} \lambda_{\phi\phi/s} - \frac{1}{2} \lambda_{\phi\phi/r} \lambda_{s\phi\phi} \right) \\ & - \lambda^{ar} \lambda^{bs} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{rs\phi\phi} - \lambda_{rs\phi/\phi} + \lambda_{r\phi/s\phi} - \lambda_{s\phi/r\phi} - \frac{1}{2} \lambda_{r\phi\phi/s} - \frac{1}{2} \lambda_{s\phi\phi/r} \right). \end{aligned}$$

Os índices r, s, t, v percorrem o vetor de parâmetros $\vartheta = (\beta, \phi)$, mas os índices a e b percorrem o vetor de interesse, e por este motivo os cumulantes $\lambda^{a\phi}$, $\lambda^{b\phi}$, $\lambda_{r\phi\phi}$, $\lambda_{s\phi/\phi}$, $\lambda_{r\phi/\phi}$, $\lambda_{s\phi/\phi}$, $\lambda_{\phi\phi/s}$, $\lambda_{\phi\phi/r}$, $\lambda_{rs\phi\phi}$, $\lambda_{rs\phi/\phi}$, $\lambda_{r\phi/s\phi}$, $\lambda_{s\phi/r\phi}$, $\lambda_{r\phi\phi/s}$ e $\lambda_{s\phi\phi/r}$ são nulos. Portanto, a função B^{ab} se reduz a

$$\lambda^{ac} \lambda^{bd} \lambda^{tv} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{c dt} \lambda_{\phi v\phi} - \lambda_{c dt} \lambda_{\phi v/\phi} + \lambda_{ct/d} \lambda_{\phi v/\phi} + \lambda_{dt/c} \lambda_{\phi v/\phi} - \frac{1}{2} \lambda_{dt/c} \lambda_{\phi v\phi} - \frac{1}{2} \lambda_{dt/c} \lambda_{\phi v\phi} \right),$$

onde a, b, c, d assumem valores em $\{1, \dots, p\}$.

Esta expressão, por sua vez, pode ser escrita como

$$\begin{aligned} B^{ab} = & \lambda^{ac} \lambda^{bd} \lambda^{ef} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{cde} \lambda_{\phi f\phi} - \lambda_{cde} \lambda_{\phi f/\phi} + \lambda_{ce/d} \lambda_{\phi f/\phi} + \lambda_{de/c} \lambda_{\phi f/\phi} \right. \\ & \left. - \frac{1}{2} \lambda_{de/c} \lambda_{\phi f\phi} - \frac{1}{2} \lambda_{de/c} \lambda_{\phi f\phi} \right) \\ & \lambda^{ac} \lambda^{bd} \lambda^{\phi\phi} \nu^{\phi\phi} \left(\frac{1}{2} \lambda_{cd\phi} \lambda_{\phi\phi\phi} - \lambda_{cd\phi} \lambda_{\phi\phi/\phi} + \lambda_{c\phi/d} \lambda_{\phi\phi/\phi} + \lambda_{d\phi/c} \lambda_{\phi\phi/\phi} \right. \\ & \left. - \frac{1}{2} \lambda_{d\phi/c} \lambda_{\phi\phi\phi} - \frac{1}{2} \lambda_{d\phi/c} \lambda_{\phi\phi\phi} \right). \end{aligned}$$

Aqui os índices e e f também assumem valores em $\{1, \dots, p\}$.

Os cumulantes $\lambda_{\phi f\phi}$, $\lambda_{\phi f/\phi}$, $\lambda_{c\phi/d}$ e $\lambda_{d\phi/c}$ são nulos e $\lambda_{\phi\phi/\phi} = \lambda_{\phi\phi\phi}$. Então,

$$B^{ab} = -\frac{1}{2} \lambda^{ac} \lambda^{bd} \lambda^{\phi\phi} \nu^{\phi\phi} \lambda_{cd\phi} \lambda_{\phi\phi\phi}$$

e, portanto,

$$\begin{aligned}\Theta(\vartheta) &= -\frac{1}{4}\lambda^{ac}\lambda^{bd}\lambda^{\phi\phi}\nu^{\phi\phi}\lambda_{cd\phi}\lambda_{\phi\phi\phi}L_aL_b \\ &= \frac{nd_1'''(\phi)}{4(nd_1''(\phi))^2}\sum_{l=1}^n w_l \sum_{a,b,c,d=1}^p \lambda^{ac}\lambda^{bd}x_{lc}x_{ld}L_aL_b.\end{aligned}$$

Sabemos também que

$$\sum_{a,b,c,d=1}^p \lambda^{ac}x_{lc}\lambda^{bd}x_{ld}L_a(\vartheta)L_b(\vartheta) = h_{ll},$$

onde h_{ll} é o l -ésimo elemento da diagonal principal da matriz H . Então, a expressão da função de ajuste é

$$\Theta(\vartheta) = \frac{1}{4n} \frac{d_1'''(\phi)}{d_1''(\phi)^2} \text{tr}\{WH\}, \quad (3.23)$$

que coincide com a expressão (3.22) quando $Z_2 = 0$, constituindo assim um indicativo de que nosso resultado está correto.

3.1.6 Exemplo: distribuição normal; média como parâmetro de interesse

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes com distribuição normal de média μ e variância σ^2 . Segundo nossa notação $\psi = \mu$ e $\zeta = \sigma^2$. É simples ver que o logaritmo da função de verossimilhança assume a forma

$$L(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \quad (3.24)$$

e que a função escore para σ^2 é

$$U(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2.$$

Fixando μ obtemos o estimador de máxima verossimilhança para σ^2 como

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu)^2$$

e, conseqüentemente, substituindo em (3.24) σ^2 por $\tilde{\sigma}^2$ obtemos a função de verossimilhança perfilada para μ :

$$L_P(\mu) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sum_{i=1}^n (y_i - \mu)^2 + \frac{n}{2} \log(n) - \frac{n}{2}.$$

Para encontrarmos a função de ajuste necessitamos encontrar a função $d_1(\phi)$ e as derivadas correspondentes. Neste sentido, lembrando que no modelo normal ϕ , o parâmetro de dispersão nos modelos lineares generalizados, é o inverso da variância ($1/\sigma^2$), temos que $d_1(\phi) = \log \phi/2$, ou seja, $d_1(\sigma^2) = -\log \sigma^2/2$. Portanto, $d_1''(\sigma^2) = -\sigma^4/2$ e $d_1'''(\sigma^2) = \sigma^6$. Desta forma, a função de ajuste é

$$\Theta(\vartheta) = \frac{d_1'''(\sigma^2)}{4nd_1''(\sigma^2)^2} \text{tr}\{WH\} = \frac{1}{n\sigma^2} \text{tr}\{WH\},$$

lembrando que aqui $\vartheta = (\mu, \sigma^2)$.

Encontremos a seguir a forma da matriz W . No modelo linear normal a função de ligação é a canônica e, por isso $\theta = \eta$ e $\mu = \eta$. Então, as matrizes V e T se reduzem à matriz identidade e, portanto, $W = I$.

Obtenhamos agora a expressão da matriz H . Neste sentido, lembremos que foi definida na Seção 3.1.3 e nesta situação seus elementos são iguais a

$$h_{lm} = (d_{1l}U(\mu))(d_{1m}U(\mu)) = d_{1l}d_{1m}U(\mu)^2,$$

onde d_{lm} é o elemento (l, m) da matriz $D = \phi^{-1}(X^T W X)^{-1} X^T = \phi^{-1}(X^T X)^{-1} X^T$ e $X = (1, \dots, 1)^T$, a matriz do modelo, de dimensão $n \times 1$. Então $D = \phi^{-1}(1/n, \dots, 1/n)$.

Por outro lado, $U(\mu) = n(\bar{y} - \mu)/\sigma^2$, do que resulta $h_{lm} = (\bar{y} - \mu)^2/(n\phi)^2$ e $\text{tr}\{H\} = (\bar{y} - \mu)^2/n\phi^2$. Portanto, a função de ajuste é

$$\Theta(\vartheta) = \frac{1}{\sigma^2} (\bar{y} - \mu)^2,$$

que avaliada em $\hat{\sigma}^2$ é da forma

$$\tilde{\Theta}(\vartheta) = \frac{n(\bar{y} - \mu)^2}{\sum_{i=1}^n (y_i - \mu)^2}. \quad (3.25)$$

Concluimos então que, utilizando nossos resultados, o logaritmo da função de verossimilhança perfilada ajustada é da forma

$$\bar{L}_P(\mu) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sum_{i=1}^n (y_i - \mu)^2 + \frac{n}{2} \log(n) - \frac{n}{2} + \frac{n(\bar{y} - \mu)^2}{\sum_{i=1}^n (y_i - \mu)^2}. \quad (3.26)$$

Desta função obtemos o estimador de máxima verossimilhança perfilado ajustado $\bar{\mu}_P$ de μ . Primeiramente encontremos $\bar{U}_P(\mu)$:

$$\begin{aligned} \frac{d}{d\mu} \bar{L}_P(\mu) &= -\frac{n-2}{2} \frac{\sum_{i=1}^n (y_i - \mu)}{\sum_{i=1}^n (y_i - \mu)^2} + \frac{-2n(\bar{y} - \mu) \sum_{i=1}^n (y_i - \mu)^2 + 2n(\bar{y} - \mu)^2 \sum_{i=1}^n (y_i - \mu)}{\{\sum_{i=1}^n (y_i - \mu)^2\}^2} \\ &= n \frac{\bar{y} - \mu}{\sum_{i=1}^n (y_i - \mu)^2} + 2n \frac{\bar{y} - \mu}{\{\sum_{i=1}^n (y_i - \mu)^2\}^2} \left\{ (\bar{y} - \mu)^2 - \sum_{i=1}^n (y_i - \mu)^2 \right\}, \end{aligned}$$

e observando que $\bar{U}_P(\bar{y}) = 0$, concluimos que $\bar{\mu}_P = \bar{y}$, ou seja, $\bar{\mu}_P$ coincide com o estimador de máxima verossimilhança de μ .

A estatística do teste da razão de verossimilhanças ajustado, que é definida como

$$\bar{W}(\mu) = 2\{\bar{L}_P(\bar{y}) - \bar{L}_P(\mu)\},$$

assume a forma

$$\bar{W}(\mu) = n \log \frac{\sum_{i=1}^n (y_i - \mu)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} - 2n \frac{(\bar{y} - \mu)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (3.27)$$

Estudemos a estatística de teste acima. Para isso definamos as variáveis aleatórias

$$F_1 = \frac{(\bar{Y} - \mu)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

e

$$F_2 = \frac{\sum_{i=1}^n (Y_i - \mu)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}.$$

É conhecido que $n(\bar{Y} - \mu)^2/\sigma^2 \sim \chi_1^2$ e, então,

$$\frac{\frac{n}{\sigma^2} (\bar{Y} - \mu)^2}{\frac{1}{(n-1)\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2} \sim F(1, n-1)$$

ou

$$n(n-1)F_1 \sim F(1, n-1),$$

onde $F(1, n - 1)$ denota a distribuição F de Fisher com 1 grau de liberdade no numerador e $n - 1$ graus de liberdade no denominador. Então, $E\{n(n - 1)F_1\} = n - 1/\{n - 3\}$ e, portanto, $E\{F_1\} = 1/\{n(n - 3)\}$.

Acerca de F_2 , observamos que

$$\frac{\sum_{i=1}^n (Y_i - \mu)^2}{\sigma^2} = \frac{n}{\sigma^2} (\bar{Y} - \mu)^2 + \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Desta forma,

$$F_2 = \frac{\frac{n}{\sigma^2} (\bar{Y} - \mu)^2}{\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2} + \frac{\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2}{\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2},$$

ou seja,

$$F_2 = nF_1 + 1,$$

e obtemos então que $\log F_2 = \log(nF_1 + 1)$. A expansão em série de Taylor de $\log(1 + nF_1)$ é da forma

$$\log(1 + nF_1) = nF_1 - \frac{1}{2}n^2F_1^2 + \frac{1}{3}n^3F_1^3 + \dots,$$

e, então, encontramos

$$E\{\log F_2\} = \frac{1}{n - 3} + O(n^{-6}).$$

A estatística de teste em (3.27) pode ser escrita como

$$\bar{\mathcal{W}}(\mu) = n \log F_2 - 2nF_1,$$

e, substituindo as esperanças das variáveis F_1 e $\log F_2$, temos que

$$E\{\bar{\mathcal{W}}(\mu)\} = \frac{n}{n - 3} - 2\frac{n}{n(n - 3)} + O(n^{-5}),$$

e, portanto,

$$E\{\bar{\mathcal{W}}(\mu)\} = 1 + O(n^{-1}). \quad (3.28)$$

Note que a média de $\bar{\mathcal{W}}(\mu)$ coincide com a da distribuição $\chi^2(1)$, sua distribuição de assintótica, com um erro de ordem $O(n^{-1})$. Esta aproximação e a de todos os momentos da estatística pelos momentos correspondentes da distribuição $\chi^2(1)$ podem ser melhoradas através do uso da correção de Bartlett. Este é o objetivo da próxima seção.

3.2 Correção de Bartlett para o teste da razão de verossimilhanças perfiladas ajustadas nos modelos lineares generalizados

Continuando nosso estudo sobre a verossimilhança perfilada nos modelos lineares generalizados, vejamos a forma da correção de Bartlett para o teste da razão de verossimilhanças perfilada ajustada. Aqui e nas seguintes seções deste capítulo aplicaremos os resultados obtidos na Seção 2.2. Seguindo as idéias precedentes, começaremos com a forma desta correção no teste do parâmetro de dispersão e posteriormente nos dedicaremos aos parâmetros de regressão.

3.2.1 Teste do parâmetro de dispersão

A forma geral do coeficiente da correção de Bartlett para o teste da razão de verossimilhanças da hipótese $H_0 : \phi = \phi^0$, baseado na verossimilhança perfilada ajustada (ver (3.4)), é dado por (2.15), sendo que o termo $\xi(\vartheta)$ pode ser encontrado em Cordeiro (1983). Para a obtenção de $\xi^\dagger(\vartheta)$ observemos que se r ou s assumem valores relativos aos parâmetros de regressão, τ^{rs} se anula. Por esta razão escrevemos de (2.17)

$$\begin{aligned} \xi^\dagger(\vartheta) = & -\tau^{\phi\phi} \nu^{tv} \nu^{uw} \left(\lambda_{\phi\phi t} \lambda_{uv/w} - \frac{1}{2} \lambda_{\phi\phi t} \lambda_{uvw} - \lambda_{tu/\phi} \lambda_{\phi v/w} + \frac{1}{2} \lambda_{tu/\phi} \lambda_{\phi vw} \right. \\ & - \frac{1}{2} \lambda_{\phi tv} \lambda_{\phi vw} + \lambda_{\phi tu} \lambda_{\phi v/w} + \lambda_{\phi t/u} \lambda_{\phi vw} - \lambda_{\phi t/u} \lambda_{\phi w/v} \\ & - \lambda_{\phi t/u} \lambda_{vw/\phi} + \frac{1}{2} \lambda_{\phi tu} \lambda_{vw/\phi} + \lambda_{\phi t/v} \lambda_{\phi u/w} - \frac{1}{2} \lambda_{\phi t/v} \lambda_{\phi uw} \\ & \left. - \frac{1}{2} \lambda_{\phi tv} \lambda_{\phi u/w} + \frac{1}{4} \lambda_{\phi tv} \lambda_{\phi uw} + \lambda_{\phi t/\phi} \lambda_{uvw} - 2 \lambda_{\phi t/\phi} \lambda_{uv/w} \right) \\ & + 2\tau^{\phi\phi} \lambda^{tv} \nu^{uw} \left(\frac{1}{4} \lambda_{\phi tv} \lambda_{\phi uw} - \frac{1}{2} \lambda_{\phi v/t} \lambda_{\phi uw} - \frac{1}{2} \lambda_{\phi tv} \lambda_{\phi u/w} + \lambda_{\phi v/t} \lambda_{\phi u/w} \right) \\ & - \tau^{\phi\phi} \nu^{tu} \left(\frac{1}{2} \lambda_{\phi\phi tu} - \lambda_{\phi\phi t/u} - \lambda_{\phi tu/\phi} + 2 \lambda_{\phi t/\phi u} \right). \end{aligned}$$

Esta expressão se reduz substancialmente já que os cumulantes $\lambda_{\phi\phi t}$, $\lambda_{\phi v/w}$, $\lambda_{\phi t/u}$, $\lambda_{\phi w/v}$, $\lambda_{\phi t/v}$, $\lambda_{\phi u/w}$, $\lambda_{\phi t/\phi}$, $\lambda_{\phi\phi t/u}$, $\lambda_{\phi tu/\phi}$ e $\lambda_{\phi t/\phi u}$ se anulam. Então,

$$\begin{aligned}\xi^\dagger(\vartheta) = & -\lambda^{\phi\phi}\nu^{tv}\nu^{uw}\left(\frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} - \frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} + \frac{1}{2}\lambda_{\phi tu}\lambda_{\phi vw} + \frac{1}{4}\lambda_{\phi tv}\lambda_{\phi uw}\right) \\ & + \lambda^{\phi\phi}\lambda^{\phi\phi}\nu^{uw}\left(\frac{1}{2}\lambda_{\phi\phi\phi}\lambda_{\phi uw} - \lambda_{\phi\phi/\phi}\lambda_{\phi uw}\right) + \frac{1}{2}\lambda^{\phi\phi}\nu^{tv}\nu^{uw}\lambda_{\phi tv}\lambda_{\phi uw},\end{aligned}$$

ou seja,

$$\xi^\dagger(\vartheta) = \frac{1}{4}\lambda_{\phi\phi}\nu^{tv}\nu^{uw}\lambda_{\phi tv}\lambda_{\phi uw} - \frac{1}{2}\lambda_{\phi\phi}\nu^{tv}\nu^{uw}\lambda_{\phi tu}\lambda_{\phi vw} - \frac{1}{2}\lambda^{\phi\phi}\lambda^{\phi\phi}\nu^{uw}\lambda_{\phi\phi\phi}\lambda_{\phi uw}.$$

Observando que nesta situação $\nu^{tv}\nu^{uw}\lambda_{\phi tv}\lambda_{\phi uw} = q^2/\phi^2$, $\nu^{tv}\nu^{uw}\lambda_{\phi tu}\lambda_{\phi vw} = q/\phi^2$ e que $\nu^{uw}\lambda_{\phi uw} = q/\phi$, concluímos que

$$\xi^\dagger(\vartheta) = \frac{1}{4}\frac{q(q-2)}{nd_1''(\phi)\phi^2} - \frac{1}{2}\frac{qd_1'''(\phi)}{nd_1''(\phi)^2\phi}. \quad (3.29)$$

Substituindo as expressões de $d_1'(\phi)$, $d_1''(\phi)$ e $d_1'''(\phi)$ no modelo de regressão normal em (3.29) obtemos $\xi^\dagger(\vartheta) = -q^2/(2n) - q/n$. O termo $\xi(\vartheta)$ da correção de Bartlett para o teste usual da razão de verossimilhanças, no modelo de regressão normal, pode ser encontrado em Cordeiro & McCullagh (1991) como sendo $\xi(\vartheta) = 1/(3n) + q^2/(2n) + q/n$. Somando $\xi(\vartheta)$ com $\xi^\dagger(\vartheta)$, obtemos que $\bar{\xi}(\vartheta) = 1/(3n)$, que coincide com a expressão (2.22), representando uma verificação parcial de nossos resultados.

Ilustremos o desempenho de diversas estatísticas de teste para $H_0 : \phi = 1$ através de simulações Monte Carlo no modelo de regressão normal. Nosso objetivo é comparar as estatísticas da razão de verossimilhanças usual, da razão de verossimilhanças perfilada ajustada, a da correção de Bartlett para a estatísticas da razão de verossimilhanças usual e da razão de verossimilhanças perfilada ajustada.

Simulamos 100.000 réplicas de amostras de tamanhos 10, 15, 20 e 25 do modelo de regressão normal linear de parâmetros $\phi = 1$ e $\beta = (1, \dots, 1)^\top$, este de dimensão $p \times 1$, e a matriz de modelo X obtida da distribuição $\mathcal{U}(0, 1)$. Em cada situação programamos as funções de verossimilhança e de verossimilhança perfilada ajustada, esta última sendo uma situação particular da função obtida em (3.4), assim como as estatísticas da razão de verossimilhanças

respectivas. Todo este trabalho foi realizado usando a linguagem matricial de programação Ox (Doornik, 2001) e é apresentado nas Tabelas 3.1 e 3.2.

Dos resultados de simulação apresentados podemos concluir que o teste da razão de verossimilhanças perfiladas ajustadas apresenta taxas de rejeição mais próximas dos níveis nominais do que o teste da razão de verossimilhanças usual. Por exemplo, observemos que escolhendo $n = 10$, $q = 1$ e nível nominal de 5%, a taxa de rejeição obtida com o teste da razão de verossimilhanças usual é de 7,4%, enquanto com o teste da razão de verossimilhanças perfiladas ajustadas é de 5,7%, este, evidentemente, muito mais próximo do nível nominal escolhido.

Observamos também destes resultados de simulação que, ao aplicar a correção de Bartlett para o teste da razão de verossimilhanças usual, as taxas de rejeição estão mais próximas dos níveis nominais do que as obtidas da estatística sem esta correção. Continuando com o

Tabela 3.1: Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_{\alpha}^2(1))$, $P(\mathcal{W}^* \geq \chi_{\alpha}^2(1))$, $P(\overline{\mathcal{W}} \geq \chi_{\alpha}^2(1))$, $P(\overline{\mathcal{W}}^* \geq \chi_{\alpha}^2(1))$, $\alpha = 0,15; 0,10; 0,05; 0,01; 0,005$ e $0,0005$. Valores percentuais em 100.000 réplicas para testar $H_0 : \phi = 1$ no modelo de regressão linear normal com β sendo o vetor de parâmetros de perturbação.

	$n = 10$						$n = 15$						
	15%	10%	5%	1%	0,5%	0,05%	15%	10%	5%	1%	0,5%	0,05%	
$q = 1$	\mathcal{W}	19,0	13,4	7,4	1,8	1,0	0,10	17,6	12,1	6,5	1,5	0,8	0,10
	\mathcal{W}^*	15,4	10,3	5,2	1,0	0,5	0,05	15,2	10,1	5,1	1,0	0,5	0,05
	$\overline{\mathcal{W}}$	16,2	11,0	5,7	1,2	0,6	0,07	15,6	10,5	5,4	1,1	0,6	0,06
	$\overline{\mathcal{W}}^*$	15,5	10,4	5,3	1,1	0,5	0,06	15,2	10,1	5,2	1,0	0,5	0,05
$q = 2$	\mathcal{W}	24,9	18,7	11,3	3,4	2,0	0,40	21,4	15,5	8,9	2,5	1,4	0,20
	\mathcal{W}^*	16,6	11,1	5,5	1,0	0,5	0,05	15,8	10,5	5,2	1,0	0,5	0,04
	$\overline{\mathcal{W}}$	17,7	12,3	6,6	1,6	0,9	0,15	16,3	11,1	5,7	1,3	0,7	0,07
	$\overline{\mathcal{W}}^*$	17,0	11,7	6,1	1,4	0,8	0,13	15,9	10,8	5,5	1,2	0,6	0,06
$q = 3$	\mathcal{W}	33,4	26,5	17,8	6,7	4,4	1,02	26,6	20,0	12,4	4,0	2,4	0,40
	\mathcal{W}^*	18,7	12,6	6,2	1,1	0,5	0,04	16,6	11,1	5,4	1,0	0,5	0,03
	$\overline{\mathcal{W}}$	20,3	14,7	8,4	2,4	1,5	0,31	17,4	12,0	6,4	1,5	0,9	0,11
	$\overline{\mathcal{W}}^*$	19,6	14,1	7,9	2,2	1,3	0,27	16,9	11,6	6,1	1,4	0,8	0,10

Tabela 3.2: Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(1))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(1))$, $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(1))$, $\alpha = 0,15; 0,10; 0,05; 0,01; 0,005$ e $0,0005$. Valores percentuais em 100.000 réplicas para testar $H_0 : \phi = 1$ no modelo de regressão linear normal com β sendo o vetor de parâmetros de perturbação.

	$n = 20$						$n = 25$						
	15%	10%	5%	1%	0,5%	0,05%	15%	10%	5%	1%	0,5%	0,05%	
$q = 1$	\mathcal{W}	17,0	11,7	6,1	1,4	0,7	0,08	16,3	11,2	5,9	1,3	0,6	0,07
	\mathcal{W}^*	15,1	10,1	5,0	1,0	0,5	0,04	14,8	10,0	5,1	1,0	0,5	0,04
	$\overline{\mathcal{W}}$	15,5	10,4	5,2	1,1	0,5	0,04	15,3	10,2	5,2	1,1	0,5	0,05
	$\overline{\mathcal{W}}^*$	15,2	10,1	5,0	1,0	0,5	0,04	15,0	10,0	5,0	1,0	0,5	0,05
$q = 2$	\mathcal{W}	19,7	14,1	7,9	2,1	1,1	0,20	18,7	13,1	7,1	1,8	1,0	0,13
	\mathcal{W}^*	15,5	10,4	5,3	1,0	0,5	0,05	15,3	10,1	5,1	1,0	0,5	0,04
	$\overline{\mathcal{W}}$	15,9	10,8	5,6	1,2	0,6	0,06	15,5	10,4	5,3	1,1	0,6	0,06
	$\overline{\mathcal{W}}^*$	15,6	10,6	5,4	1,1	0,6	0,06	15,3	10,2	5,2	1,1	0,6	0,06
$q = 3$	\mathcal{W}	23,5	17,3	10,2	3,0	1,7	0,30	21,6	15,7	9,1	2,4	1,4	0,21
	\mathcal{W}^*	16,0	10,6	5,3	1,0	0,4	0,04	15,6	10,4	5,1	1,0	0,5	0,04
	$\overline{\mathcal{W}}$	16,4	11,1	5,9	1,3	0,7	0,08	15,9	10,8	5,5	1,2	0,6	0,08
	$\overline{\mathcal{W}}^*$	16,0	10,9	5,7	1,2	0,6	0,07	15,7	10,6	5,4	1,1	0,6	0,07
$q = 4$	\mathcal{W}	28,6	21,9	13,7	4,6	2,8	0,52	25,7	19,3	11,8	3,6	2,2	0,37
	\mathcal{W}^*	16,7	11,1	5,4	0,9	0,4	0,03	16,2	10,9	5,4	0,9	0,4	0,03
	$\overline{\mathcal{W}}$	17,3	11,9	6,4	1,5	0,8	0,11	16,5	11,3	5,9	1,3	0,7	0,10
	$\overline{\mathcal{W}}^*$	16,9	11,6	6,1	1,4	0,8	0,11	16,2	11,1	5,8	1,3	0,7	0,09

exemplo, se $n = 10$, $q = 1$ e nível nominal de 5% a taxa de rejeição obtida com a estatística corrigida é de 5,2%.

Quando aplicamos a correção de Bartlett à estatística da razão de verossimilhanças perfiladas ajustadas $\overline{\mathcal{W}}^*$ observamos que esta tem desempenho semelhante à estatística da razão de verossimilhanças usual corrigida \mathcal{W}^* .

3.2.2 Teste dos parâmetros de regressão

Nosso objetivo agora é obter a expressão de $\xi^\dagger(\vartheta)$ quando consideramos β_1 como sendo o vetor de parâmetros de interesse e β_2 ou (β_2, ϕ) , o vetor de parâmetros de perturbação. Lembremos

que a dimensão de β_1 é p e a de β_2 é q .

Aqui procederemos de maneira similar à situação em que obtivemos a expressão para $\Theta(\vartheta)$ no sentido de encontrar os termos que, somados, formarão $\xi^\dagger(\vartheta)$, fixando valores dos sub-índices.

Se o parâmetro de dispersão for conhecido, os sub-índices r, s, t, u, v e w somente assumirão valores correspondentes aos componentes do vetor de parâmetros da regressão, os quais sabemos que não são ortogonais.

A seguir escreveremos matricialmente cada um dos 24 termos de $\xi^\dagger(\vartheta)$. A soma destes constituirá, nesta situação, a expressão do termo $\xi^\dagger(\vartheta)$ da correção de Bartlett para o teste da razão de verossimilhanças perfilada ajustada (ver 2.15):

(1)

$$\begin{aligned}
& -\tau^{rs}\nu^{tv}\nu^{uw}\lambda_{rst}\lambda_{uv/w} = \\
& = -\phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \phi^{-1} \sum_{l,m} (f+g)_l (f+2g)_m z_{2lm} z_{2mm} (z - z_2)_{ll} \\
& = \phi^{-1} \mathbf{1}^\top (F + G) (Z_d - Z_{2d}) Z_2 Z_{2d} (F + 2G) \mathbf{1}.
\end{aligned}$$

(2)

$$\begin{aligned}
& \frac{1}{2} \tau^{rs}\nu^{tv}\nu^{uw}\lambda_{rst}\lambda_{uvw} = \\
& = \frac{1}{2} \phi^2 \sum_{l,m} (f+2g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -\frac{1}{2} \phi^{-1} \mathbf{1}^\top (F + 2G) (Z_d - Z_{2d}) Z_2 Z_{2d} (F + 2G) \mathbf{1}.
\end{aligned}$$

(3)

$$\begin{aligned}
& \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{tu/s} \lambda_{rv/w} = \\
& = \phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{mr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = -\phi^{-1} 1^\top (F+G)(Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(4)

$$\begin{aligned}
& -\frac{1}{2} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{tu/s} \lambda_{rvw} = \\
& = -\frac{1}{2} \phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{mr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = \frac{1}{2} \phi^{-1} 1^\top (F+G)(Z-Z_2) \otimes Z_2 \otimes Z_2 (F+2G) 1.
\end{aligned}$$

(5)

$$\begin{aligned}
& \frac{1}{2} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rtu} \lambda_{svw} = \\
& = \frac{1}{2} \phi^2 \sum_{l,m} (f+2g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = -\frac{1}{2} \phi^{-1} 1^\top (F+2G)(Z-Z_2) \otimes Z_2 \otimes Z_2 (F+2G) 1.
\end{aligned}$$

(6)

$$\begin{aligned}
& -\tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rtu} \lambda_{sv/w} = \\
& = -\phi^2 \sum_{l,m} (f+2g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = \phi^{-1} 1^\top (F+2G)(Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(7)

$$\begin{aligned}
& -\tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/u} \lambda_{svw} = \\
& = -\phi^2 \sum_{l,m} (f+2g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = \phi^{-1} 1^\top (F+2G)(Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(8)

$$\begin{aligned}
& \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/u} \lambda_{sw/v} = \\
& = \phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = -\phi^{-1} 1^\top (F+G) (Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(9)

$$\begin{aligned}
& \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/u} \lambda_{vw/s} = \\
& = \phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = -\phi^{-1} 1^\top (F+G) (Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(10)

$$\begin{aligned}
& -\frac{1}{2} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rtu} \lambda_{vw/s} = \\
& = -\frac{1}{2} \phi^2 \sum_{l,m} (f+2g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{lu} x_{mw} \\
& = \frac{1}{2} \phi^{-1} 1^\top (F+2G) (Z-Z_2) \otimes Z_2 \otimes Z_2 (F+G) 1.
\end{aligned}$$

(11)

$$\begin{aligned}
& -\tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/v} \lambda_{su/w} = \\
& = -\phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \phi^{-1} 1^\top (F+G) Z_{2d} (Z-Z_2) Z_{2d} (F+G) 1.
\end{aligned}$$

(12)

$$\begin{aligned}
& \frac{1}{2} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/v} \lambda_{suw} = \\
& = \frac{1}{2} \phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -\frac{1}{2} \phi^{-1} 1^\top (F+G) Z_{2d} (Z-Z_2) Z_{2d} (F+2G) 1.
\end{aligned}$$

(13)

$$\begin{aligned}
& \frac{1}{2} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rtv} \lambda_{su/w} = \\
& = \frac{1}{2} \phi^2 \sum_{l,m} (f+2g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -\frac{1}{2} \phi^{-1} 1^\top (F+2G) Z_{2d} (Z-Z_2) Z_{2d} (F+G) 1.
\end{aligned}$$

(14)

$$\begin{aligned}
& -\frac{1}{4} \tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rtv} \lambda_{suw} = \\
& = -\frac{1}{4} \phi^2 \sum_{l,m} (f+2g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \nu^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \frac{1}{4} \phi^{-1} 1^\top (F+2G) Z_{2d} (Z-Z_2) Z_{2d} (F+2G) 1.
\end{aligned}$$

(15)

$$\begin{aligned}
& -\tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/s} \lambda_{uvw} = \\
& = -\phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \phi^{-1} 1^\top (F+G) Z_2 (Z_d - Z_{2d}) Z_{2d} (F+2G) 1.
\end{aligned}$$

(16)

$$\begin{aligned}
& 2\tau^{rs} \nu^{tv} \nu^{uw} \lambda_{rt/s} \lambda_{uv/w} = \\
& = 2\phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,v} \nu^{tv} x_{lt} x_{mv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -2\phi^{-1} 1^\top (F+G) Z_2 (Z_d - Z_{2d}) Z_{2d} (F+G) 1.
\end{aligned}$$

(17)

$$\begin{aligned}
& \frac{1}{2} \tau^{rs} \lambda^{tv} \nu^{uw} \lambda_{rtv} \lambda_{suw} = \\
& = \frac{1}{2} \phi^2 \sum_{l,m} (f+2g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -\frac{1}{2} \phi^{-1} 1^\top (F+2G) Z_d (Z-Z_2) Z_{2d} (F+2G) 1.
\end{aligned}$$

(18)

$$\begin{aligned}
& -\tau^{rs}\lambda^{tv}\nu^{uw}\lambda_{rv/t}\lambda_{suw} = \\
& = -\phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \phi^{-1} \mathbf{1}^\top (F+G) Z_d (Z-Z_2) Z_{2d} (F+2G) \mathbf{1}.
\end{aligned}$$

(19)

$$\begin{aligned}
& -\tau^{rs}\lambda^{tv}\nu^{uw}\lambda_{rv/t}\lambda_{su/w} = \\
& = \phi^2 \sum_{l,m} (f+g)_l (f+2g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = \phi^{-1} \mathbf{1}^\top (F+G) Z_d (Z-Z_2) Z_{2d} (F+2G) \mathbf{1}.
\end{aligned}$$

(20)

$$\begin{aligned}
& 2\tau^{rs}\lambda^{tv}\nu^{uw}\lambda_{rv/t}\lambda_{su/w} = \\
& = 2\phi^2 \sum_{l,m} (f+g)_l (f+g)_m \sum_{r,s} \tau^{rs} x_{lr} x_{ms} \sum_{t,v} \lambda^{tv} x_{lt} x_{lv} \sum_{u,w} \nu^{uw} x_{mu} x_{mw} \\
& = -2\phi^{-1} \mathbf{1}^\top (F+G) Z_d (Z-Z_2) Z_{2d} (F+G) \mathbf{1}.
\end{aligned}$$

(21)

$$-\frac{1}{2}\tau^{rs}\lambda^{tu}\lambda_{rstu} = \frac{1}{2}\phi \sum_l e_l \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,u} \nu^{tu} x_{lt} x_{lu} = \frac{1}{2}\phi^{-1} \mathbf{1}^\top E_d (Z_d - Z_{2d}) Z_{2d} \mathbf{1}.$$

(22)

$$\tau^{rs}\nu^{tu}\lambda_{rst/u} = -\phi \sum_l (f'+2j)_l \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,u} \nu^{tu} x_{lt} x_{lu} = -\phi^{-1} \mathbf{1}^\top (F'+2J) (Z_d - Z_{2d}) Z_{2d} \mathbf{1},$$

onde $f' = df/d\eta$ e $F' = \text{diag}\{f'_1, \dots, f'_n\}$.

(23)

$$\tau^{rs}\nu^{tu}\lambda_{rtu/s} = -\phi \sum_l (f'+2j)_l \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,u} \nu^{tu} x_{lt} x_{lu} = -\phi^{-1} \mathbf{1}^\top (F'+2J) (Z_d - Z_{2d}) Z_{2d} \mathbf{1}.$$

(24)

$$-2\tau^{rs}\nu^{tu}\lambda_{rt/su} = 2\phi \sum_l (f'+j)_l \sum_{r,s} \tau^{rs} x_{lr} x_{ls} \sum_{t,u} \nu^{tu} x_{lt} x_{lu} = 2\phi^{-1} 1^\top (F'+J)(Z_d - Z_{2d}) Z_{2d} 1.$$

Somando estes termos obtemos

$$\begin{aligned} \xi^\dagger(\vartheta) &= \phi^{-1} \left[1^\top (F + 2G) \left\{ \frac{1}{4} Z_{2d} (Z - Z_2) Z_{2d} - \frac{1}{2} (Z_d - Z_{2d}) Z_2 Z_{2d} \right. \right. \\ &\quad \left. \left. - \frac{1}{2} (Z - Z_2) \otimes Z_2 \otimes Z_2 - \frac{1}{2} Z_d (Z - Z_2) Z_{2d} \right\} (F + 2G) 1 \right. \\ &\quad \left. - 1^\top G \{ Z_{2d} (Z - Z_2) Z_{2d} - 2(Z - Z_2)_d Z_2 Z_{2d} - (Z - Z_2) \otimes Z_2 \otimes Z_2 \right. \\ &\quad \left. - 2Z_d (Z - Z_2) Z_{2d} \} (F + G) 1 + \frac{1}{2} \text{tr} \{ (E - 4J)(Z_d - Z_{2d}) Z_{2d} \} \right], \end{aligned} \quad (3.30)$$

uma vez que $1^\top (E - 4J)(Z_d - Z_{2d}) Z_{2d} 1 = \text{tr} \{ (E - 4J)(Z_d - Z_{2d}) Z_{2d} \}$. Observemos que a função de cumulantes f' não aparece na expressão geral (3.30). Isto é devido a que, ao somar (22), (23) e (24), os termos onde a matriz F' aparece se anulam.

Uma situação particular interessante é considerar a função de ligação canônica. Neste caso, temos $\theta = \eta$ e então g , que pode ser escrito como $d^2\theta/d\eta^2 d\mu/d\eta$ (ver Apêndice B), é nulo e portanto $G = 0$ e $J = 0$.

Obtemos então que, se a função de ligação é canônica, $\xi^\dagger(\vartheta)$ se reduz a

$$\begin{aligned} \xi^\dagger(\vartheta) &= \phi^{-1} \left[1^\top F \left\{ \frac{1}{4} Z_{2d} (Z - Z_2) Z_{2d} - \frac{1}{2} (Z_d - Z_{2d}) Z_2 Z_{2d} \right. \right. \\ &\quad \left. \left. - \frac{1}{2} (Z - Z_2) \otimes Z_2 \otimes Z_2 - \frac{1}{2} Z_d (Z - Z_2) Z_{2d} \right\} F 1 \right. \\ &\quad \left. + \frac{1}{2} \text{tr} \{ E(Z_d - Z_{2d}) Z_{2d} \} \right]. \end{aligned}$$

Considerando agora o parâmetro de dispersão desconhecido nos deparamos com uma situação similar à da obtenção do acréscimo na função de ajuste $\Theta(\vartheta)$, abordada na Seção 3.1.4. Queremos dizer que procederemos a seguir a encontrar o acréscimo em (3.30) provocado por termos um parâmetro de perturbação a mais.

Observando a expressão de ξ^\dagger concluímos que somente os índices t , v , u e w podem assumir ϕ , já que se algum dos índices r ou s assumir o valor correspondente a ϕ , τ é zero, e então, não temos termos a adicionar.

Assim,

$$\begin{aligned}
\xi^\dagger(\vartheta) = & -\tau^{rs}\nu^{tv}\nu^{\phi\phi} \left(\lambda_{rst}\lambda_{\phi v/\phi} + \frac{1}{2}\lambda_{rst}\lambda_{\phi v\phi} - \lambda_{t\phi/s}\lambda_{rv/\phi} - \frac{1}{2}\lambda_{t\phi/s}\lambda_{rv\phi} \right. \\
& - \frac{1}{2}\lambda_{rt\phi}\lambda_{sv\phi} + \lambda_{rt\phi}\lambda_{sv/\phi} + \lambda_{rt/\phi}\lambda_{sv\phi} - \lambda_{rt/\phi}\lambda_{s\phi/v} - \lambda_{rt/\phi}\lambda_{v\phi/s} \\
& + \frac{1}{2}\lambda_{rt\phi}\lambda_{v\phi/s} + \lambda_{rt/v}\lambda_{s\phi/\phi} - \frac{1}{2}\lambda_{rt/v}\lambda_{s\phi\phi} - \frac{1}{2}\lambda_{rtv}\lambda_{s\phi/\phi} + \frac{1}{4}\lambda_{rtv}\lambda_{s\phi\phi} \\
& \left. + \lambda_{rt/s}\lambda_{\phi v\phi} - 2\lambda_{rt/s}\lambda_{\phi v/\phi} \right) \\
& -\tau^{rs}\nu^{\phi\phi}\nu^{uw} \left(\lambda_{rs\phi}\lambda_{u\phi/w} - \frac{1}{2}\lambda_{rs\phi}\lambda_{u\phi w} - \lambda_{\phi u/s}\lambda_{r\phi/w} + \frac{1}{2}\lambda_{\phi u/s}\lambda_{r\phi w} \right. \\
& - \frac{1}{2}\lambda_{r\phi u}\lambda_{s\phi w} + \lambda_{r\phi u}\lambda_{s\phi/w} + \lambda_{r\phi/u}\lambda_{s\phi w} - \lambda_{r\phi/u}\lambda_{sw/\phi} - \lambda_{r\phi/u}\lambda_{\phi w/s} \\
& + \frac{1}{2}\lambda_{r\phi u}\lambda_{\phi w/s} + \lambda_{r\phi/\phi}\lambda_{su/w} - \frac{1}{2}\lambda_{r\phi/\phi}\lambda_{suw} - \frac{1}{2}\lambda_{r\phi\phi}\lambda_{su/w} + \frac{1}{4}\lambda_{r\phi\phi}\lambda_{suw} \\
& \left. + \lambda_{r\phi/s}\lambda_{u\phi w} - 2\lambda_{r\phi/s}\lambda_{u\phi/w} \right) \\
& -\tau^{rs}\nu^{\phi\phi}\nu^{\phi\phi} \left(\lambda_{rs\phi}\lambda_{\phi\phi/\phi} - \frac{1}{2}\lambda_{rs\phi}\lambda_{\phi\phi\phi} - \lambda_{\phi\phi/s}\lambda_{r\phi/\phi} + \frac{1}{2}\lambda_{\phi\phi/s}\lambda_{r\phi\phi} \right. \\
& - \frac{1}{2}\lambda_{r\phi\phi}\lambda_{s\phi\phi} + \lambda_{r\phi\phi}\lambda_{s\phi/\phi} + \lambda_{r\phi/\phi}\lambda_{s\phi\phi} - \lambda_{r\phi/\phi}\lambda_{s\phi/\phi} - \lambda_{r\phi/\phi}\lambda_{\phi\phi/s} \\
& + \frac{1}{2}\lambda_{r\phi\phi}\lambda_{\phi\phi/s} + \lambda_{r\phi/\phi}\lambda_{s\phi/\phi} - \frac{1}{2}\lambda_{r\phi/\phi}\lambda_{s\phi\phi} - \frac{1}{2}\lambda_{r\phi\phi}\lambda_{s\phi/\phi} + \frac{1}{4}\lambda_{r\phi\phi}\lambda_{s\phi\phi} \\
& \left. + \lambda_{r\phi/s}\lambda_{\phi\phi\phi} - 2\lambda_{r\phi/s}\lambda_{\phi\phi/\phi} \right) \\
& + 2\tau^{rs}\lambda^{tv}\nu^{\phi\phi} \left(\lambda_{rv/t}\lambda_{s\phi/\phi} - \frac{1}{2}\lambda_{rv/t}\lambda_{s\phi\phi} - \frac{1}{2}\lambda_{rtv}\lambda_{s\phi/\phi} + \frac{1}{4}\lambda_{rtv}\lambda_{s\phi\phi} \right) \\
& + 2\tau^{rs}\lambda^{\phi\phi}\nu^{uw} \left(\lambda_{r\phi/\phi}\lambda_{su/\phi} - \frac{1}{2}\lambda_{r\phi/\phi}\lambda_{suw} - \frac{1}{2}\lambda_{r\phi\phi}\lambda_{su/w} + \frac{1}{4}\lambda_{r\phi\phi}\lambda_{suw} \right) \\
& + 2\tau^{rs}\lambda^{\phi\phi}\nu^{\phi\phi} \left(\lambda_{r\phi/\phi}\lambda_{s\phi/\phi} - \frac{1}{2}\lambda_{r\phi/\phi}\lambda_{s\phi\phi} - \frac{1}{2}\lambda_{r\phi\phi}\lambda_{s\phi/\phi} + \frac{1}{4}\lambda_{r\phi\phi}\lambda_{s\phi\phi} \right) \\
& - \tau^{rs}\nu^{\phi\phi} \left(\frac{1}{2}\lambda_{rs\phi\phi} - \lambda_{rs\phi/\phi} - \lambda_{r\phi\phi/s} + 2\lambda_{r\phi/s\phi} \right).
\end{aligned}$$

Eliminando os cumulantes que se anulam, obtemos

$$\begin{aligned}\xi^\dagger(\vartheta) = & -\tau^{rs}\nu^{tv}\nu^{\phi\phi}\left(-\frac{1}{2}\lambda_{rt\phi}\lambda_{sv\phi} + \lambda_{rt\phi}\lambda_{sv/\phi} + \lambda_{rt/\phi}\lambda_{sv\phi}\right) \\ & -\tau^{rs}\nu^{\phi\phi}\nu^{uw}\left(-\frac{1}{2}\lambda_{rs\phi}\lambda_{u\phi w} - \frac{1}{2}\lambda_{r\phi u}\lambda_{s\phi w}\right) - \tau^{rs}\nu^{\phi\phi}\nu^{\phi\phi}\left(\lambda_{rs\phi}\lambda_{\phi\phi/\phi} - \frac{1}{2}\lambda_{rs\phi}\lambda_{\phi\phi\phi}\right).\end{aligned}$$

Lembrando que $\lambda_{sv/\phi} = \lambda_{sv\phi}$, $\lambda_{rt/\phi} = \lambda_{rt\phi}$ e que $\lambda_{\phi\phi/\phi} = \lambda_{\phi\phi\phi}$ e, agrupando adequadamente os termos na expressão anterior, temos

$$\xi^\dagger(\vartheta) = -\tau^{rs}\nu^{tv}\nu^{\phi\phi}\lambda_{\phi rt}\lambda_{\phi sv} + \frac{1}{2}\tau^{rs}\nu^{\phi\phi}\nu^{uw}\lambda_{\phi rs}\lambda_{\phi uw} - \frac{1}{2}\tau^{rs}\nu^{\phi\phi}\nu^{\phi\phi}\lambda_{\phi rs}\lambda_{\phi\phi\phi}. \quad (3.31)$$

Mas

$$\begin{aligned}\tau^{rs}\lambda_{\phi rs} &= -\sum_{r,s}\tau^{rs}\sum_l w_l x_{lr}x_{ls} = -\sum_l w_l \sum_{r,s}\tau^{rs}x_{lr}x_{ls} \\ &= \phi^{-1}\sum_l w_l(z_{ll} - z_{2ll}) = \phi^{-1}\text{tr}\{W(Z - Z_2)\},\end{aligned}$$

e dado que $\text{tr}\{W(Z - Z_2)\} = p$ obtemos $\tau^{rs}\lambda_{\phi rs} = p/\phi$. Então,

$$-\frac{1}{2}\tau^{rs}\nu^{\phi\phi}\nu^{\phi\phi}\lambda_{\phi rs}\lambda_{\phi\phi\phi} = -\frac{1}{2}\frac{pd_1'''(\phi)}{n\phi d_1''(\phi)^2}.$$

Por outro lado,

$$\begin{aligned}\tau^{rs}\nu^{tv}\lambda_{\phi rt}\lambda_{\phi sv} &= \sum_{l,m} w_l w_m \sum_{r,s}\tau^{rs}x_{lr}x_{ms} \sum_{t,v}\nu^{tv}x_{lt}x_{mv} = \\ &= \phi^{-2}\mathbf{1}^\top W(Z - Z_2)Z_2W\mathbf{1} = 0\end{aligned}$$

e

$$\begin{aligned}\tau^{rs}\nu^{uw}\lambda_{\phi rs}\lambda_{\phi uw} &= \sum_{l,m} w_l w_m \sum_{r,s}\tau^{rs}x_{lr}x_{ls} \sum_{t,v}\nu^{uw}x_{mu}x_{mv} = \\ &= \phi^{-2}\text{tr}\{W(Z - Z_2)\}\text{tr}\{WZ_2\} = \phi^{-2}pq.\end{aligned}$$

Então,

$$\frac{1}{2}\tau^{rs}\nu^{uw}\nu^{\phi\phi}\lambda_{\phi rs}\lambda_{\phi uw} = \frac{1}{2}\frac{pq}{n\phi^2 d_1''(\phi)}.$$

Podemos afirmar que o acréscimo na correção de Bartlett para a estatística ajustada de Stern nos modelos lineares generalizados quando o parâmetro de dispersão é desconhecido assume a forma

$$\frac{pq}{2n\phi^2 d_1''(\phi)} - \frac{1}{2}\frac{pd_1'''(\phi)}{n\phi d_1''(\phi)^2}.$$

Portanto, nesta situação

$$\begin{aligned}
\xi^\dagger(\vartheta) &= \phi^{-1} \left[1^\top (F + 2G) \left\{ \frac{1}{4} Z_{2d}(Z - Z_2)Z_{2d} - \frac{1}{2} (Z_d - Z_{2d})Z_2Z_{2d} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} (Z - Z_2) \otimes Z_2 \otimes Z_2 - \frac{1}{2} Z_d(Z - Z_2)Z_{2d} \right\} (F + 2G)1 \right. \\
&\quad \left. - 1^\top G \{ Z_{2d}(Z - Z_2)Z_{2d} - 2(Z - Z_2)_d Z_2 Z_{2d} - (Z - Z_2) \otimes Z_2 \otimes Z_2 \right. \\
&\quad \left. - 2Z_d(Z - Z_2)Z_{2d} \} (F + G)1 + \frac{1}{2} \text{tr}\{(E - 4J)(Z_d - Z_{2d})Z_{2d}\} \right] \\
&\quad + \frac{pq}{2n\phi^2 d_1''(\phi)} - \frac{1}{2} \frac{pd_1'''(\phi)}{n\phi d_1''(\phi)^2}.
\end{aligned} \tag{3.32}$$

3.2.3 Exemplo: modelo de regressão normal

Consideremos novamente o modelo da Seção 3.1.2, que é amplamente utilizado em situações práticas. Por esta razão e pelo fato de podermos obter expressões fechadas para os termos do ajuste para a verossimilhança perfilada, voltamos a ele como exemplo de modelos de regressão.

Como situação particular dos modelos lineares generalizados vamos considerar a função de ligação canônica. Isto implica uma grande simplificação nos resultados já que os cumulantes necessários apresentam expressões simplificadas devido, fundamentalmente, a que $g = 0$.

A hipótese em questão é $H_0 : \beta_1 = \beta_1^0$, onde $\beta_1^{(0)} = (1, 0, \dots, 0)^\top$, de dimensão $p \times 1$, ou seja, estamos interessados em verificar se o intercepto do modelo de regressão é 1 e se os restantes parâmetros de interesse são zero. Desta forma o preditor linear sob a hipótese nula é da forma

$$\eta_l = 1 + \beta^{p+1} x_{l,p+1} + \dots + \beta^{p+q} x_{l,p+q},$$

e, sob a hipótese alternativa,

$$\eta_l = \beta^1 + \beta^2 x_{l,2} + \dots + \beta^p x_{l,p} + \beta^{p+1} x_{l,p+1} + \dots + \beta^{p+q} x_{l,p+q},$$

para $l = 1, \dots, n$. Identificamos, então, que o vetor de parâmetros de perturbação é constituído pelos parâmetros de regressão $\beta^{p+1}, \dots, \beta^{p+q}$ e pelo parâmetro de dispersão ϕ .

Nas Tabelas 3.3 e 3.4 apresentamos o resultado de simulações de Monte Carlo com o objetivo de comparar o desempenho dos testes baseados nas estatísticas da razão de verossimilhanças, na razão de verossimilhanças perfiladas ajustadas e nas correspondentes estatísticas com correção de Bartlett para testar a hipótese $H_0 : \beta_1 = \beta_1^{(0)}$.

Encontraremos a seguir as expressões das funções de ajuste $\Lambda(\vartheta)$ e $\Theta(\vartheta)$ e do termo $\xi^\dagger(\vartheta)$, da correção de Bartlett. A forma da função de ajuste $\Lambda(\vartheta)$ foi obtida em (3.7) para o caso de ϕ conhecido e, como foi comentado na Seção 3.1.4, sua expressão não é afetada pelo desconhecimento da dispersão. Agora, a função de ligação canônica no modelo de regressão normal é a identidade e, portanto, $F = G = 0$ e $W = V = T = I$ e, então, $\Lambda(\vartheta) = 0$.

Tabela 3.3: Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$, valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com ϕ como o parâmetro de perturbação ($\phi = 1$) no modelo de regressão linear normal, número de parâmetros de interesse $p = 1, 2, 3$ e 4 e tamanho de amostra $n = 10$ e 15.

		$n = 10$					$n = 15$				
		15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%
$p = 1$	\mathcal{W}	18,3	12,8	7,0	1,8	1,0	17,3	12,1	6,4	1,5	0,8
	\mathcal{W}^*	15,3	10,3	5,3	1,1	0,6	15,3	10,3	5,2	1,1	0,6
	$\overline{\mathcal{W}}$	14,3	9,5	4,9	1,0	0,6	14,6	9,7	4,9	1,0	0,5
	$\overline{\mathcal{W}}^*$	15,2	10,4	5,4	1,3	0,7	15,2	10,3	5,3	1,2	0,6
$p = 2$	\mathcal{W}	21,9	15,8	9,1	2,5	1,4	19,4	13,7	7,5	1,9	1,0
	\mathcal{W}^*	16,1	10,9	5,7	1,2	0,6	15,6	10,5	5,3	1,1	0,6
	$\overline{\mathcal{W}}$	16,3	11,3	6,2	1,5	0,8	15,7	10,6	5,5	1,2	0,6
	$\overline{\mathcal{W}}^*$	13,7	9,2	4,6	1,0	0,5	15,7	10,6	5,5	1,2	0,6
$p = 3$	\mathcal{W}	26,7	19,9	12,1	3,8	2,3	21,9	15,8	9,0	2,4	1,4
	\mathcal{W}^*	17,5	12,1	6,4	1,4	0,8	16,1	10,9	5,5	1,2	0,6
	$\overline{\mathcal{W}}$	19,9	14,3	8,2	2,3	1,4	17,4	12,1	6,5	1,6	0,8
	$\overline{\mathcal{W}}^*$	14,1	9,5	4,9	1,1	0,6	13,6	9,0	4,4	0,9	0,4
$p = 4$	\mathcal{W}	-	-	-	-	-	25,1	18,5	11,0	3,3	1,9
	\mathcal{W}^*	-	-	-	-	-	16,8	11,5	6,0	1,3	0,7
	$\overline{\mathcal{W}}$	-	-	-	-	-	19,9	14,2	8,1	2,1	1,2
	$\overline{\mathcal{W}}^*$	-	-	-	-	-	13,8	9,3	4,6	1,0	0,5

Com relação à função $\Theta(\vartheta)$ dada em (3.22), vemos que considerar a função de ligação identidade implica que esta se reduz a

$$\Theta(\vartheta) = \frac{1}{4n} \left\{ \frac{d_1'''(\phi)}{d_1''(\phi)^2} - \frac{q}{\phi d_1''(\phi)} \right\} \text{tr}\{H\} + \frac{1}{2n\phi d_1''(\phi)} \text{tr}\{Z_2 H\},$$

já que também $E = J = 0$.

Por outro lado, no modelo de regressão normal, sabemos que $d_1''(\phi) = -1/2\phi^2$ e $d_1'''(\phi) = 1/\phi^3$. Então,

$$\Theta(\vartheta) = \frac{\phi}{2n} (q + 2) \text{tr}\{H\} - \frac{\phi}{n} \text{tr}\{Z_2 H\},$$

e dado que $\text{tr}\{Z_2 H\} = \text{tr}\{X_2(X_2^\top X_2)^{-1} X_2^\top H\} = \text{tr}\{X_2^\top X_2(X_2^\top X_2)^{-1} H\} = \text{tr}\{H\}$, temos

$$\Theta(\vartheta) = \frac{q\phi}{2n} \text{tr}\{H\}.$$

Com relação à correção de Bartlett para a estatística do teste da razão de verossimilhanças perfiladas ajustadas, temos que $\xi^\dagger(\vartheta)$ em (3.32) se reduz a

$$\xi^\dagger(\vartheta) = \frac{pq}{2n\phi^2 d_1''(\phi)} - \frac{1}{2} \frac{p d_1'''(\phi)}{n\phi d_1''(\phi)^2},$$

pelo fato da função de ligação canônica ser a identidade e, no modelo normal,

$$\xi^\dagger(\vartheta) = \frac{p(q-2)}{n},$$

obtendo-se também que $\xi(\vartheta) = (p-q)/n$.

Utilizamos novamente a linguagem matricial de programação Ox (Doornik, 2001) para gerar 10.000 réplicas do modelo de regressão normal sob a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$, com $\phi = 1$ e $\beta_2 = (1, \dots, 1)^\top$, de dimensão $q \times 1$. O tamanho da amostra em cada réplica foi escolhido com $n = 10, 15, 20$ e $n = 25$ e o número de parâmetros de interesse e perturbação entre 1 e 5. Os valores de $x_{i,2}, \dots, x_{i,q}$ foram obtidos através de simulações independentes da distribuição $\chi^2(2)$.

Observamos destes resultados de simulação que, como esperado, quanto maior o número de parâmetros a estimar pior é o desempenho dos testes, sendo necessário o uso de correções. Quando comparamos as estatísticas \mathcal{W} e $\overline{\mathcal{W}}$ percebemos a importância da utilização de funções de verossimilhanças ajustadas, embora os comportamentos das correspondentes estatísticas com correção de Bartlett sejam similares.

Tabela 3.4: Estimativas de Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$, valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com ϕ como o parâmetro de perturbação ($\phi = 1$) no modelo de regressão linear normal, número de parâmetros de interesse $p = 1, 2, 3, 4$ e 5 e tamanho de amostra $n = 20$ e $n = 25$.

		$n = 20$					$n = 25$				
		15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%
$p = 1$	\mathcal{W}	16,5	11,3	5,9	1,4	0,7	16,2	11,0	5,8	1,3	0,7
	\mathcal{W}^*	15,0	10,1	5,1	1,1	0,5	15,0	10,0	5,1	1,0	0,5
	$\overline{\mathcal{W}}$	14,4	9,6	4,9	1,0	0,5	14,5	9,7	5,0	1,0	0,5
	$\overline{\mathcal{W}}^*$	14,9	10,1	5,1	1,1	0,5	15,0	10,0	5,1	1,1	0,5
$p = 2$	\mathcal{W}	18,2	12,6	6,8	1,6	0,8	17,3	12,0	6,3	1,4	0,8
	\mathcal{W}^*	15,4	10,3	5,3	1,0	0,5	15,0	10,1	5,1	1,0	0,5
	$\overline{\mathcal{W}}$	15,4	10,3	5,3	1,1	0,5	15,0	10,1	5,1	1,1	0,5
	$\overline{\mathcal{W}}^*$	14,1	9,3	4,6	0,9	0,4	14,0	9,2	4,5	0,9	0,4
$p = 3$	\mathcal{W}	20,0	14,1	7,7	2,0	1,1	18,7	13,1	7,0	1,6	0,9
	\mathcal{W}^*	15,6	10,5	5,3	1,1	0,6	15,2	10,2	5,1	1,0	0,5
	$\overline{\mathcal{W}}$	16,6	11,3	5,9	1,4	0,7	16,0	10,8	5,6	1,2	0,6
	$\overline{\mathcal{W}}^*$	13,6	8,9	4,4	0,9	0,4	13,6	8,9	4,4	0,8	0,4
$p = 4$	\mathcal{W}	22,1	15,8	9,0	2,4	1,3	20,3	14,3	7,9	2,0	1,2
	\mathcal{W}^*	16,0	10,8	5,5	1,2	0,6	15,4	10,4	5,2	1,1	0,6
	$\overline{\mathcal{W}}$	18,1	12,7	6,8	1,6	0,9	17,2	11,8	6,2	1,5	0,8
	$\overline{\mathcal{W}}^*$	13,5	9,0	4,5	0,9	0,4	13,5	8,9	4,4	0,9	0,4
$p = 5$	\mathcal{W}	-	-	-	-	-	22,7	16,3	9,2	2,5	1,4
	\mathcal{W}^*	-	-	-	-	-	16,3	10,9	5,6	1,2	0,6
	$\overline{\mathcal{W}}$	-	-	-	-	-	19,1	13,3	7,2	1,8	1,0
	$\overline{\mathcal{W}}^*$	-	-	-	-	-	16,4	11,1	5,8	1,3	0,7

3.2.4 Exemplo: modelo de regressão gama

Um outro exemplo dentro dos modelos lineares generalizados é considerar a variável resposta com distribuição gama de média μ e coeficiente de variação ϕ^{-1} . A função de densidade é dada por

$$f(y; \mu, \phi) = \exp \left[\phi \left(-\frac{y}{\mu} + \log \frac{1}{\mu} \right) - \log \Gamma(\phi) + \phi \log(\phi y) - \log y \right],$$

onde $y \geq 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi)$ a função gama. A densidade exponencial é um caso especial quando $\phi = 1$. Para $\phi > 1$ a densidade assume zero na origem, tem um máximo em $y = \mu - \mu/\phi$ e depois decresce. A distribuição $\chi^2(k)$ é um outro caso especial quando $\phi = k/2$ e $\mu = k$. Note que $\phi = E\{Y\}^2/\text{var}\{Y\}$ é o inverso do coeficiente de variação de Y ao quadrado. A função de variância da distribuição gama é dada por $V(\mu) = \mu^2$.

Similarmente ao exemplo anterior, interessa-nos uma parte do vetor de parâmetros de regressão. Assim, a hipótese a ser testada continua sendo $H_0 : \beta_1 = \beta_1^0$, onde $\beta_1^{(0)} = (1, \dots, 0)^\top$. O objetivo com este exemplo é avaliar, através de simulações de Monte Carlo, o comportamento do teste da razão de verossimilhanças, do teste da razão de verossimilhanças perfiladas ajustadas, e dos correspondentes testes corrigidos, via correção de Bartlett. Os resultados das simulações são apresentados nas Tabelas 3.5, 3.6, 3.7 e 3.8.

Tabela 3.5: Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4, número de parâmetros de perturbação $q = 2$ e 3, $\phi = 1$ e tamanho de amostra $n = 15$.

	$q = 2$					$q = 3$					
	15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%	
$p = 2$	\mathcal{W}	21,8	15,7	8,9	2,4	1,29	23,1	17,2	10,2	3,2	1,90
	\mathcal{W}^*	16,0	10,6	5,2	1,1	0,43	17,6	12,3	6,6	1,6	0,88
	$\overline{\mathcal{W}}$	18,2	12,6	6,6	1,5	0,83	20,0	14,3	8,1	2,2	1,38
	$\overline{\mathcal{W}}^*$	16,1	10,6	5,1	1,1	0,59	18,0	12,7	6,8	1,7	0,97
$p = 3$	\mathcal{W}	23,3	17,5	10,2	3,2	1,83	26,6	20,5	13,0	4,2	2,75
	\mathcal{W}^*	16,2	10,8	5,6	1,3	0,74	18,2	12,9	6,9	1,7	1,13
	$\overline{\mathcal{W}}$	19,6	14,2	7,8	2,2	1,25	21,2	15,6	9,1	2,7	1,67
	$\overline{\mathcal{W}}^*$	16,0	10,7	5,6	1,4	0,77	18,4	13,2	7,3	2,1	1,28
$p = 4$	\mathcal{W}	27,1	20,4	12,8	4,3	2,55	32,4	25,1	16,5	6,1	4,05
	\mathcal{W}^*	16,4	11,2	6,1	1,4	0,75	18,4	12,5	7,1	1,9	1,14
	$\overline{\mathcal{W}}$	21,8	16,1	9,4	2,7	1,73	26,6	20,1	12,4	4,3	2,79
	$\overline{\mathcal{W}}^*$	16,1	11,0	6,0	1,6	0,85	18,3	13,0	7,4	2,1	1,26

Utilizamos a linguagem de programação matricial Ox (Doornik, 2001) para gerar 10.000

réplicas do modelo de regressão gama sob a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$, com $\phi = 1$ e $\beta_2 = (1, \dots, 1)^T$, de dimensão $q \times 1$. O tamanho da amostra em cada réplica foi escolhido com $n = 15, 20, 25$ e 30 . O número de parâmetros de interesse sendo $p = 2, 3$ e 4 e o número de parâmetros de perturbação sendo $q = 2, 3, 4$ e 5 .

Para gerar as amostras do modelo de regressão, utilizamos a matriz de modelo X_1 obtida de amostras independentes da distribuição $\mathcal{U}(0, 1)$ e a matriz X_2 foi obtida de amostras independentes da distribuição $\chi^2(2)$, exceto a primeira coluna de uns, correspondente ao intercepto no preditor linear.

Para obter os resultados de simulação utilizamos o método de penalização (Luenberger, 1973), que é um procedimento para aproximar problemas de otimização com restrições a problemas de otimização sem restrições. Para maximizar as funções de verossimilhança e verossimilhança perfilada ajustada, dispomos em Ox da função MaxBFGS que nos permite maximizar funções multidimensionais sem restrições, através do método Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Fletcher, 1987).

Na Tabela 3.5 podemos observar o comportamento das estatísticas de teste quando o tamanho da amostra é $n = 15$. Os resultados mostrados são as médias amostrais das taxas de rejeição em 10.000 réplicas Monte Carlo. O comportamento destes testes, com tamanhos de amostra maiores, é apresentado nas Tabelas 3.6, 3.7 e 3.8. Todos os resultados de simulação apresentados nesta seção correspondem ao modelo de regressão gama com função de ligação canônica, ou seja, função de ligação inversa.

Com o objetivo de visualizar melhor os resultados obtidos e apresentados nestas tabelas, escolhemos duas situações, a saber $p = 4$, $q = 3$, com $n = 20$ e $n = 25$. Construímos então gráficos das curvas das discrepâncias de quantis de cada teste, mostrados nas Figuras 3.2 e 3.3.

Nestas figuras apresentamos a diferença entre os quantis amostrais sob H_0 e o valor nominal segundo a distribuição de referência. Portanto, quanto mais perto esteja cada curva da linha horizontal em zero, menor será a discrepância entre a taxa de rejeição média do teste e o nível nominal calculado da distribuição de referência.

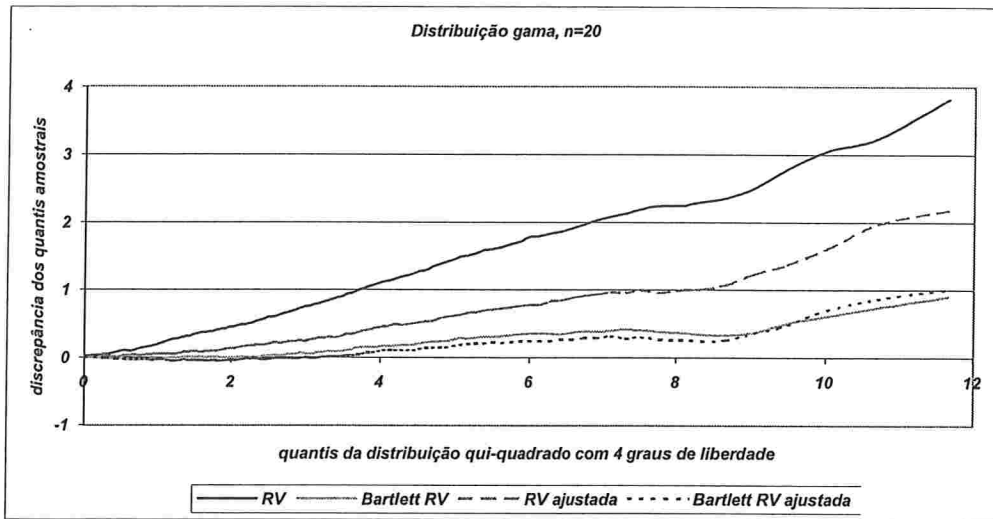


Figura 3.2: Média, em cem amostras, da discrepância dos quantis amostrais das estatísticas da razão de verossimilhanças perfiladas, perfiladas ajustadas e as correspondentes estatísticas com correção de Bartlett no modelo de regressão gama. Vetor de parâmetros de interesse $\beta = (0, 0, 0, 0)$ ($p = 4$), $n = 20$ e parâmetros de perturbação $\beta = (1, 1, 1)$ e $\phi = 1$, estes desconhecidos.

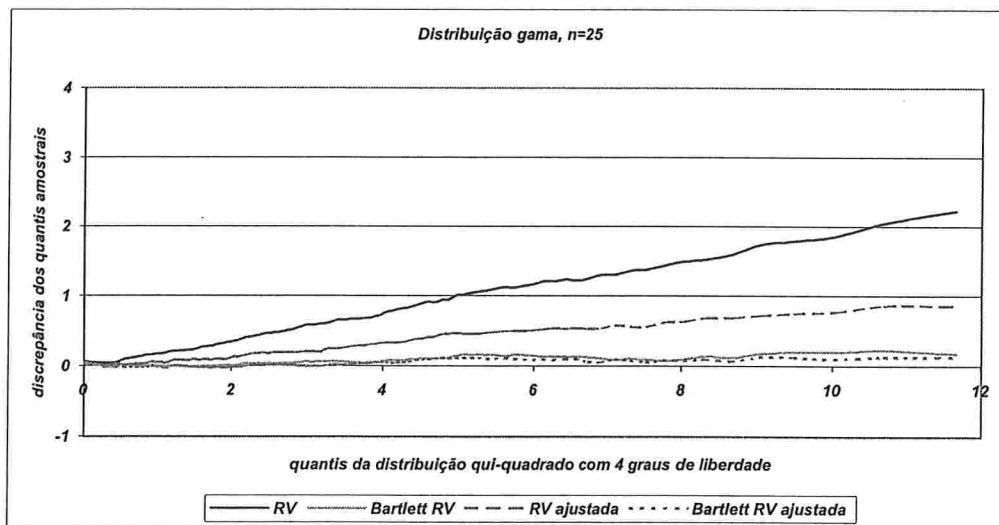


Figura 3.3: Média, em cem amostras, da discrepância dos quantis amostrais das estatísticas da razão de verossimilhanças perfiladas, perfiladas ajustadas e as correspondentes estatísticas com correção de Bartlett no modelo de regressão gama. Vetor de parâmetros de interesse $\beta = (0, 0, 0, 0)$ ($p = 4$), $n = 25$ e vetor de parâmetros de regressão $\beta = (1, 1, 1)$ e $\phi = 1$, estes desconhecidos.

Tabela 3.6: Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4 , número de parâmetros de perturbação $q = 2, 3, 4$ e 5 , $\phi = 1$ e tamanho de amostra $n = 20$.

		$q = 2$					$q = 3$				
		15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%
$p = 2$	\mathcal{W}	18,3	12,5	6,8	1,5	0,74	19,7	14,1	7,9	2,2	1,30
	\mathcal{W}^*	14,9	10,1	5,0	0,8	0,47	16,5	11,5	5,9	1,5	0,80
	$\overline{\mathcal{W}}$	15,5	10,4	5,2	1,0	0,57	16,7	11,5	6,2	1,6	0,94
	$\overline{\mathcal{W}}^*$	14,7	9,9	4,9	0,9	0,48	16,4	11,2	6,0	1,5	0,88
		$q = 4$					$q = 5$				
$p = 2$	\mathcal{W}	20,8	15,0	8,7	2,5	1,46	24,8	18,5	11,1	3,5	1,97
	\mathcal{W}^*	16,9	11,8	6,5	1,5	0,70	18,4	12,8	6,7	1,5	0,88
	$\overline{\mathcal{W}}$	18,0	12,7	6,9	1,8	0,94	15,7	10,9	5,6	1,5	0,81
	$\overline{\mathcal{W}}^*$	17,0	12,0	6,5	1,6	0,78	17,5	12,0	6,5	1,7	0,99
		$q = 2$					$q = 3$				
$p = 3$	\mathcal{W}	19,1	13,7	7,8	1,8	1,00	23,7	17,5	10,6	3,2	1,86
	\mathcal{W}^*	14,3	9,8	5,0	0,8	0,42	17,0	11,9	6,5	1,4	0,73
	$\overline{\mathcal{W}}$	16,8	11,9	6,4	1,4	0,69	17,5	12,3	7,0	1,7	0,97
	$\overline{\mathcal{W}}^*$	14,3	9,6	5,2	0,9	0,41	16,7	11,7	6,7	1,5	0,84
		$q = 4$					$q = 5$				
$p = 3$	\mathcal{W}	25,2	18,7	11,1	3,1	1,83	26,8	20,2	12,5	4,2	2,88
	\mathcal{W}^*	17,5	12,2	6,2	1,4	0,70	18,3	12,7	7,1	1,9	1,06
	$\overline{\mathcal{W}}$	20,3	14,5	7,9	2,0	1,16	19,1	13,8	7,8	2,5	1,50
	$\overline{\mathcal{W}}^*$	17,5	12,2	6,4	1,5	0,78	17,8	12,7	7,1	2,2	1,25
		$q = 2$					$q = 3$				
$p = 4$	\mathcal{W}	21,7	15,5	8,7	2,4	1,40	26,5	19,9	11,9	3,7	2,37
	\mathcal{W}^*	14,2	9,4	5,0	1,0	0,54	17,2	11,8	5,9	1,5	0,91
	$\overline{\mathcal{W}}$	18,7	12,8	7,3	1,9	1,03	20,2	14,4	7,6	2,5	1,42
	$\overline{\mathcal{W}}^*$	14,2	9,5	4,9	1,1	0,62	16,7	11,3	5,8	1,6	0,98

Uma primeira observação destas figuras é a necessidade de ajustar a função de verossimilhança perfilada para uma melhor construção de testes para os parâmetros de interesse, assim como também percebemos a importância da correção de Bartlett, tanto para o teste usual da razão de verossimilhanças como para o teste correspondente construído de verossi-

Tabela 3.7: Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 1, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4, número de parâmetros de perturbação $q = 2, 3, 4$ e 5, $\phi = 1$ e tamanho de amostra $n = 25$.

		$q = 2$					$q = 3$				
		15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%
$p = 2$	\mathcal{W}	16,6	11,2	5,9	1,3	0,64	18,3	12,8	6,9	1,8	0,90
	\mathcal{W}^*	15,2	10,2	5,3	1,1	0,49	15,8	10,8	5,5	1,3	0,50
	$\overline{\mathcal{W}}$	16,0	10,5	5,5	1,2	0,56	15,7	10,8	5,5	1,3	0,59
	$\overline{\mathcal{W}}^*$	15,5	10,1	5,4	1,1	0,55	15,9	11,0	5,6	1,3	0,60
		$q = 4$					$q = 5$				
$p = 2$	\mathcal{W}	20,1	14,2	7,9	2,1	1,22	21,4	15,5	9,0	2,4	1,34
	\mathcal{W}^*	16,7	11,4	5,9	1,4	0,78	16,8	11,9	6,5	1,3	0,79
	$\overline{\mathcal{W}}$	17,9	12,2	6,7	1,7	0,99	15,4	10,6	5,4	1,1	0,62
	$\overline{\mathcal{W}}^*$	16,6	11,4	6,2	1,5	0,84	16,6	11,7	6,1	1,5	0,76
		$q = 2$					$q = 3$				
$p = 3$	\mathcal{W}	18,2	12,6	6,8	1,7	0,85	21,2	15,4	8,5	2,1	1,17
	\mathcal{W}^*	14,7	9,5	5,1	0,9	0,56	16,5	11,2	5,7	1,1	0,64
	$\overline{\mathcal{W}}$	17,0	11,5	6,1	1,4	0,70	17,3	11,7	6,3	1,3	0,71
	$\overline{\mathcal{W}}^*$	14,9	9,6	5,2	1,0	0,53	16,6	11,0	5,9	1,2	0,65
		$q = 4$					$q = 5$				
$p = 3$	\mathcal{W}	22,5	16,0	9,1	2,4	1,39	22,7	16,4	9,6	2,7	1,62
	\mathcal{W}^*	17,1	11,2	5,9	1,3	0,57	16,9	11,5	5,9	1,4	0,73
	$\overline{\mathcal{W}}$	18,2	12,4	6,6	1,6	0,87	18,5	13,1	7,0	1,9	1,05
	$\overline{\mathcal{W}}^*$	16,9	11,4	6,0	1,3	0,59	17,1	11,6	6,2	1,6	0,79
		$q = 2$					$q = 3$				
$p = 4$	\mathcal{W}	20,0	14,3	8,1	1,8	1,03	22,9	16,2	9,1	2,4	1,29
	\mathcal{W}^*	14,8	10,1	5,0	0,9	0,36	15,8	10,3	5,3	1,0	0,49
	$\overline{\mathcal{W}}$	17,9	12,5	6,8	1,5	0,73	18,3	12,4	6,6	1,5	0,79
	$\overline{\mathcal{W}}^*$	14,8	10,3	5,2	1,0	0,41	15,5	10,3	5,3	1,1	0,63
		$q = 4$					$q = 5$				
$p = 4$	\mathcal{W}	24,7	17,9	10,5	3,1	1,87	-	-	-	-	-
	\mathcal{W}^*	16,9	11,5	6,1	1,3	0,61	-	-	-	-	-
	$\overline{\mathcal{W}}$	20,0	14,0	7,8	2,2	1,08	-	-	-	-	-
	$\overline{\mathcal{W}}^*$	16,9	11,4	6,0	1,4	0,60	-	-	-	-	-

Tabela 3.8: Estimativas Monte Carlo de $P(\mathcal{W} \geq \chi_\alpha^2(p))$, $P(\mathcal{W}^* \geq \chi_\alpha^2(p))$, $P(\overline{\mathcal{W}} \geq \chi_\alpha^2(p))$ e $P(\overline{\mathcal{W}}^* \geq \chi_\alpha^2(p))$. Valores percentuais em 10.000 réplicas para testar $H_0 : \beta^1 = 0, \dots, \beta^p = 0$ através da função de verossimilhança perfilada com $\beta^{p+1}, \dots, \beta^{p+q}$, como o vetor de parâmetros de perturbação no modelo de regressão gama, número de parâmetros de interesse $p = 2, 3$ e 4, número de parâmetros de perturbação $q = 2, 3, 4$ e 5, $\phi = 1$ e tamanho de amostra $n = 30$.

		$q = 2$					$q = 3$				
		15%	10%	5%	1%	0,5%	15%	10%	5%	1%	0,5%
$p = 2$	\mathcal{W}	16,9	11,6	6,0	1,2	0,67	17,9	12,5	6,3	1,6	0,76
	\mathcal{W}^*	15,8	10,6	5,2	1,0	0,57	15,8	10,7	5,2	1,1	0,55
	$\overline{\mathcal{W}}$	16,4	11,1	5,6	1,1	0,13	16,4	10,9	5,6	1,2	0,66
	$\overline{\mathcal{W}}^*$	15,9	10,6	5,3	1,0	0,54	16,0	10,7	5,3	1,2	0,61
		$q = 4$					$q = 5$				
$p = 2$	\mathcal{W}	19,2	13,6	7,6	2,0	1,09	20,2	14,5	7,9	2,1	1,22
	\mathcal{W}^*	16,5	11,4	5,7	1,4	0,64	17,4	11,9	6,2	1,5	0,73
	$\overline{\mathcal{W}}$	17,6	12,2	6,5	1,6	0,88	16,5	11,1	5,6	1,3	0,67
	$\overline{\mathcal{W}}^*$	16,7	11,4	5,8	1,4	0,75	17,5	11,9	6,2	1,5	0,84
		$q = 2$					$q = 3$				
$p = 3$	\mathcal{W}	17,4	11,7	6,3	1,4	0,81	19,4	13,8	7,5	1,7	1,10
	\mathcal{W}^*	15,2	10,2	5,2	1,1	0,56	15,8	10,8	5,4	1,2	0,67
	$\overline{\mathcal{W}}$	16,4	11,0	5,8	1,3	0,66	17,0	11,8	6,0	1,4	0,80
	$\overline{\mathcal{W}}^*$	15,3	10,2	5,3	1,1	0,56	15,8	11,0	5,5	1,2	0,72
		$q = 4$					$q = 5$				
$p = 3$	\mathcal{W}	20,8	15,0	8,4	2,1	1,19	20,3	14,7	7,9	2,1	1,14
	\mathcal{W}^*	16,8	11,7	6,0	1,3	0,74	16,8	11,5	5,9	1,2	0,72
	$\overline{\mathcal{W}}$	17,5	12,3	6,6	1,4	0,88	17,4	12,1	6,4	1,5	0,77
	$\overline{\mathcal{W}}^*$	16,9	11,7	6,1	1,3	0,79	16,9	11,7	6,2	1,4	0,74
		$q = 2$					$q = 3$				
$p = 4$	\mathcal{W}	18,1	12,6	6,5	1,5	0,81	20,8	14,8	8,1	2,2	1,24
	\mathcal{W}^*	14,1	9,5	4,6	0,9	0,45	15,7	10,5	5,4	1,2	0,48
	$\overline{\mathcal{W}}$	16,6	11,3	5,6	1,3	0,63	17,7	12,0	6,3	1,7	0,72
	$\overline{\mathcal{W}}^*$	14,3	9,4	4,7	0,9	0,50	15,8	10,5	5,3	1,3	0,52
		$q = 4$					$q = 5$				
$p = 4$	\mathcal{W}	21,5	15,7	8,6	2,4	1,29	-	-	-	-	-
	\mathcal{W}^*	16,3	10,8	5,7	1,2	0,63	-	-	-	-	-
	$\overline{\mathcal{W}}$	18,4	12,8	6,9	1,7	0,89	-	-	-	-	-
	$\overline{\mathcal{W}}^*$	16,2	10,8	5,5	1,2	0,60	-	-	-	-	-

milhanças perfiladas ajustadas. Isto porque as curvas correspondentes estão mais perto da linha horizontal em zero.

Observamos destes resultados de simulação que, como esperado, quanto maior o número de parâmetros a estimar pior é o desempenho dos testes, sendo necessária a correção delas. Quando comparamos as estatística de teste \mathcal{W} e $\overline{\mathcal{W}}$ percebemos a importância da utilização de funções de verossimilhanças ajustadas. Por exemplo, quando $n = 20$, $p = 2$ e $q = 3$, o desempenho do teste $\overline{\mathcal{W}}^*$, para níveis nominais de 15% e 10%, é ligeiramente melhor do que do teste \mathcal{W}^* . Nos outros níveis nominais, inverte-se este comportamento. Esta alternância de desempenho se mantém, praticamente, em todas as outras situações consideradas.

Algumas situações não foram consideradas no trabalho de simulação por considerarmos que o número de parâmetros do modelo proposto é inconsistente com o tamanho de amostra. Por exemplo, na Tabela 3.3 a situação $n = 10$ e $p = 4$ não foi considerada e na Tabela 3.4, a situação $n = 20$ e $p = 5$ também não foi considerada, ambas situações no modelos de regressão normal. No modelo de regressão gama, tanto com tamanho de amostra $n = 25$ quanto com tamanho de amostra $n = 30$, não consideramos nas simulações as situações em que $p = 4$ e $q = 5$, por considerarmos modelos com excesso de parâmetros relativamente ao tamanho de amostra.

Concluimos destes resultados que ao utilizarmos o teste da razão de verossimilhanças perfiladas ajustadas obteremos taxas de rejeição sob H_0 mais próximas do nível nominal do que se usarmos o teste da razão de verossimilhanças usual, este sendo marcadamente anti-conservador, já que em todas as situações consideradas as taxas de rejeição excedem consideravelmente os níveis nominais correspondentes. Concluimos também que a correção de Bartlett para o teste usual da razão de verossimilhanças apresenta comportamento similar à correção correspondente do teste da razão verossimilhanças perfiladas ajustadas.

Capítulo 4

Considerações finais e conclusões

Inferências em modelos paramétricos na presença de parâmetros de perturbação envolvem, em geral, o uso de verossimilhanças perfiladas. Estas dependem apenas dos parâmetros de interesse, já que são obtidas das correspondentes verossimilhanças originais com os parâmetros de perturbação substituídos por suas estimativas de máxima verossimilhança para cada valor fixado dos parâmetros de interesse. Embora as verossimilhanças perfiladas gozem de algumas propriedades semelhantes às válidas para as verossimilhanças genuínas, elas podem apresentar, por exemplo, vícios na função escore e na informação, estes sendo de ordem $O(1)$.

Ajustes para verossimilhanças perfiladas têm sido propostos na literatura. Um deles, muito atrativo por ser aplicável em contextos bem gerais, é o de Stern (1997). Nesta tese, estudamos em detalhes o ajuste de Stern e mostramos que seus resultados contêm erros. Um dos objetivos da tese foi, então, desenvolver novamente a proposta de Stern, obtendo assim o ajuste correto que faz com que a verossimilhança perfilada ajustada tenha vício da função escore e da informação de ordem $O(n^{-1})$. Através de alguns exemplos, temos indicação de que nossos resultados estão corretos. Esse objetivo foi alcançado no Capítulo 2.

É bem conhecido o fato de que, em problemas regulares, a estatística da razão de verossimilhanças tem, sob a hipótese nula, distribuição assintótica qui-quadrado com um número de graus de liberdade adequado e com um erro de ordem $O(n^{-1})$. O mesmo ocorre para a estatística da razão de verossimilhanças baseada em verossimilhanças perfiladas ajustadas.

Ainda no Capítulo 2, obtivemos uma correção de Bartlett que, aplicada nesta última estatística, faz com que o erro da aproximação torne-se de ordem $O(n^{-2})$.

A aplicação dos resultados teóricos obtidos em modelos de regressão não é trivial. Foi objetivo do Capítulo 3 aplicá-los aos modelos lineares generalizados e apresentá-los de forma tal que as fórmulas possam ser implementadas em qualquer software que tenha facilidades para executar operações com matrizes e vetores.

Também no Capítulo 3, apresentamos estudos de simulação para avaliar o desempenho dos testes da razão de verossimilhanças usual e o baseado em verossimilhanças perfiladas ajustadas e suas respectivas versões corrigidas via correção de Bartlett. Concluímos que o teste da razão de verossimilhanças usual é muito anticonservativo rejeitando a hipótese nula com uma frequência bem maior do que esperado com base no nível nominal do teste. O ajuste de Stern, na sua forma correta obtida nesta tese, torna o teste mais confiável no sentido de que, embora ainda seja anticonservativo, tem tamanho mais próximo do nível nominal escolhido. A aplicação do ajuste de Stern deixa ainda lugar para mais correção. Esta é obtida através do uso da correção de Bartlett aplicada ao teste ajustado desenvolvido nesta tese. Notamos que a correção de Bartlett aplicada à estatística da razão de verossimilhanças usual ou à estatística ajustada leva a testes com desempenhos semelhantes.

Nesta tese o foco esteve, primordialmente, em testes de hipóteses. Em pesquisas futuras, o ajuste para a verossimilhança perfilada de Stern deve ser estudado com mais detalhes no que diz respeito à estimação de parâmetros. Além disso, é conveniente compará-lo com outras propostas de ajuste apresentadas na literatura.

É de relevância também estudar ajustes para verossimilhanças perfiladas em problemas em que o número de parâmetros de perturbação cresce com o tamanho da amostra (Neyman & Scott, 1948) e em outros modelos de regressão como, por exemplo, os modelos de dispersão (Jørgensen, 1987, 1997) e os modelos mistos (McCulloch & Searle, 2001).

Apêndice A

Algumas relações entre cumulantes

A.1 Forma alternativa da soma $\sigma_{ab}\lambda^{br}$

Sabemos que,

$$\lambda = \begin{pmatrix} (\lambda_{ab}) & (\lambda_{aj}) \\ (\lambda_{ib}) & (\lambda_{ij}) \end{pmatrix}$$

e

$$\lambda^{-1} = \begin{pmatrix} (\lambda^{ab}) & (\lambda^{aj}) \\ (\lambda^{ib}) & (\lambda^{ij}) \end{pmatrix},$$

segundo a partição induzida pelos vetores de interesse e perturbação. Desta forma, podemos escrever

$$\lambda^{-1}\lambda = \begin{pmatrix} (\lambda^{ab})(\lambda_{ab}) + (\lambda^{aj})(\lambda_{ib}) & (\lambda^{ab})(\lambda_{aj}) + (\lambda^{aj})(\lambda_{ij}) \\ (\lambda^{ib})(\lambda_{ab}) + (\lambda^{ij})(\lambda_{aj}) & (\lambda^{ib})(\lambda_{aj}) + (\lambda^{ij})(\lambda_{ij}) \end{pmatrix}.$$

Se denotamos por I a matriz identidade, então, da equação

$$I = \lambda^{-1}\lambda$$

obtemos

$$(\lambda^{ab})(\lambda_{aj}) = -(\lambda^{aj})(\lambda_{ij}), \tag{A.1}$$

dentre outras equações.

Observemos que

$$\begin{aligned} (\sigma_{ab})(\lambda^{br}) &= \lambda^{ab^{-1}} \begin{pmatrix} \lambda^{ab} & \lambda^{aj} \end{pmatrix} \\ &= \begin{pmatrix} I & \lambda^{ab^{-1}} \lambda^{aj} \end{pmatrix}. \end{aligned}$$

Por outro lado, multiplicando por $\lambda^{ab^{-1}}$ a esquerda em (A.1) temos

$$\lambda_{aj} = -\lambda^{ab^{-1}} \lambda^{aj} \lambda_{ij};$$

portanto, multiplicando a equação anterior por λ_{ij}^{-1} a direita vemos que

$$\lambda_{aj} \lambda_{ij}^{-1} = -\lambda^{ab^{-1}} \lambda^{aj}.$$

Assim conseguimos provar que

$$(\sigma_{ab})(\lambda^{br}) = \begin{pmatrix} I & -\lambda^{aj} \lambda^{ij^{-1}} \end{pmatrix}$$

e isto nos permite escrever, por exemplo, que

$$\sigma_{ab} \lambda^{br} l_r = l_a - \nu^{st} \lambda^{as} l_t.$$

A.2 Matriz de cumulantes ν

Nesta seção provaremos que

$$\nu^{st} = \lambda^{rs} - \tau^{rs}$$

definido em (1.5) assume a forma

$$\nu = \begin{pmatrix} 0 & 0 \\ 0 & \lambda_{ij}^{-1} \end{pmatrix}.$$

Segundo foi definido em (1.4),

$$\tau^{rs} = \lambda^{ra} \lambda^{sb} \sigma_{ab},$$

que, matricialmente, pode ser escrito como

$$\begin{aligned} \tau &= \begin{pmatrix} \lambda^{ab} \lambda^{ab^{-1}} \lambda^{ab} & \lambda^{ab} \lambda^{ab^{-1}} \lambda^{aj} \\ \lambda^{ib} \lambda^{ab^{-1}} \lambda^{ab} & \lambda^{ib} \lambda^{ab^{-1}} \lambda^{aj} \end{pmatrix} \\ &= \begin{pmatrix} \lambda^{ab} & \lambda^{aj} \\ \lambda^{ib} & \lambda^{ib} \lambda^{ab^{-1}} \lambda^{aj} \end{pmatrix}. \end{aligned}$$

Então,

$$\begin{aligned}
\nu &= \lambda^{-1} - \tau \\
&= \begin{pmatrix} \lambda^{ab} & \lambda^{aj} \\ \lambda^{ib} & \lambda^{ij} \end{pmatrix} - \begin{pmatrix} \lambda^{ab} & \lambda^{aj} \\ \lambda^{ib} & \lambda^{ij} - \lambda^{ib}\lambda^{ab^{-1}}\lambda^{aj} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & \lambda^{ij} - \lambda^{ib}\lambda^{ab^{-1}}\lambda^{aj} \end{pmatrix}.
\end{aligned}$$

Provemos agora que

$$\lambda_{ij}^{-1} = \lambda^{ij} - \lambda^{ib}\lambda^{ab^{-1}}\lambda^{aj}.$$

Utilizando a expressão para a inversa de uma matriz particionada (Rao, 1973, pg.33), obtemos

$$(\lambda^{-1})^{-1} = \begin{pmatrix} \lambda^{ab^{-1}} + FE^{-1}F & -FE^{-1} \\ -E^{-1}F^T & E^{-1} \end{pmatrix},$$

onde

$$E = \lambda^{ij} - \lambda^{ib}\lambda^{ab^{-1}}\lambda^{aj}$$

e

$$F = \lambda^{ab^{-1}}\lambda^{aj}.$$

Desta forma, vemos que

$$\lambda_{ij} = E^{-1}$$

ou $\lambda_{ij}^{-1} = E$ e, portanto, $\lambda_{ij}^{-1} = \lambda^{ij} - \lambda^{ib}\lambda^{ab^{-1}}\lambda^{aj}$, como se queria demonstrar.

A.3 Obtenção de A_r^c

A função A^c é definida no Teorema 2.1.1 como sendo

$$A^c = \lambda^{cr} \nu^{st} \left(\lambda_{rs/t} - \frac{1}{2} \lambda_{rst} \right).$$

Posteriormente, na Seção 2.1.2, durante a demonstração do Teorema 2.1.2 é necessário encontrar A_r^c , ou seja, a derivada em relação ao r -ésimo componente do vetor ϑ . Estudaremos

aqui as derivadas de matrizes e depois aplicaremos os resultados para encontrar a expressão de A_r^c .

Seja $V(\vartheta)$ uma matriz não-singular de ordem $p + q \times p + q$, de elementos $v_{i,j}$, $i, j = 1, \dots, p + q$. Sabemos que $V^{-1}V = I$ e, derivando V^{-1} com relação a ϑ , obtemos

$$\frac{\partial V^{-1}}{\partial \vartheta} V + V^{-1} \frac{\partial V}{\partial \vartheta} = 0.$$

Então,

$$\frac{\partial V^{-1}}{\partial \vartheta} = -V^{-1} \frac{\partial V}{\partial \vartheta} V^{-1},$$

onde $\partial V / \partial \vartheta = (\partial v_{i,j} / \partial \vartheta)$. Desta forma, podemos escrever

$$\frac{\partial}{\partial \vartheta^u} \lambda^{cr} = \lambda^{cs} \lambda_{st/r} \lambda^{tu}$$

e

$$\frac{\partial}{\partial \vartheta^w} \nu^{st} = \nu^{vt} \lambda_{ts/r} \nu^{sw}.$$

Utilizando estas expressões das derivadas de cumulantes, obtemos que

$$A_r^c = - \left(\lambda^{cs} \lambda_{ts/r} \lambda^{tu} \nu^{vw} + \lambda^{cu} \nu^{tv} \lambda_{ts/r} \nu^{sw} \right) \left(\lambda_{uv/w} - \frac{1}{2} \lambda_{uvw} \right) + \lambda^{cu} \nu^{uw} \left(\lambda_{uv/wr} - \frac{1}{2} \lambda_{uvw/r} \right).$$

Apêndice B

Cumulantes de ordem superior nos modelos lineares generalizados

De grande importância é estabelecer uma notação para os cumulantes do logaritmo da função de verossimilhança nos modelos lineares generalizados (Seção 1.1.2) e é esse o objetivo deste Apêndice. Primeiramente, obteremos derivadas de ordem superior do logaritmo da função de verossimilhança dos modelos lineares generalizados.

A notação que utilizaremos é similar àquela apresentada em diferentes trabalhos como Cordeiro (1983), Cordeiro & McCullagh (1991), Cordeiro *et al.* (1993), Cribari-Neto & Ferrari (1995) e Botter & Cordeiro (1997), dentre outros.

O logaritmo da função de verossimilhança dos modelos lineares generalizados, segundo discutido na Seção 1.3, assume a forma

$$L(\beta, \phi) = \phi \sum_{l=1}^n \{y_l \theta^l - b(\theta^l) + c(y_l)\} + \sum_{l=1}^n d_1(y_l) + n d_2(\phi).$$

A partir desta função, por simples diferenciação, chega-se a que

$$\frac{\partial L(\beta, \phi)}{\partial \beta^r} = \phi \sum_{l=1}^n \{y_l - \mu^l\} \frac{1}{V_l} \frac{d\mu^l}{d\eta^l} x_{lr},$$
$$\frac{\partial^2 L(\beta, \phi)}{\partial \beta^r \partial \beta^s} = \phi \sum_{l=1}^n \left\{ \{y_l - \mu^l\} \frac{1}{V_l} \frac{d^2 \mu^l}{d\eta^{l2}} - \{y_l - \mu^l\} \frac{1}{V_l^2} \frac{dV_l}{d\mu^l} \left(\frac{d\mu^l}{d\eta^l} \right)^2 - \frac{1}{V_l} \left(\frac{d\mu^l}{d\eta^l} \right)^2 \right\} x_{lr} x_{ls}$$

e que

$$\begin{aligned} \frac{\partial^3 L(\beta, \phi)}{\partial \beta^r \partial \beta^s \partial \beta^t} = & \phi \sum_{l=1}^n \left\{ \{y_l - \mu^l\} \frac{2}{V_l^3} \left(\frac{dV_l}{d\mu^l} \right)^2 \left(\frac{d\mu^l}{d\eta^l} \right)^3 - \{y_l - \mu^l\} \frac{1}{V_l^2} \frac{d^2 V_l}{d\mu^l{}^2} \left(\frac{d\mu^l}{d\eta^l} \right)^3 \right. \\ & - \frac{3}{V_l} \frac{d\mu^l}{d\eta^l} \frac{d^2 \mu^l}{d\eta^l{}^2} + \{y_l - \mu^l\} \frac{1}{V_l} \frac{d^3 \mu^l}{d\eta^l{}^3} - \{y_l - \mu^l\} \frac{3}{V_l^2} \frac{dV_l}{d\mu^l} \frac{d\mu^l}{d\eta^l} \frac{d^2 \mu^l}{d\eta^l{}^2} \\ & \left. + \frac{2}{V_l^2} \frac{dV_l}{d\mu^l} \left(\frac{d\mu^l}{d\eta^l} \right)^3 \right\} x_{lr} x_{ls} x_{lt}. \end{aligned}$$

Tomando esperanças nas duas últimas expressões encontram-se os cumulantes

$$\lambda_{rs} = -\phi \sum_{l=1}^n \frac{1}{V_l} \left(\frac{d\mu^l}{d\eta^l} \right)^2 x_{lr} x_{ls}$$

e

$$\lambda_{rst} = -\phi \sum_{l=1}^n \left\{ \frac{3}{V_l} \frac{d\mu^l}{d\eta^l} \frac{d^2 \mu^l}{d\eta^l{}^2} - \frac{2}{V_l^2} \frac{dV_l}{d\mu^l} \left(\frac{d\mu^l}{d\eta^l} \right)^3 \right\} x_{lr} x_{ls} x_{lt}.$$

Definindo as funções

$$f = \frac{d\mu}{d\eta} \frac{d^2 \mu}{d\eta^2} \frac{1}{V} \quad (\text{B.1})$$

e

$$g = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2 \mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3, \quad (\text{B.2})$$

podemos escrever os cumulantes como

$$\lambda_{rs} = -\phi \sum_{l=1}^n w_l x_{lr} x_{ls}, \quad (\text{B.3})$$

lembrando que $w = (d\mu/d\eta)^2/V$, e

$$\lambda_{rst} = -\phi \sum_{l=1}^n (f + 2g)_l x_{lr} x_{ls} x_{lt}. \quad (\text{B.4})$$

Derivando λ_{rs} em relação a β^t temos

$$\lambda_{rs/t} = \frac{\partial}{\partial \beta^t} \lambda_{rs},$$

que assume a forma

$$\lambda_{rs/t} = -\phi \sum_{l=1}^n \left\{ 2 \frac{d\mu^l}{d\eta^l} \frac{d^2\mu^l}{d\eta^{l^2}} \frac{d\eta^l}{d\beta^t} \frac{d\theta^l}{d\mu^l} + \left(\frac{d\mu^l}{d\eta^l} \right)^2 \frac{d^2\theta^l}{d\mu^{l^2}} \frac{d\mu^l}{d\eta^l} \frac{d\eta^l}{d\beta^t} \right\} x_{lr} x_{ls}.$$

Observando que

$$\frac{d\theta}{d\eta} = \frac{d\theta}{d\mu} \frac{d\mu}{d\eta} = \frac{1}{V} \frac{d\mu}{d\eta},$$

então

$$\frac{d^2\theta}{d\eta^2} = \frac{d^2\theta}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^2 + \frac{d\theta}{d\mu} \frac{d^2\mu}{d\eta^2},$$

logo

$$\frac{d^2\theta}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^2 = \frac{d^2\theta}{d\eta^2} - \frac{d\theta}{d\mu} \frac{d^2\mu}{d\eta^2}$$

e multiplicando ambos os lados desta última equação por $d\mu/d\eta$ temos

$$\frac{d^2\theta}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d\mu}{d\eta} = \frac{d^2\theta}{d\eta^2} \frac{d\mu}{d\eta} - \frac{d\theta}{d\mu} \frac{d^2\mu}{d\eta^2} \frac{d\mu}{d\eta}.$$

Vemos também que g dado em (B.2) pode ser escrito como

$$g = \frac{d^2\theta}{d\eta^2} \frac{d\mu}{d\eta}$$

e, portanto, temos

$$\frac{d^2\theta}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d\mu}{d\eta} = g - \frac{d^2\mu}{d\eta^2} \frac{d\theta}{d\eta}$$

ou ainda

$$\frac{d^2\theta}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d\mu}{d\eta} = g - f.$$

Podemos então escrever

$$\lambda_{rs/t} = -\phi \sum_{l=1}^n (f + g)_l x_{lr} x_{ls} x_{lt}, \quad (\text{B.5})$$

e utilizando o fato de que

$$\frac{d^2\theta}{d\mu^2} = -\frac{1}{V^2} \frac{dV}{d\mu},$$

o cumulante em (B.5) assume a forma

$$\lambda_{rs/t} = -\phi \sum_{l=1}^n \left\{ 2 \frac{d\mu^l}{d\eta^l} \frac{d^2\mu^l}{d\eta^{l^2}} \frac{1}{V_l} - \frac{1}{V_l^2} \frac{dV_l}{d\mu^l} \left(\frac{d\mu^l}{d\eta^l} \right)^3 \right\} x_{lr} x_{ls} x_{lt}.$$

Com o objetivo de encontrar a expressão do cumulante $\lambda_{rs/tu}$, derivamos $\lambda_{sr/t}$ com relação a β^u . Desta forma temos

$$\lambda_{rs/tu} = -\phi \sum_{l=1}^n \frac{\partial}{\partial \beta^u} (f+g)_l x_{lr} x_{ls} x_{lt},$$

na qual

$$\begin{aligned} \frac{\partial}{\partial \beta^u} (f+g)_l (rst)_l &= \frac{d}{d\eta} (f+g)_l x_{lr} x_{ls} x_{lt} \frac{\partial \eta}{\partial \beta^u} \\ &= \frac{d}{d\eta} (f+g)_l x_{lr} x_{ls} x_{lt} x_{lu}, \end{aligned}$$

obtendo-se

$$\lambda_{rs/tu} = -\phi \sum_{l=1}^n \frac{d}{d\eta} (f+g)_l x_{lr} x_{ls} x_{lt} x_{lu}.$$

Dado que

$$\frac{dg}{d\eta} = \frac{d^2\mu}{d\eta^2} \frac{d^2\theta}{d\eta^2} + \frac{d^3\theta}{d\eta^3} \frac{d\mu}{d\eta}$$

e

$$\frac{df}{d\eta} = \frac{d^3\mu}{d\eta^3} \frac{d\theta}{d\eta} + \frac{d^2\theta}{d\eta^2} \frac{d^2\mu}{d\eta^2},$$

temos

$$\lambda_{rs/tu} = -\phi \sum_{l=1}^n \left\{ \frac{d^3\theta}{d\eta^3} \frac{d\mu}{d\eta} + 2 \frac{d^2\mu}{d\eta^2} \frac{d^2\theta}{d\eta^2} + \frac{d^3\mu}{d\eta^3} \frac{d\theta}{d\eta} \right\}^l x_{lr} x_{ls} x_{lt} x_{lu}.$$

Observemos que

$$\frac{d^3\theta}{d\eta^3} = \frac{d^3\theta}{d\mu^3} \left(\frac{d\mu}{d\eta} \right)^2 + 2 \frac{d^2\theta}{d\mu^2} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \frac{d^2\mu}{d\eta^2} + \frac{1}{V} \frac{d^3\mu}{d\eta^3}$$

e

$$\frac{d^3\theta}{d\mu^3} = \frac{2}{V^3} \left(\frac{dV}{d\mu} \right)^2 - \frac{1}{V^2} \frac{d^2V}{d\mu^2}.$$

Substituindo a expressão anterior e aquela para $d^2\theta/d\mu^2$ obtemos

$$\begin{aligned} \lambda_{rs/tu} = & -\phi \sum_{l=1}^n \left\{ \frac{2}{V_l^3} \left(\frac{dV_l}{d\mu^l} \right)^2 \left(\frac{d\mu^l}{d\eta^l} \right)^3 - \frac{1}{V_l^2} \frac{d^2V_l}{d\mu^{l^2}} \left(\frac{d\mu^l}{d\eta^l} \right)^3 - \frac{1}{V_l^2} \frac{dV_l}{d\mu^l} \frac{d\mu^l}{d\eta^l} \frac{d^2\mu^l}{d\eta^{l^2}} \right. \\ & \left. - \frac{4}{V_l^2} \frac{dV_l}{d\mu^l} \left(\frac{d\mu^l}{d\eta^l} \right)^2 \frac{d^2\mu^l}{d\eta^{l^2}} + \frac{2}{V_l} \frac{d\mu^l}{d\eta^l} \frac{d^3\mu^l}{d\eta^{l^3}} + \frac{2}{V_l} \left(\frac{d^2\mu^l}{d\eta^{l^2}} \right)^2 \right\} x_{lr} x_{ls} x_{lt} x_{lu}. \end{aligned}$$

Observemos que as funções $df/d\eta$ e $dg/d\eta$ podem ser escritas como

$$\frac{df}{d\eta} = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} + \frac{1}{V} \left\{ \left(\frac{d^2\mu}{d\eta^2} \right)^2 + \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3} \right\}$$

e

$$\begin{aligned} \frac{dg}{d\eta} = & -\frac{4}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} + \frac{1}{V} \left\{ \left(\frac{d^2\mu}{d\eta^2} \right)^2 + \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3} \right\} \\ & + \frac{1}{V^2} \left(\frac{d\mu}{d\eta} \right)^4 \left\{ \frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - \frac{d^2V}{d\mu^2} \right\}. \end{aligned}$$

B.1 Obtenção dos cumulantes $\lambda_{rst/u}$ e λ_{rstu}

Sabemos que

$$\lambda_{rst/u} = \frac{d}{d\eta} \lambda_{rst} \frac{\partial}{\partial \beta^u} \eta$$

e, portanto,

$$\lambda_{rst/u} = -\phi \sum_{l=1}^n \frac{d}{d\eta} (f + 2g)_l x_{lr} x_{ls} x_{lt} x_{lu}.$$

Utilizando as expressões das derivadas de f e g na seção anterior, obtemos

$$\lambda_{rst/u} = -\phi \sum_{l=1}^n \left\{ \frac{d^3\mu}{d\eta^3} \frac{d\theta}{d\eta} + 3 \frac{d^2\mu}{d\eta^2} \frac{d^2\theta}{d\eta^2} + 2 \frac{d\mu}{d\eta} \frac{d^3\theta}{d\eta^3} \right\}^l x_{lr} x_{ls} x_{lt} x_{lu}$$

e substituindo as expressões para $d\theta/d\eta$, $d^2\theta/d\eta^2$ e $d^3\theta/d\eta^3$ temos

$$\begin{aligned} \lambda_{rst/u} = & -\phi \sum_{l=1}^n \left\{ \frac{4}{V_l^3} \left(\frac{dV_l}{d\mu^l} \right)^2 \left(\frac{d\mu^l}{d\eta^l} \right)^3 - \frac{7}{V_l^2} \frac{dV_l}{d\mu^l} \frac{d^2\mu^l}{d\eta^{l2}} \left(\frac{d\mu^l}{d\eta^l} \right)^2 - \frac{2}{V_l^2} \frac{d^2V_l}{d\mu^{l2}} \left(\frac{d\mu^l}{d\eta^l} \right)^3 \right. \\ & \left. - \frac{2}{V_l^2} \frac{dV_l}{d\mu^l} \frac{d\mu^l}{d\eta^l} \frac{d^2\mu^l}{d\eta^{l2}} + \frac{3}{V_l} \frac{d\mu^l}{d\eta^l} \frac{d^3\mu^l}{d\eta^{l3}} + \frac{3}{V_l} \left(\frac{d^2\mu^l}{d\eta^{l2}} \right)^2 \right\} x_{lr} x_{ls} x_{lt} x_{lu}. \end{aligned}$$

Para encontrar a forma do cumulante λ_{rstu} primeiramente observamos que

$$\frac{\partial L(\beta, \phi)}{\partial \beta_r \partial \beta_s \partial \beta_t} = \phi \sum_{l=1}^n \left\{ \{y_l - \mu^l\} \frac{d^3\theta^l}{d\eta^{l3}} - 2g_l - f_l \right\} x_{lr} x_{ls} x_{lt},$$

e derivando em relação a β^u temos

$$\frac{\partial L(\beta, \phi)}{\partial \beta^r \partial \beta^s \partial \beta^t \partial \beta^u} = \phi \sum_{l=1}^n \left\{ \{y_l - \mu^l\} \frac{d^4\theta^l}{d\eta^{l4}} \frac{\partial \eta^l}{\partial \beta^u} - \frac{d\mu^l}{d\eta^l} \frac{d^3\theta^l}{d\eta^{l3}} \frac{\partial \eta^l}{\partial \beta^u} - 2 \frac{dg_l}{d\eta^l} \frac{\partial \eta^l}{\partial \beta^u} - \frac{df_l}{d\eta^l} \frac{\partial \eta^l}{\partial \beta^u} \right\} x_{lr} x_{ls} x_{lt}.$$

Tomando esperança e observando que $\partial\eta^l/\partial\beta^u = x_{lu}$ e $E\{y - \mu\} = 0$, chegamos a

$$E \left\{ \frac{\partial L(\beta, \phi)}{\partial\beta^r \partial\beta^s \partial\beta^t \partial\beta^u} \right\} = -\phi \sum_{l=1}^n \left\{ \frac{d\mu^l}{d\eta^l} \frac{d^3\theta^l}{d\eta^{l^3}} + 2 \frac{dg_l}{d\eta^l} - \frac{df_l}{d\eta^l} \right\} x_{lr} x_{ls} x_{lt} x_{lu}.$$

Substituindo as expressões de $dg/d\eta$ e $df/d\eta$ e nestas, as correspondentes expressões de $d\theta/d\eta$, $d^2\theta/d\eta^2$ e $d^3\theta/d\eta^3$ obtemos

$$\lambda_{rstu} = -\phi \sum_{l=1}^n e_l x_{lr} x_{ls} x_{lt} x_{lu},$$

onde

$$e = \frac{4}{V} \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3} + \frac{3}{V} \left(\frac{d^2\mu}{d\eta^2} \right)^2 + \frac{6}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4 - \frac{12}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} - \frac{3}{V^2} \frac{d^2V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4. \quad (\text{B.6})$$

Apêndice C

Propriedades do produto Hadamard e da função traço

Em algumas situações precisamos calcular

$$1^\top W(Z \otimes Z)W1, \quad (\text{C.1})$$

onde W é uma matriz diagonal de pesos, $Z = X(X^\top X)^{-1}X^\top$ definida na Seção 3.1.1 e \otimes o produto direto de matrizes ou produto de Hadamard.

Aqui estudaremos propriedades que relacionam o produto Hadamard com a função traço para escrever de maneira reduzida (C.1) nas expressões em que apareça. Detalhes sobre as propriedades do produto Hadamard ou produto direto e da função traço pode ser encontrado em Rao (1973).

Provemos primeiramente que

$$W(Z \otimes Z)W = (WZ) \otimes (ZW).$$

Observemos que

$$W(Z \otimes Z)W = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{pmatrix} \begin{pmatrix} z_{11}^2 & \cdots & z_{1n}^2 \\ \vdots & \ddots & \vdots \\ z_{n1}^2 & \cdots & z_{nn}^2 \end{pmatrix} \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{pmatrix},$$

que pode ser escrito como

$$W(Z \otimes Z)W = \begin{pmatrix} w_1 z_{11}^2 & \cdots & w_1 z_{1n}^2 \\ \vdots & \ddots & \vdots \\ w_n z_{n1}^2 & \cdots & w_n z_{nn}^2 \end{pmatrix} \begin{pmatrix} w_1 & & \\ & \cdots & \\ & & w_n \end{pmatrix} = \begin{pmatrix} w_1 z_{11}^2 w_1 & \cdots & w_1 z_{1n}^2 w_n \\ \vdots & \ddots & \vdots \\ w_n z_{n1}^2 w_1 & \cdots & w_n z_{nn}^2 w_n \end{pmatrix}.$$

De outra forma,

$$\begin{aligned} (WZ) \otimes (ZW) &= \begin{pmatrix} w_1 z_{11} & \cdots & w_1 z_{1n} \\ \vdots & \ddots & \vdots \\ w_n z_{n1} & \cdots & w_n z_{nn} \end{pmatrix} \otimes \begin{pmatrix} z_{11} w_1 & \cdots & z_{1n} w_n \\ \vdots & \ddots & \vdots \\ z_{n1} w_1 & \cdots & z_{nn} w_n \end{pmatrix} \\ &= \begin{pmatrix} w_1 z_{11}^2 w_1 & \cdots & w_1 z_{1n}^2 w_n \\ \vdots & \ddots & \vdots \\ w_n z_{n1}^2 w_1 & \cdots & w_n z_{nn}^2 w_n \end{pmatrix}. \end{aligned}$$

Portanto, provamos que

$$W(Z \otimes Z)W = (WZ) \otimes (ZW)$$

e, devido a $(ZW) = (ZW)^\top$, temos

$$W(Z \otimes Z)W = (WZ) \otimes (ZW)^\top.$$

Utilizando a seguinte propriedade do produto Hadamard

$$1^\top (A \otimes B^\top) 1 = \text{tr}(AB),$$

que pode ser encontrada em Rao (1973, pg. 30), obtemos

$$\begin{aligned} 1^\top W(Z \otimes Z)W 1 &= 1^\top (WZ) \otimes (ZW) 1 \\ &= \text{tr}(WZ) \text{tr}(ZW) \\ &= \text{tr}(WX(X^\top WX)^{-1} X^\top WX(X^\top WX)^{-1} X^\top) \\ &= \text{tr}(WX(X^\top WX)^{-1} X^\top) \\ &= \text{tr}(X^\top WX(X^\top WX)^{-1}). \end{aligned}$$

Desta forma,

$$1^\top W(Z \otimes Z)W1 = \dim(X^\top WX). \quad (\text{C.2})$$

Na situação em que a matriz de pesos W se reduz à matriz identidade I , temos

$$1^\top I(Z \otimes Z)I1 = \text{tr}(Z).$$

Por outro lado, a função traço satisfaz

$$\text{tr}(AB) = \text{tr}(BA),$$

e observando que

$$\text{tr}(Z) = \text{tr}(X(X^\top X)^{-1}X^\top),$$

obtemos que

$$\text{tr}(Z) = \text{tr}(X^\top X(X^\top X)^{-1}) = \dim(X^\top X).$$

Referências Bibliográficas

- Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters, based on the standardized log likelihood ratio. *Biometrika*, **73**(2), 307–322.
- Barndorff-Nielsen, O.E. (1992). Adjusted likelihood inference about interest parameters. *Theory of Probability and its Applications*, **38**(2), 179–193.
- Barndorff-Nielsen, O.E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society, B*, **56**(1), 125–140.
- Barndorff-Nielsen, O.E. & Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.
- Barndorff-Nielsen, O.E. & Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistics. *Biometrika*, **75**(2), 374–378.
- Bartlett, M.S. (1937). Properties of sufficiency and statistical test. *Proceeding of the Royal Society, A*, **160**, 268–282.
- Bartlett, M.S. (1953a). Approximate confidence intervals. *Biometrika*, **40**(1-2), 12–19.
- Bartlett, M.S. (1953b). Approximate confidence intervals II. More than one unknown parameter. *Biometrika*, **40**(3-4), 306–317.

- Bartlett, M.S. (1955). Approximate confidence intervals, III. A bias correction. *Biometrika*, **40**(1-2), 306–317.
- Bishop, Y., Fienberg, S. & Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.
- Botter, D.A. & Cordeiro, G.M. (1997). Bartlett corrections for generalized linear models with dispersion covariates. *Communications in Statistics, Theory and Methods*, **26**(2), 279–307.
- Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society, B*, **45**(3), 404–413.
- Cordeiro, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**(2), 265–274.
- Cordeiro, G.M. & Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika*, **78**(3), 573–582.
- Cordeiro, G.M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, B*, **55**(3), 629–643.
- Cordeiro, G.M. & Paula, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, **76**(1), 93–100.
- Cordeiro, G.M., Ferrari, S.L.P. & Paula, G.A. (1993). Improved score test for generalized linear models. *Journal of the Royal Statistical Society, B*, **55**(3), 661–674.
- Cordeiro, G.M., Paula, G.A. & Botter, D.A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review*, **62**(2), 257–274.
- Cox, D.R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, B*, **49**(1), 1–39.

- Cox, D., Hinkley, D., Reid, N. & Snell, E. (1991). *Statistical Theory and Modelling*. London: Chapman and Hall.
- Cribari-Neto, F. & Ferrari, S.L.P. (1995). Second order asymptotics for score test in generalized linear models. *Biometrika*, **82**(2), 426–432.
- Davison, A.C. (1988). Approximate conditional inference in generalized linear models. *Journal of the Royal Statistical Society, B*, **50**(3), 445–461.
- Davison, A.C. & Stafford, J.E. (1998). The score function and a comparison of various adjustments of the profile likelihood. *The Canadian Journal of Statistics*, **26**(1), 139–148.
- DiCiccio, T.J. & Stern, E.E. (1994). Frequentist and bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *Journal of the Royal Statistical Society, B*, **56**(2), 397–408.
- DiCiccio, T.J., Martin, M.A., Stern, E.E. & Young, G.A. (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, B*, **58**(1), 189–203.
- Doornik, J.A. (2001). *Ox: an Object-oriented Matrix Programming Language*. London: Timberlake Consultants and Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik/>, fourth edition.
- Fahrmeir, L. & Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*, **13**(1), 342–368.
- Ferrari, S.L.P. (1991). *Aperfeiçoamento do teste escore, aplicações e extensões*. Ph.D. thesis, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Fletcher, R. (1987). *Practical Methods of Optimization*. New York: John Wiley & Sons.
- Fraser, D. (1989). Adjustments to profile likelihood. *Biometrika*, **76**(3), 477–488.
- Gnedenko, B.V. & Kolmogorov, A.N. (1962). *Limits Distributions for Sums of Independent Random Variables*. Reading: Addison-Wesley, second edition.

- Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**(4), 1208–1212.
- Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society*, B, **49**(2), 127–162.
- Jørgensen, B. (1997). *Theory of Dispersion Models*. London: Chapman & Hall.
- Knudsen, S.J. (1998). *Estimating Functions and Separate Inference*. Ph.D. thesis, Institut for Matematiske Fag, Århus Universitet, Århus, Danmark.
- Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**(3-4), 295–303.
- Luenberger, D.G. (1973). *Introducton to Linear and Nonlinear Programming*. New York: Addison-Wesley Publishing Company, Inc.
- McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. Oxford: Chapman & Hall.
- McCullagh, P. & Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society*, B, **52**(2), 325–344.
- McCulloch, C.E. & Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.
- Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, A, **14**(135), 370–384.
- Neyman, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1–32.
- Peters, W. (1987). *Counting for Something*. New York: Springer-Verlag.

- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. New York: John Wiley and Sons, second edition.
- Sartori, N., Bellio, R. & Salvan, A. (1999). The directed modified profile likelihood in models with many nuisance parameters. *Biometrika*, **86**(3), 735–742.
- Sen, P.K. & Singer, J. (1993). *Large Sample Methods in Statistics: an Introduction with Applications*. New York: Chapman and Hall.
- Severini, T.A. (1998a). An approximation to the modified profile likelihood function. *Biometrika*, **85**(2), 403–411.
- Severini, T.A. (1998b). Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, **85**(3), 507–522.
- Spivak, M. (1970). *Calculus*. Barcelona: Editorial Reverté.
- Stern, S.E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society, B*, **59**(3), 653–665.
- Stuart, A. & Ord, J. (1987). *Kendall's Advanced Theory of Statistics*, volume 1. London: Charles Griffin, fifth edition.
- Sweeting, T.J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Annals of Statistics*, **8**(6), 1375–1381.
- Wedderburn, R.W.M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, **63**(1), 27–32.
- Whittaker, E. & Watson, G. (1952). *A Course of Modern Analysis*. New York: Cambridge University Press.
- Wilks, S.S. (1962). *Mathematical Statistics*. New York: John Wiley & Sons.