APERFEIÇOAMENTO
DE TESTES ESCORE,
APLICAÇÕES E EXTENSÕES

Silvia Lopes de Paula Ferrari

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

Orientador: **Prof. Dr. Gauss Moutinho Cordeiro**

– São Paulo, 1991 –

ERRATA

Página	Linha	Onde se lê	Leia-se
Resumo ⁽¹⁾	8	são verificadas	é verificada
Abstract ⁽²⁾	7	are	is
4	27	Mantel-Haenzel	Mantel-Haenszel
5	10	$n^{3/2}$	$n^{-3/2}$
6	25	normalizadas	normalizadoras
20	5	$B_3 - 2t$	$B_3 - at$
26	6	matriz	matriz identidade
31	3	Mantel-Haenzel	Mantel-Haenszel
45	19	n_i	n_i^{-1}
70	14	$\kappa_{\phi,\phi}^{-1}$	$K_{\phi,\phi}^{-1}$
70	15	$\kappa_{\beta,\beta}^{-1}$	$K_{\beta,\beta}^{-1}$
75	19	$(y_i - \mu^{(0)})$	$(y_i - \mu^{(0)})^2$
78	23	que S_R tem	que $n^{-1}S_R$ tem
78	24	$(n - 1)/1$	$(n - 1)/2$
83	20	indenpendentes	independentes
90	17, 19	$K = nI$	$K = nI_p$
94	12, 13	re-spectivamente	res-pectivamente
95	2, 3	ortogonal-idade	ortogona-lidade
100	23	$b(x_\alpha n)$	$b(x_\alpha, n)$
102	17	indice	índice
109	22	$\phi(\log \phi - \psi(\phi))$. Aqui	$\phi \log \phi - \log \Gamma(\phi)$. Aqui
116	5	modelos	modelo
121	18	$1^T W \{Q_2$	$1^T F \{Q_2$
121	18	$[(Z^* - Z_2^* \otimes C_2)]F1$	$[(Z^* - Z_2^*) \otimes C_2]^T \} W1$
123	20	$1^T W [Q_2$	$1^T F [Q_2$
123	20	$\{(Z^* - Z_2^*) \otimes C_2\} F1$	$\{(Z^* - Z_2^*) \otimes C_2\}^T \} W1$
124	16	$Z_2^* = Z_2$	$Z^* - Z_2^* = Z - Z_2$
126	20	$\chi^i \chi^2$	χ^2

(1) Segunda página do RESUMO.

(2) Segunda página do ABSTRACT.

*Ao meu marido
Alessandro*

*Aos meus pais
Lucy e Edinésio*

*À minha avó
Hilda*

AGRADECIMENTOS

Esta tese é resultado de um encontro. Um encontro entre meu lado profissional, exercido com vontade e dedicação, e meu lado pessoal, pelo gratificante contato com as pessoas com as quais convivi na sua elaboração. Estas pessoas estão aqui, presentes neste trabalho, fornecendo-me inspiração, esclarecendo dúvidas, transmitindo seu conhecimento e incentivo. Assim,

ao Professor **Gauss M. Cordeiro**, com sua competente e dinâmica orientação, seu profundo conhecimento e sua entusiasta crença em minha capacidade,

ao Professor **Gilberto A. Paula**, responsável pelo meu contato inicial com o Professor Gauss, e pelo trabalho em conjunto na obtenção dos resultados do Capítulo 3,

à colega **Denise A. Botter**, pelo seu companheirismo e pela contribuição em alguns exemplos do Capítulo 5 e

à colega **Mônica Inés Casajús**, responsável pela digitação,

quero expressar forte agradecimento pela dedicação e contribuição profissional e profunda emoção pelo sentimento sempre presente de amizade, o que confere a esta tese uma dimensão de caráter pessoal que muito me enriquece.

De meu orientador Professor Gauss M. Cordeiro, permanece a sensação de ter reduzido a poucos metros a grande distância que nos separa, com suas cartas, mensagens e telefonemas, quase diários nas ocasiões mais necessárias, sempre respondendo prontamente às minhas colocações ou fornecendo seus comentários e sugestões. Aprendi, também com meu orientador, um sentido mais profundo de amor e dedicação à pesquisa que com certeza irá influenciar todo meu futuro profissional.

Manifesto, também, meu reconhecimento ao **CNPq**, à **FAPESP** e à direção do **IMPA** que prestaram apoio financeiro.

RESUMO

O principal objetivo desta tese é obter um fator de correção para a estatística escore de forma similar ao enfoque dado à da razão de verossimilhança. Sabe-se que, para melhorar a aproximação por qui-quadrado para a distribuição da estatística da razão de verossimilhança, pode-se multiplicar a estatística por um fator de correção, conhecido como correção de Bartlett. Desta maneira, é obtida uma estatística ajustada tendo distribuição qui-quadrado até ordem n^{-1} sob a hipótese nula, onde n é o tamanho da amostra.

A impossibilidade de se obter um fator de correção para a estatística escore tem sido citada frequentemente na literatura. Entretanto, mostramos que existe uma correção do tipo Bartlett para a estatística escore. Assim, obtemos uma estatística escore modificada tendo distribuição qui-quadrado até ordem n^{-1} sob a hipótese nula cuja forma é, de certa maneira, semelhante à da estatística da razão de verossimilhança ajustada, porque é dada por uma multiplicação da estatística não modificada por um fator de correção. Enquanto a correção de Bartlett não depende do valor da estatística da razão de verossimilhança, o fator da correção para a estatística escore é dado por um polinômio na própria estatística. A distribuição da estatística escore modificada sob a hipótese alternativa também é considerada neste trabalho.

A obtenção do fator de correção para a estatística escore depende do modelo considerado para os dados. É bastante difícil obter a correção em alguns problemas já que esta requer o cálculo de cumulantes conjuntos de derivadas da log-verossimilhança total.

Derivamos fórmulas especiais para os fatores de correção em alguns testes em modelos lineares generalizados, modelos próprios de dispersão e modelos não-lineares da família exponencial. As fórmulas são dadas em notação matricial e têm vantagens do ponto de vista numérico. Também são suficientemente simples para serem usadas analiticamente para se obter expressões em forma fechada para os fatores de correção, análogos às correções de Bartlett, em diversos testes escore importantes. Testes escore aperfeiçoados são discutidos e várias aplicações a testes especiais são dadas ao longo deste trabalho. A exatidão das fórmulas obtidas são verificadas em alguns casos particulares. Ilustramos o uso das fórmulas através de exemplos numéricos. Algumas simulações são apresentadas para mostrar que os fatores de correção para estatísticas escore têm importância do ponto de vista prático.

Fórmulas especiais para obtenção de fatores de correção para problemas envolvendo parâmetros ortogonais são derivadas. Além disto, damos uma fórmula geral em notação matricial para o fator de correção para ser usada em um sistema computacional algébrico como o REDUCE ou uma linguagem numérica que permita álgebra matricial como, por exemplo, GAUSS ou APL.

Estendemos nossos resultados obtendo fatores de correção para uma classe mais ampla de estatísticas que têm distribuição qui-quadrado assintoticamente. A famosa correção de Bartlett pode ser obtida como um caso especial destes fatores de correção. Mostramos que tais fatores podem ser facilmente obtidos de expansões até ordem n^{-1} de um certo número de momentos destas estatísticas.

Finalmente, apresentamos algumas conclusões e sugerimos alguns tópicos para pesquisa futura.

ABSTRACT

The main purpose of this thesis is to obtain a correction factor for the score statistic, in similarity to the classical approach given for the likelihood ratio statistic. In order to improve the chi-squared approximation for the distribution of the likelihood ratio statistic, it is widely known that one may multiply the statistic by a correction factor, known as Bartlett correction. This way one gets an adjusted statistic having chi-squared distribution to order n^{-1} under the null hypothesis, where n is the sample size.

The impossibility in developing a correction factor for the score statistic has been claimed frequently in the literature. However, we show that a Bartlett-type correction is available for the score statistic. Thus, we obtain a modified score statistic having chi-squared distribution to order n^{-1} under the null hypothesis whose form is, to some extent, very close to the adjusted likelihood ratio statistic, because it is given by a multiplication of the unmodified statistic by a correction factor. While the Bartlett correction does not depend on the value of the likelihood ratio statistic, the correction factor for the score statistic is given by a polynomial of the statistic itself. The distribution of the modified score statistic under the null hypothesis is also considered in this work.

The computation of the correction factor for the score statistic depends on the model considered for the data. It is very difficult to obtain the correction in some testing problems since it requires the computation of joint cumulants of log-likelihood derivatives for the full data. We derive special formulae for the correction factors in some score

testing problems in generalized linear models, proper dispersion models and non-linear exponential family models. The formulae are given in matrix notation and have advantages for numerical purposes. They are also simple enough to be used analytically to obtain several closed-form expressions for the correction factors, analogous to Bartlett corrections, in a variety of important score tests. Improved score tests for these models are discussed and several applications for special tests are given throughout this work. The accuracy of the obtained formulae are checked in some particular cases. We illustrate the formulae through numerical examples. Some simulations are performed to show that the correction factors for score statistics are of practical importance.

Special formulae for computing the correction factors for score testing problems involving orthogonal parameters are derived. Also, we give a general matrix formula for the correction factor to be used in a computer algebra system such as REDUCE or a language supporting numerical matrix algebra as, for example, GAUSS or APL.

We extend our results by obtaining correction factors for a wider class of statistics which have asymptotic chi-squared distribution. The famous Bartlett correction can be obtained as a special case of these correction factors. We show that such factors can be easily obtained from expansions to order n^{-1} of a certain number of moments of these statistics.

Finally, we give some conclusions and suggest some subjects for future research.

ÍNDICE

1.	INTRODUÇÃO	1
2.	UMA ESTATÍSTICA ESCORE APERFEIÇOADA TENDO DISTRIBUIÇÃO χ^2 ATÉ ORDEM n^{-1}	11
2.1.	Introdução	11
2.2.	Expansão assintótica para a função distribuição da estatística da razão de verossimilhança	13
2.3.	Expansões assintóticas de Harris	15
2.4.	Uma estatística escore aperfeiçoada	18
2.5.	O caso uniparamétrico	24
2.6.	Um comentário sobre o poder dos testes escore aperfeiçoados	25
3.	APERFEIÇOAMENTO DE TESTES ESCORE EM MODELOS LINEARES GENERALIZADOS	29
3.1.	Introdução	29
3.2.	Testes escore em modelos lineares generalizados	32
3.3.	Testes escore aperfeiçoados	35
3.4.	Alguns comentários sobre os coeficientes	38
3.5.	Cálculo dos coeficientes no sistema GLIM	39

Índice

3.6. Alguns casos especiais	41
3.7. Modelo de classificação a um fator	43
3.8. Teste da variável adicionada	49
3.8.1. Modelo de regressão com dois parâmetros	53
3.8.2. Modelo de classificação a um fator com uma variável adicionada	59
3.9. Poder dos testes escore aperfeiçoados em modelos lineares generalizados	59
4. FÓRMULAS ESPECIAIS PARA O FATOR DE CORREÇÃO DA ESTATÍSTICA ESCORE	65
4.1. Modelos com parâmetros globalmente ortogonais	65
4.1.1. Introdução	65
4.1.2. Construção de parâmetros globalmente ortogonais	67
4.1.3. Testes escore aperfeiçoados	69
4.1.4. Casos especiais	71
4.1.4.1. Dois parâmetros unidimensionais ortogonais	71
4.1.4.2. Parâmetro de interesse unidimensional	79
4.1.4.3. Parâmetro de perturbação unidimensional	80
4.2. Problemas com parâmetros localmente ortogonais	81
4.3. Fórmulas matriciais para o fator de correção da estatística escore	83
4.3.1. Introdução	83
4.3.2. Fórmulas matriciais para os coeficientes	83
4.3.3. Uma aplicação	89
4.4. Algumas propriedades da estatística escore aperfeiçoada	90
5. TESTES ESCORE APERFEIÇOADOS EM MODELOS LINEARES GENERALIZADOS COM DISPERSÃO DESCONHECIDA E OUTRAS EXTENSÕES	93
5.1. Introdução	93
5.2. Testes de hipóteses sobre o parâmetro ϕ	95

Índice

5.3. Testes de hipóteses sobre β com ϕ desconhecido	103
5.4. Extensão aos modelos próprios de dispersão	105
5.4.1. A classe dos modelos próprios de dispersão	105
5.4.2. Testes de hipóteses sobre o parâmetro ϕ	107
5.4.3. Testes de hipóteses para β com ϕ desconhecido	115
5.5. Testes escore aperfeiçoados em modelos não-lineares da família exponencial	118
5.5.1. Introdução	118
5.5.2. Testes escore aperfeiçoados	120
5.5.3. Alguns casos especiais	123
6 FATORES DE CORREÇÃO DE BARTLETT GENERALIZADOS	125
6.1. Introdução	125
6.2. Estatísticas modificadas tendo distribuição χ^2 até ordem n^{-1}	126
6.2.1. Obtenção das estatísticas aperfeiçoadas	126
6.2.2. Casos especiais	129
6.3. Fatores de correção de Bartlett desenvolvidos a partir de momentos das estatísticas não modificadas	129
6.3.1. Obtenção de fatores de correção de Bartlett generalizados	129
6.3.2. Casos especiais	134
6.4. Obtenção de expansão assintótica da distribuição de uma estatística a partir de seus momentos	135
6.5. Aplicações	136
CONCLUSÕES	144
APÊNDICE 1	146
APÊNDICE 2	149

Índice

APÊNDICE 3	154
APÊNDICE 4	157
APÊNDICE 5	164
BIBLIOGRAFIA	172

CAPÍTULO 1

INTRODUÇÃO

A grande maioria dos testes de hipóteses estatísticas em problemas envolvendo grandes amostras é baseada em um dos seguintes princípios: razão de verossimilhança, Wald e escore.

O teste escore foi originalmente introduzido por Rao (1948) e posteriormente por Aitchison e Silvey (1958) e Silvey (1959) sob a denominação de teste do multiplicador de Lagrange e faz parte da classe dos testes $C(\alpha)$ sugerida por Neyman (1959, 1979). No passado, os testes da razão de verossimilhança e de Wald foram mais difundidos que o teste escore. No entanto, mais recentemente, as pesquisas envolvendo testes escore tiveram grande impulso quer em teoria quer em aplicações. Ainda, em muitas situações, o teste escore tem grande vantagem computacional sobre os demais. Enquanto os testes da razão de verossimilhança e de Wald envolvem estimação dos parâmetros sob a hipótese

alternativa, o teste escore, requer estimação somente sob a hipótese nula. Sendo assim, nos casos em que a estimação sob a hipótese alternativa é complicada envolvendo, por exemplo, estimação não linear e processos iterativos, o teste escore pode apresentar um custo computacional bastante inferior aos demais. No caso em que a hipótese nula é simples, o teste escore não requer estimação alguma.

Sabe-se que, em problemas regulares, as estatísticas da razão de verossimilhança (ω), de Wald (W) e escore (S_R) têm, sob a hipótese nula, distribuição assintótica χ_r^2 , ou seja, χ^2 com r graus de liberdade, onde r é a diferença entre as dimensões dos espaços paramétricos sob as hipóteses alternativa e nula. Assim, os testes são baseados na comparação das estatísticas com o valor crítico obtido da distribuição χ^2 de referência para um nível de significância nominal fixado. Valores altos das estatísticas indicam fraca consistência dos dados com a hipótese nula. Propriedades de otimalidade assintótica destes testes são encontradas em Cox e Hinkley (1974, Seção 9.3) e Engle (1984, Seção 6).

As estatísticas ω , W e S_R são assintoticamente equivalentes no sentido de que têm a mesma distribuição assintótica não só sob a hipótese nula mas também sob alternativas contíguas (*) (Cox e Hinkley, 1974, Seção 9.3). A escolha entre as estatísticas precisa, então, ser baseada em critérios como vantagens computacionais, comparação de poderes até segunda ou terceira ordem e desempenho dos testes em amostras pequenas e moderadas, investigados, em geral, via simulação (Godfrey, 1981; Bera e McKenzie, 1986; Ohtani e Toyoda, 1985).

As pesquisas sobre os poderes dos testes sob alternativas contíguas têm sido desenvolvidas sob dois enfoques diferentes. O primeiro leva em conta que, até primeira ordem, os três testes têm o mesmo poder e que, até ordem $n^{-1/2}$, têm o mesmo tamanho; faz sen-

(*) Uma hipótese alternativa contígua à hipótese nula $H_0 : \theta = \theta^{(0)}$ tem a forma $H : \theta = \theta^{(0)} + \xi$, onde ξ é de ordem $n^{-1/2}$ com n representando o tamanho da amostra. Aqui θ é um parâmetro desconhecido e $\theta^{(0)}$ é um número real fixado. A definição é estendida de maneira evidente para parâmetros multidimensionais.

tido, portanto, comparar as parcelas de ordem $n^{-1/2}$ das expansões dos poderes dos testes. Sob este enfoque, o trabalho pioneiro é o artigo de Peers (1971) que desenvolve expansões até ordem $n^{-1/2}$ para a função de poder dos três testes sob alternativas contíguas, no caso de hipóteses nulas simples. Os trabalhos que se seguiram são extensões dos resultados de Peers a hipóteses compostas: Hayakawa (1975) trata dos testes da razão de verossimilhança e de Wald enquanto que Harris e Peers (1980) tratam do teste escore. De uma análise destes artigos, conclui-se que não é possível, em geral, afirmar que um dos testes é mais poderoso que os demais. Somente em situações particulares podem ser feitas comparações. O segundo enfoque foi proposto por Chandra e Joshi (1983). A idéia é modificar convenientemente as regiões críticas dos testes escore e de Wald para que os testes modificados venham a ter o mesmo tamanho que o da razão de verossimilhança até ordem n^{-1} . Ao menos no caso uniparamétrico, os três testes passam a ter o mesmo poder até ordem $n^{-1/2}$ e torna-se relevante comparar as parcelas de ordem n^{-1} das expansões dos poderes dos testes. Desta comparação, Chandra e Joshi (1983) concluem que o teste escore é o mais poderoso quando termos de ordem n^{-1} são levados em conta. Estendendo este trabalho, Chandra e Mukerjee (1984, 1985) mostram a superioridade do teste escore em uma ampla classe de estatísticas. Extensões para o caso multiparamétrico são encontradas nos artigos de Mukerjee (1990a, b).

A estatística escore é invariante com respeito a transformação de parâmetros (del Pino, 1990) e também a certas modificações na hipótese alternativa. Godfrey (1981) observou que, para determinados problemas, as estatísticas escore para testar uma mesma hipótese nula contra diferentes alternativas são idênticas. Esta falta de sensibilidade da estatística S_R a mudanças na alternativa parecia, a princípio, levar a um mau desempenho dos testes escore em amostras pequenas e moderadas. No entanto, os estudos de simulação citados acima não têm sustentado esta suspeita.

Breusch (1979), generalizando os trabalhos de Savin (1976) e Berndt e Savin (1977), mostrou que, em testes de restrições lineares dos parâmetros da regressão em

certos modelos normais lineares, é válida a relação $W \geq \omega \geq S_R$ para quaisquer conjuntos de dados (vide também Evans e Savin, 1982). Isto indica que o teste escore é o mais conservativo. É conveniente salientar, no entanto, que não se pode concluir que o teste escore é o menos poderoso pois os testes têm tamanhos diferentes. Claro está que os testes baseados em W , ω e S_R em amostras finitas podem levar a decisões conflitantes para certos níveis nominais. A igualdade $W = \omega = S_R$ entre as três estatísticas é válida para log-verossimilhanças quadráticas (Engle, 1984).

Os testes escore têm sido bastante utilizados em várias áreas da Estatística. Na análise de dados sob a forma de contagens dispostos em tabelas de contingência, os testes usuais conhecidos como χ^2 de Pearson são testes escore. Comparações entre a estatística de Pearson e outros critérios competitivos são encontradas em Chapman (1976), Larntz (1978), Margolin e Light (1974), Fienberg (1979), Koehler e Larntz (1980), Lawal (1984), Cressie e Read (1984) e referências citadas nestes artigos. O teste da razão de verossimilhança é, sem dúvida, o mais utilizado em modelos lineares generalizados (MLGs) (McCullagh e Nelder, 1989), entretanto, inúmeras aplicações de testes escore nestes modelos são encontradas na literatura. Pregibon (1982) mostrou que a estatística escore pode ser calculada no GLIM (Payne, 1986), sistema computacional largamente difundido na análise de dados via os MLGs, com a mesma facilidade que se calcula a estatística da razão de verossimilhança. No Capítulo 3, fazemos uma breve revisão bibliográfica sobre o uso de testes escore em MLGs.

Encontram-se também aplicações de testes escore em modelos de séries temporais (Hosking, 1980, 1981 e Poskitt e Tremayne 1981, 1982) e, em particular, em modelos econométricos. Breusch e Pagan (1980) e Engle (1984) fazem uma revisão bibliográfica sobre testes escore e apresentam vários exemplos em Econometria. Em análise de sobrevivência, o livro de Lawless (1982) (vide também O'Quigley e Pessione, 1989) mostra várias aplicações de testes escore nos modelos de riscos proporcionais de Cox. Em particular, o teste usual de Mantel-Haenzel é um teste escore (Day e Byar, 1979). Schiffman (1986)

ênfatiza a simplicidade de cálculo da estatística escore para testar se a taxa é constante em modelos com taxa de falha discreta.

Há, em geral, grande dificuldade em se determinar as distribuições exatas das estatísticas ω , W e S_R . Esta é a razão pela qual os testes têm sido construídos com base em resultados assintóticos. Como já mencionamos, as três estatísticas têm, aproximadamente, distribuição χ_r^2 em amostras grandes. Há várias décadas, surgiu a preocupação de se tentar melhorar a qualidade da aproximação por χ^2 para a distribuição da estatística da razão de verossimilhança. A idéia, originalmente sugerida por Bartlett (1937) para o seu teste de homogeneidade de variâncias, é baseada em um princípio bastante simples. Suponha que, sob a hipótese nula, $E(\omega) = r\{1 + c + O(n^{3/2})\}$, onde n é o tamanho da amostra e c é uma constante de ordem n^{-1} . Então, a média da estatística modificada $\omega' = \omega(1 + c)^{-1}$ ou, equivalentemente, $\omega^* = \omega(1 - c)$ é mais próxima daquela da distribuição χ_r^2 do que a média de ω . A constante c é, em geral, função de parâmetros desconhecidos que devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula.

Lawley (1956) derivou uma fórmula para c em função de momentos das quatro primeiras derivadas da log-verossimilhança e mostrou que, em problemas regulares, a estatística ω^* tem todos os seus momentos coincidindo com os da distribuição χ^2 de referência até termos de ordem n^{-1} . Posteriormente, Hayakawa (1977) derivou uma expansão assintótica para a distribuição de ω sob a hipótese nula. Hayakawa mostrou que, se a hipótese nula for simples, a estatística $\omega^* = \omega(1 - c)$ tem distribuição χ^2 até ordem n^{-1} . A constante c , embora escrita de forma diferente, coincide com aquela obtida por Lawley. Entretanto, para hipóteses compostas, só seria possível obter o fator de correção se um determinado coeficiente da expansão fosse nulo. Isto parecia estar em conflito com os resultados de Lawley. A reconciliação dos trabalhos de Hayakawa e Lawley só foi estabelecida vários anos depois devido ao artigo de Cordeiro (1987) que mostrou que, sob condições gerais de regularidade, o tal coeficiente é identicamente zero.

Quase 45 anos se passaram, desde a idéia inicial de Bartlett em 1937, para que a pesquisa em fatores de correção para a estatística da razão de verossimilhança tivesse acentuado impulso e se consolidasse. Este impulso foi motivado na década de 80, mais precisamente, a partir de 1982, em especial, na Inglaterra e Dinamarca. Nessa década, as seguintes pessoas (em ordem alfabética) contribuíram decisivamente para o crescimento das pesquisas na área de correções de Bartlett: O.E.Barndorff-Nielsen, P.Blaesild, T.K.Chandra, G.M.Cordeiro, D.R.Cox, T.J.DiCiccio, J.K.Ghosh, P.Hall, P.Harris, J.L.Jensen, S.Johansen, P.McCullagh, J.Møller, B.T.Porteous, W.H.Ross, I.N.Skovgaard e P.W.Vos. No passado, o próprio M.S.Bartlett, G.E.P.Box, T.Hayakawa, D.N.Lawley, S.A.Sharp e D.A.Williams deram os primeiros passos em trabalhos nesta área.

Atualmente, as pesquisas em fatores de Bartlett seguem quatro direções principais: a primeira, na obtenção de fórmulas algébricas simples das correções em modelos especiais; a segunda, no desenvolvimento de métodos alternativos gerais de cálculo destes fatores; a terceira, em estabelecer a interpretação dos fatores de Bartlett à luz da geometria diferencial e em conectar estes fatores com tópicos de interesse recente, como ortogonalização de parâmetros, verossimilhança perfilada, tensores, etc; a quarta, em aplicações numéricas a diversas áreas e estudos de simulação.

Na primeira direção, aparecem maior número de artigos. Convém aqui citar os fatores de Bartlett desenvolvidos para os modelos lineares generalizados e extensões (Cordeiro, 1983, 1985, 1987; Cordeiro e Paula, 1989b; Cordeiro, Paula e Botter, 1991) e para modelos multivariados normais com estruturas de covariância especiais (Porteous, 1985, Møller, 1986).

Na segunda direção o artigo de Barndorff-Nielsen e Cox (1984) é o mais importante e trata de um método indireto para obter o fator de Bartlett em termos das constantes normalizadas da distribuição condicional do estimador de máxima verossimilhança. Outras estratégias para a obtenção do fator de Bartlett foram desenvolvidas por Barndorff-Nielsen e Blaesild (1986) e McCullagh e Cox (1986). Vários trabalhos, ainda não publicados, foram

desenvolvidos mais recentemente por diversos pesquisadores (D.F.Andrews, J.E.Stafford, Y.Wang, J.L.Jensen e G.M.Cordeiro entre outros) visando tornar as correções de Bartlett mais atraentes sob o ponto de vista de cálculo algébrico.

Na terceira direção os artigos de McCullagh e Cox (1986), Barndorff-Nielsen (1986) e Ross (1987) são os mais importantes por interpretarem os fatores de Bartlett em termos de quantidades de interesse da geometria diferencial. Outras conexões interessantes dos fatores de Bartlett podem ser vistas em McCullagh (1984), Cox (1984) e Ross (1987). Convém aqui salientar o Capítulo 7 do livro de McCullagh (1987) e o Capítulo 10 do livro recente em honra a Sir David Cox (Barndorff-Nielsen, 1990).

Na quarta direção muitos pesquisadores estão atuando no momento mas nenhuma publicação de peso tem sido encontrada.

Excelentes revisões bibliográficas sobre correções de Bartlett para a estatística da razão de verossimilhança são encontradas nos artigos de Reid (1988), Cordeiro e Paula (1989b) e Cordeiro, Paula e Botter (1991).

A estatística escore não admite um fator de correção simples como o da razão de verossimilhança. Este fato tem sido frequentemente citado na literatura (Harris, 1985, 1987; Cox, 1988; Barndorff-Nielsen, 1990).

Harris (1985), seguindo as idéias de Hayakawa (1977), desenvolve uma expansão assintótica até ordem n^{-1} para a distribuição da estatística escore. Dada a impossibilidade de obter um fator de correção simples para S_R , Harris propôs um teste escore aperfeiçoado baseado na própria estatística escore e em uma expansão, até ordem n^{-1} , para seus quantis. Em outras palavras, em lugar de modificar a estatística do teste, Harris propôs a modificação do valor crítico obtido da distribuição χ^2 de referência para um nível de significância nominal fixado.

Chandra e Mukerjee (1991) encontraram uma estatística escore modificada tendo distribuição χ^2 até ordem n^{-1} apenas no caso de hipóteses simples. Não obtiveram, no entanto, tal estatística para situações mais gerais como, por exemplo, aquela em que a

hipótese nula é composta e há mais do que um parâmetro de perturbação.

No Capítulo 2, discutimos a expansão assintótica para a distribuição de S_R obtida por Harris e mostramos que é sempre possível encontrar uma estatística escore aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} . A estatística modificada é dada por S_R multiplicada por um fator de correção que envolve um polinômio (de segundo grau na própria estatística e cujos coeficientes são, em geral, funções de parâmetros desconhecidos que devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula. Os resultados obtidos são bastante abrangentes no sentido de que são válidos para problemas uni e multiparamétricos, para hipóteses simples e compostas e generalizam os resultados recentes de Chandra e Mukerjee (1991). A obtenção da estatística aperfeiçoada pode ser feita por dois enfoques diferentes. O primeiro segue da expansão da função geradora de momentos da estatística escore modificada. O segundo parte do Teorema 1 de Cox e Reid (1987a). Indo mais além, mostramos que existe uma estreita conexão entre a expressão da estatística modificada e a expansão assintótica dos quantis de S_R obtida por Harris (1985) através da fórmula geral de inversão de expansões de Hill e Davis (1968). Mostramos também que a expansão de Harris pode ser obtida diretamente de Teorema 1 de Cox e Reid (1987a) sem necessidade do uso da fórmula de Hill e Davis. Ainda neste capítulo, fazemos alguns comentários sobre os poderes dos testes escore aperfeiçoados.

O fator de correção para a estatística escore envolve, em geral, funções complicadas de cumulantes conjuntos de derivadas da log-verossimilhança até quarta ordem. É fundamental, portanto, derivar fórmulas simplificadas para estes fatores em classes importantes de modelos. No Capítulo 3, deduzimos fórmulas matriciais para os fatores de correção de testes escores sobre parâmetros da regressão em MLGs com dispersão conhecida. As fórmulas são de fácil computação pois envolvem somente operações simples entre matrizes e vetores. Apresentamos aplicações a uma série de modelos especiais e estudos de simulação. Além disto, obtemos fórmulas, também em notação matricial, para os poderes dos

testes escore aperfeiçoados.

No Capítulo 4, encontramos fórmulas para os fatores de correção para problemas com parâmetros ortogonais (Cox e Reid, 1987b). Obtemos também fórmulas matriciais para o teste de uma hipótese nula geral simples ou composta que podem ser de grande utilidade no desenvolvimento de programas computacionais para obtenção de fórmulas algébricas para os fatores de correção. Ainda neste capítulo, analisamos algumas características da estatística escore aperfeiçoada.

No Capítulo 5, com base nos resultados obtidos para problemas com parâmetros ortogonais, derivamos fatores de correção para o teste do parâmetro de dispersão e do conjunto completo dos parâmetros da regressão em MLGs com dispersão desconhecida. Estendemos os resultados à classe dos modelos próprios de dispersão (Barndorff-Nielsen e Jørgensen, 1990) e apresentamos um estudo de simulação. Ainda neste capítulo, obtemos fatores de correção para testes em modelos não-lineares da família exponencial.

Temos forte convicção de que novos resultados poderão ser obtidos de forma análoga ao que tem sido feito para a razão de verossimilhança. Esta nossa crença deve-se às inúmeras similaridades entre os fatores de correção da estatística escore e os tradicionais fatores de Bartlett. Acreditamos ainda que as quatro direções em que têm caminhado as pesquisas sobre os fatores de correção para a razão de verossimilhança serão seguidas, a médio prazo, pelos trabalhos em correções para a estatística escore.

No Capítulo 6, estendemos os resultados do Capítulo 2 a uma ampla classe de estatísticas que têm distribuição χ^2 assintoticamente. Esta classe envolve não só as estatísticas escore e da razão de verossimilhança, como também a de Wald e formas assintoticamente equivalentes à estatística escore (Bera e McKenzie, 1986). Novamente, o Teorema 1 de Cox e Reid (1987a) é a base da obtenção do que chamamos de “fatores de correção de Bartlett generalizados” para as estatísticas desta classe. Propomos ainda uma forma alternativa de obter tais fatores a partir de expansões de momentos das estatísticas.

Finalmente, apresentamos nossas conclusões e sugerimos diversas linhas de pesquisa que podem ser desenvolvidas a partir das idéias desta tese.

CAPÍTULO 2

UMA ESTATÍSTICA ESCORE APERFEIÇOADA TENDO DISTRIBUIÇÃO χ^2 ATÉ ORDEM n^{-1}

2.1. INTRODUÇÃO

Neste capítulo, apresentamos uma expansão assintótica até ordem n^{-1} , onde n é o tamanho da amostra, para a distribuição da estatística escore obtida por Harris (1985) e mostramos que é sempre possível encontrar uma estatística escore modificada tendo distribuição χ^2 até esta ordem.

Consideremos o vetor $y = (y_1, \dots, y_n)^T$ representando n observações independentes com densidades contínuas e cujo logaritmo da função de verossimilhança total $L(\beta)$ depende de um parâmetro desconhecido $\beta = (\beta_1, \dots, \beta_p)^T$ com p componentes. Assume-se que a função $L(\beta)$ seja regular com respeito às derivadas em relação aos componentes de β até quarta ordem.

$$L(\beta) = \sum_{l=1}^n \log \pi(y_l, \beta).$$

O problema considerado é o de testar a hipótese nula simples ou composta $H_0 : \beta_1 = \beta_1^{(0)}$ contra a hipótese alternativa composta $H : \beta_1 \neq \beta_1^{(0)}$, onde $\beta = (\beta_1^T, \beta_2^T)^T$, $\beta_1 = (\beta_1, \dots, \beta_r)^T$ é o vetor r dimensional de parâmetros de interesse, $\beta_2 = (\beta_{r+1}, \dots, \beta_p)^T$ é o vetor de parâmetros de perturbação e $\beta_1^{(0)}$ é um vetor especificado de dimensão r ($r \leq p$).

Seja $U(\beta) = \partial L(\beta) / \partial \beta = (U_1^T(\beta_1, \beta_2), U_2^T(\beta_1, \beta_2))^T$ a função escore total para β particionada da mesma maneira que β . Denotamos por $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$ o estimador de máxima verossimilhança irrestrito de β e por $\tilde{\beta}_2$ o estimador de máxima verossimilhança de β_2 sob H_0 . Funções avaliadas em $\beta = (\beta_1^{(0)T}, \tilde{\beta}_2^T)^T$ serão denotadas com um til. A partição $\beta = (\beta_1^T, \beta_2^T)^T$ conduz às seguintes partições

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}, \quad K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}, \quad (2.1)$$

onde $K = E\{U(\beta)U^T(\beta)\}$ é a matriz de informação total de Fisher que supomos ser positiva definida e K^{-1} é a sua inversa.

A estatística escore S_R (Rao, 1948) para testar H_0 versus H é $S_R = \tilde{U}^T \tilde{K}^{-1} \tilde{U}$. Buse (1982) dá uma interpretação geométrica no caso de hipótese nula simples para a estatística escore e compara-a com as estatísticas da razão de verossimilhança e de Wald. Cox e Hinkley (1974, Seção 9.3.) expressam S_R como

$$S_R = \tilde{U}_1^T \tilde{K}^{11} \tilde{U}_1, \quad (2.2)$$

onde $K^{11} = K^{11}(\beta_1, \beta_2)$ é a matriz de covariância assintótica de $\hat{\beta}_1$ obtida de (2.1). A equivalência das duas expressões para S_R é facilmente observada notando-se que $\tilde{U}_2$ é igual a um vetor nulo.

Em grandes amostras, satisfeitas certas condições de regularidade, a distribuição sob H_0 de S_R , bem como das estatísticas da razão de verossimilhança e de Wald, é aproximadamente χ_r^2 , isto é, χ^2 com r graus de liberdade.

A estatística da razão de verossimilhança $\omega = 2\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1^{(0)}, \tilde{\beta}_2)\}$ para testar H_0 versus H pode ser aperfeiçoada por um fator de correção de Bartlett que faz com que a estatística modificada $\omega^* = \omega\{1 - b(\beta)\}$ tenha distribuição χ_r^2 até ordem n^{-1} sob H_0 . Em outras palavras, a expansão da função distribuição de ω^* , sob a hipótese nula, coincide com a da distribuição χ_r^2 quando termos de ordem inferior a n^{-1} são desprezados. Este fator de correção está relacionado com a expansão até ordem n^{-1} da média de ω sob H_0 dada por $r\{1 + b(\beta)\}$. Lawley (1956) mostrou que a multiplicação de ω pelo fator $1 - b(\beta)$ tem a propriedade de remover o termo de ordem n^{-1} das expansões de todos os momentos de ω^* . No entanto, a estatística S_R não admite um fator multiplicativo com esta propriedade (Harris, 1985, 1987; Cox, 1988; Barndorff-Nielsen, 1990). Neste capítulo mostramos que é possível encontrar um fator de correção do tipo Bartlett para S_R que é, no entanto, função da própria estatística.

Na Seção 2.2, apresentamos os resultados de Hayakawa (1977) relativos à distribuição assintótica da estatística da razão de verossimilhança. Na Seção 2.3, as expansões para a função distribuição da estatística escore e para seus quantis são apresentadas. Na Seção 2.4, com base no trabalho de Harris, derivamos uma estatística escore aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} e, na Seção 2.5, exibimos esta estatística no caso uniparamétrico e damos um exemplo para a distribuição exponencial. Finalmente, na Seção 2.6, fazemos um comentário sobre o poder dos testes escore aperfeiçoados.

2.2. EXPANSÃO ASSINTÓTICA PARA A FUNÇÃO DISTRIBUIÇÃO DA ESTATÍSTICA DA RAZÃO DE VEROSSIMILHANÇA

Hayakawa (1977), estendendo os trabalhos de Box (1949) e Peers (1971), obteve uma expansão assintótica até ordem n^{-1} para a função distribuição da estatística da razão de verossimilhança sob a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$. A expressão “até ordem n^{-1} ” é utilizada para indicar que termos de ordem inferior a n^{-1} são ignorados. Hayakawa

mostrou que a função distribuição de ω , sob H_0 , pode ser escrita como

$$F_{\omega}(x) = G_r(x) + \frac{1}{24} \{a_2 G_{r+4}(x) - (2a_2 - a_1) G_{r+2}^{(\alpha)} + (a_2 - a_1) G_r(x)\} \quad (2.3)$$

até ordem n^{-1} . Aqui $G_m(\cdot)$ denota a função distribuição da χ_m^2 , e a_1 e a_2 são funções de cumulantes conjuntos de derivadas da log-verossimilhança e são de ordem n^{-1} . O procedimento utilizado por Hayakawa para a obtenção de (2.3) é o seguinte: (i) desenvolve-se uma expansão em série para ω em termos de derivadas de $L(\beta)$ até quarta ordem, retendo-se termos até $O_p(n^{-1})$; (ii) obtém-se uma expansão de Edgeworth multivariada para a densidade conjunta destas derivadas; (iii) a partir desta expansão, encontra-se a função geradora de momentos até ordem n^{-1} da expansão em série de ω ; (iv) finalmente, invertendo a função geradora de momentos, chega-se à expansão da função distribuição de ω dada em (2.3). A validade deste procedimento, baseado em expansões formais, foi estabelecida por Chandra e Ghosh (1979) sob condições de regularidade adequadas (vide também os trabalhos de Chandra, 1985 e Jensen, 1991, Seção 1.2).

Hayakawa observa que se e somente se $a_2 = 0$, a estatística da razão de verossimilhança ω pode ser aperfeiçoada pelo fator de correção de Bartlett $1 - b(\beta)$, citado anteriormente, de forma que $\omega^* = \omega\{1 - b(\beta)\}$ tenha distribuição χ^2 até ordem n^{-1} . Isto parecia estar em conflito com os resultados de Lawley (1956) que mostrou que ω^* tem todos os seus momentos coincidindo, até ordem n^{-1} , com os correspondentes momentos da distribuição χ^2 . No entanto, Cordeiro (1987) prova que $a_2 = 0$ e faz, portanto, uma reconciliação entre os resultados de Hayakawa e Lawley. Assim sendo, a expansão em (2.1) reduz-se a

$$F_{\omega}(x) = G_r(x) + \frac{1}{24} \{a_1 G_{r+2}(x) - a_1 G_r(x)\} \quad (2.4)$$

até ordem n^{-1} .

Barndorff-Nielsen e Hall (1988) mostram que o erro da expansão em (2.4) é $O(n^{-2})$ e que, portanto,

$$F_{\omega^*}(x) = G_r(x) + O(n^{-2}).$$

Para finalizar esta seção, é conveniente salientar que a expansão em (2.4) não é válida para distribuições “lattice” e que, portanto, o fator de Bartlett não deve, em geral, ter a propriedade de melhorar a qualidade da aproximação da distribuição da estatística por qui-quadrado para dados discretos. No entanto, Cordeiro (1982) faz vários estudos de simulação envolvendo distribuições multinomial e de Poisson que mostram que os testes modificados, baseados em ω^* , apresentam taxas de rejeição da hipótese nula bem mais próximas dos respectivos níveis nominais do que os testes não modificados, baseados em ω . Por outro lado, Frydenberg e Jensen (1989) mostram, via simulação, que, em situações particulares, os desempenhos dos testes modificados chegam a ser piores que os dos não modificados. Logo, há que se verificar com maior profundidade em que situações o uso de fatores de Bartlett é adequado no caso discreto. Já no caso misto, isto é, quando a log-verossimilhança envolve variáveis discretas e contínuas, a validade de (2.4) é verificada desde que a estatística ω dependa essencialmente da parte contínua (Jensen, 1991). Este resultado é particularmente útil em modelos de análise de sobrevivência com dados censurados e em regressão logística.

2.3. EXPANSÕES ASSINTÓTICAS DE HARRIS

A expansão assintótica até ordem n^{-1} da distribuição da estatística escore S_R para o teste discutido na Seção 2.1, sob a hipótese nula, foi obtida por Harris (1985), utilizando um procedimento análogo ao de Hayakawa (1977) como descrevemos acima. Daqui em diante, assumimos satisfeitas as condições de regularidade que garantem a validade das expansões de Hayakawa e Harris (Chandra e Ghosh, 1979).

Introduzimos a seguinte notação para as derivadas da log-verossimilhança total nas quais todos os sufixos variam pelos inteiros $1, \dots, p$: $U_r = \partial L / \partial \beta_r$, $U_{rs} = \partial^2 L / \partial \beta_r \partial \beta_s$, etc. Adotamos a notação tensorial para os cumulantes conjuntos de derivadas da log-verossimilhança (Lawley, 1956; Peers, 1971; Hayakawa, 1977; Harris, 1985; McCullagh,

1984, 1987): $\kappa_{rs} = E(U_{rs})$, $\kappa_{rst} = E(U_{rst})$, $\kappa_{r,s} = E(U_r U_s)$, $\kappa_{rs,t} = E(U_{rs} U_t)$, $\kappa_{rs,tu} = E(U_{rs} U_{tu}) - \kappa_{rs} \kappa_{tu}$, $\kappa_{r,s,tu} = E(U_r U_s U_{tu}) - \kappa_{r,s} \kappa_{tu}$ e $\kappa_{r,s,t,u} = E(U_r U_s U_t U_u) - \kappa_{r,s} \kappa_{t,u} - \kappa_{r,t} \kappa_{s,u} - \kappa_{r,u} \kappa_{s,t}$. Definimos as derivadas dos cumulantes por $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \beta_t$, $\kappa_{rs}^{(tu)} = \partial^2 \kappa_{rs} / \partial \beta_t \partial \beta_u$, etc. Todos os κ 's referem-se a um total sobre a amostra e são, em geral, de ordem n . A matriz de informação total de Fisher K tem elementos $\kappa_{r,s} = -\kappa_{rs}$, sendo $\kappa^{r,s} = -\kappa^{rs}$ os correspondentes elementos de sua inversa.

Os cumulantes satisfazem relações (Bartlett, 1953a, b; Lawley, 1956) que frequentemente facilitam seus cálculos, tais como $-\kappa_{r,s} = \kappa_{rs}$, $\kappa_{st}^{(r)} + \kappa_{r,s,t} + \kappa_{rs,t} + \kappa_{rt,s} = 0$, $\kappa_{rst}^{(u)} = \kappa_{rstu} + \kappa_{rst,u}$, $\kappa_{r,s,t} = -\kappa_{rst} - \sum_{(3)} \kappa_{r,st}$, $\kappa_{r,s,t} = 2\kappa_{rst} - \sum_{(3)} \kappa_{rs}^{(t)}$ e $\kappa_{r,s,t,u} = -3\kappa_{rstu} + 2 \sum_{(4)} \kappa_{rst}^{(u)} - \sum_{(6)} \kappa_{rs}^{(tu)} + \sum_{(3)} \kappa_{rs,tu}$, e $\kappa_{r,s,tu} = 2\kappa_{rstu} - \kappa_{rtu}^{(s)} + \kappa_{tu}^{(rs)} - \kappa_{rs,tu}$, onde $\sum_{(k)}$ denota somatório que varre todas as k permutações de índices. Assim, $\sum_{(3)} \kappa_{rs,tu} = \kappa_{rs,tu} + \kappa_{rt,su} + \kappa_{ru,st}$.

$\kappa_{rs,tu} - \kappa_{rstu} = -\kappa_{rstu}$

Sejam F_{S_R} a função distribuição de S_R , G_m a função distribuição da χ_m^2 , e f_{S_R} e g_m suas respectivas funções densidade. Harris (1985) mostrou que até ordem n^{-1} e sob H_0

$$F_{S_R}(x) = G_r(x) + \frac{1}{24} \{ A_3 G_{r+6}(x) + (A_2 - 3A_3) G_{r+4}(x) + (3A_3 - 2A_2 + A_1) G_{r+2}(x) + (A_2 - A_1 - A_3) G_r(x) \}, \quad (2.5)$$

onde

$$A_1 = 3 \sum' (\kappa_{ijk} + 2\kappa_{i,jk}) (\kappa_{rst} + 2\kappa_{rs,t}) a_{ij} a_{st} m_{kr} - 6 \sum' (\kappa_{ijk} + 2\kappa_{i,jk}) \kappa_{r,s,t} a_{ij} a_{kr} m_{st} + 6 \sum' (\kappa_{i,jk} - \kappa_{i,j,k}) (\kappa_{rst} + 2\kappa_{rs,t}) a_{js} a_{kt} m_{ir} - 6 \sum' (\kappa_{i,j,k,r} + \kappa_{i,j,kr}) a_{kr} m_{ij}, \quad (2.6)$$

$$A_2 = -3 \sum' \kappa_{i,j,k} \kappa_{r,s,t} a_{kr} m_{ij} m_{st} + 6 \sum' (\kappa_{ijk} + 2\kappa_{i,jk}) \kappa_{r,s,t} a_{ij} m_{kr} m_{st} - 6 \sum' \kappa_{i,j,k} \kappa_{r,s,t} a_{kt} m_{ir} m_{js} + 3 \sum' \kappa_{i,j,k,r} m_{ij} m_{kr} \quad (2.7)$$

e

$$A_3 = 3 \sum' \kappa_{i,j,k} \kappa_{r,s,t} m_{ij} m_{kr} m_{st} + 2 \sum' \kappa_{i,j,k} \kappa_{r,s,t} m_{ir} m_{js} m_{kt}, \quad (2.8)$$

sendo a_{ij} e m_{ij} os elementos (i, j) das matrizes

$$A = \begin{pmatrix} 0 & 0 \\ 0 & K_{22}^{-1} \end{pmatrix}, \quad M = K^{-1} - A \quad (2.9)$$

respectivamente. Usamos $\sum'$ para indicar o somatório sobre todos os β 's. Note que o termo $-6K_{\cdot,\cdot,\cdot} * M * M * J * K_{\cdot,\cdot,\cdot}$ na expressão de A_2 de Harris está incorreto devendo ser substituído por $-6\kappa_{\cdot,\cdot,\cdot} * M * M * J * \kappa_{\cdot,\cdot,\cdot}$. Os A 's são de ordem n^{-1} e são avaliados sob H_0 . As expressões em (2.6)–(2.8) parecem ser funções bastante complicadas dos cumulantes κ 's da log-verossimilhança total. No entanto, temos notado que expressões em forma fechada podem ser obtidas para modelos especiais. Nos Capítulos 3 a 5 encontramos expressões relativamente simples para os A 's para testes em diversas classes de modelos. Harris notou que, no caso de hipótese nula simples, $A_1 = 0$ e A_2 e A_3 refletem aspectos de não-normalidade da distribuição, sob H_0 , da função escore $U(\beta)$. Para hipóteses nulas compostas a interpretação dos A 's é complicada (Harris, 1985).

Ele também obteve expansões até ordem n^{-1} para os três primeiros momentos de S_R . Das suas expansões tem-se que, até esta ordem e sob H_0 , os três primeiros momentos de S_R são dados por

$$\begin{aligned} \mu'_1(S_R) &= r + \frac{A_1}{12}, \\ \mu'_2(S_R) &= r(r+2) + \frac{A_1(r+2) + 2A_2}{6n}, \\ \mu'_3(S_R) &= r(r+2)(r+4) + \frac{A_1(r+2)(r+4) + 4A_2(r+4) + 8A_3}{4}. \end{aligned} \quad (2.10)$$

Harris observou que não é possível obter um fator de correção simples para a estatística escore de tal forma que a estatística modificada, dada por S_R multiplicada por este fator, tenha distribuição χ^2 até ordem n^{-1} . Esta mesma observação é feita por Cox (1988). Mesmo mais recentemente, Barndorff-Nielsen (1990; pg 249), referindo-se ao fato

de se poder aperfeiçoar a estatística da razão de verossimilhança pelo fator de correção de Bartlett, escreve: “ This is due to the special character of the likelihood ratio statistic, and the same device applied to, for instance, the score test does not have a similar effect”. Então em lugar de modificar a estatística, Harris obteve uma expansão para os quantis de S_R . Através da fórmula de inversão de expansões de Hill e Davis (1968), ele mostrou que, desprezando quantidades de ordem inferior a n^{-1} , pode-se escrever

$$z_\alpha = x_\alpha + \frac{1}{12} \left[\frac{A_3 x_\alpha}{r(r+2)(r+4)} \{x_\alpha^2 + (r+4)x_\alpha + (r+2)(r+4)\} + \frac{x_\alpha(x_\alpha + r+2)}{r(r+2)} (A_2 - 3A_3) + \frac{x_\alpha}{r} (3A_3 - 2A_2 + A_1) \right] \quad (2.11)$$

com $\alpha = P(S_R > z_\alpha) = 1 - G_r(x_\alpha)$.

Desde que os A 's possam ser calculados, um teste escore aperfeiçoado pode ser conduzido encontrando o nível descritivo do teste através da expansão assintótica da função distribuição de S_R em (2.5) ou fixando-se um nível de significância α e comparando o valor observado de S_R com o valor crítico modificado z_α em (2.11).

2.4. UMA ESTATÍSTICA ESCORE APERFEIÇOADA

Mostramos nesta seção que, sob condições gerais de regularidade, a estatística escore S_R pode ser modificada por um fator de correção que envolve constantes de ordem n^{-1} e um polinômio de segundo grau na própria estatística produzindo uma estatística ajustada com distribuição χ^2 até ordem n^{-1} segundo a hipótese nula.

Escrevendo a expansão assintótica (2.5) em função da densidade de S_R e usando a relação de recorrência $g_{m+2}(x) = \frac{x}{m} g_m(x)$ encontra-se, até ordem n^{-1} ,

$$f_{S_R}(x) = g_r(x)(1 + B_0 + B_1 x + B_2 x^2 + B_3 x^3), \quad (2.12)$$

onde $B_0 = (A_2 - A_1 - A_3)/24$, $B_1 = (3A_3 - 2A_2 + A_1)/(24r)$, $B_2 = (A_2 - 3A_3)/\{24r(r+2)\}$ e $B_3 = A_3/\{24r(r+2)(r+4)\}$.

A partir da expansão assintótica em (2.4) para a função distribuição da estatística da razão de verossimilhança e da relação de recorrência acima, vemos que a expansão para sua função densidade sob H_0 é dada por $g_r(x)(1 + b_0 + b_1 x)$, onde b_0 e b_1 são de ordem n^{-1} . Como já mencionamos a estatística modificada $\omega^* = \omega(1 - b)$, onde b é de ordem n^{-1} , tem distribuição χ^2 até esta ordem. Aqui, $1 - b$ é o fator de correção de Bartlett usual para a estatística da razão de verossimilhança e é, em geral, função de parâmetros desconhecidos. Em vista disto e ao notar que (2.12) envolve um polinômio de terceiro grau, parece ser plausível definir uma estatística escore modificada por

$$S_R^* = S_R \{1 - (c + bS_R + aS_R^2)\}, \quad (2.13)$$

onde o fator multiplicativo entre chaves é um tipo de correção de Bartlett como função da própria estatística S_R . A idéia é determinar a , b e c , de ordem n^{-1} , como função dos B 's de modo que a função distribuição de S_R^* coincida com a da χ_r^2 até ordem n^{-1} . Isto pode ser feito expandindo a função geradora de momentos de S_R^* até ordem n^{-1} ou através do Teorema 1 de Cox e Reid (1987a) sob condições apropriadas de regularidade.

A função geradora de momentos de S_R^* sob H_0 e até ordem n^{-1} é

$$M_{S_R^*}(t) = \int_0^\infty \exp\{[x - (cx + bx^2 + ax^3)]t\} f_{S_R}(x) dx.$$

Usando (2.12) e após expandir $\exp\{-(cx + bx^2 + ax^3)t\}$ em série de Taylor obtemos até ordem n^{-1}

$$\begin{aligned} M_{S_R^*}(t) &= \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} \int_0^\infty \exp\left\{\left(t - \frac{1}{2}\right)x\right\} x^{\frac{r}{2}-1} \\ &\quad \times \{(B_3 - at)x^3 + (B_2 - bt)x^2 + (B_1 - ct)x + 1 + B_0\} dx. \end{aligned}$$

A transformação para uma nova variável $y = x(1 - 2t)$ conduz a

$$M_{S_R^*}(t) = M_r(t) + (1 - 2t)^{-\frac{r}{2}} I,$$

onde $M_r(\cdot)$ denota a função geradora de momentos da distribuição χ_r^2 e

$$I = \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} \int_0^\infty \exp\left(-\frac{y}{2}\right) y^{\frac{r}{2}-1} \left\{ \frac{B_3 - at}{(1 - 2t)^3} y^3 + \frac{B_2 - bt}{(1 - 2t)^2} y^2 + \frac{B_1 - ct}{1 - 2t} y + B_0 \right\} dy$$

é $O(n^{-1})$. Cada parcela de I pode ser facilmente obtida através dos momentos de distribuições χ^2 com graus de liberdade adequados. Assim,

$$I = B_0 + \frac{r(B_1 - ct)}{1 - 2t} + \frac{r(r+2)(B_2 - bt)}{(1 - 2t)^2} + \frac{r(r+2)(r+4)(B_3 - at)}{(1 - 2t)^3}.$$

A relação $M_{S_R^*}(t) = M_r(t)$ vale até ordem n^{-1} se e somente se $I = 0$, isto é,

$$B_0(1 - 2t)^3 + r(B_1 - ct)(1 - 2t)^2 + r(r+2)(B_2 - bt)(1 - 2t) + r(r+2)(r+4)(B_3 - at) = 0.$$

Então, a , b e c devem satisfazer

$$B_0 + rB_1 + r(r+2)B_2 + r(r+2)(r+4)B_3 = 0,$$

$$-6B_0 - rc - 4rB_1 - r(r+2)b - 2r(r+2)B_2 - r(r+2)(r+4)a = 0,$$

$$12B_0 + 4rc + 4rB_1 + 2r(r+2)b = 0,$$

$$-8B_0 - 4rc = 0.$$

A primeira equação é diretamente satisfeita pelos B 's e a única solução, dada em função de A_1 , A_2 e A_3 , para as três últimas equações é

$$\begin{aligned} a &= \frac{A_3}{12r(r+2)(r+4)}, \\ b &= \frac{A_2 - 2A_3}{12r(r+2)}, \\ c &= \frac{A_1 - A_2 + A_3}{12r}. \end{aligned} \quad (2.14)$$

Portanto, de (2.13) e (2.14), conclui-se que uma estatística escore modificada tendo distribuição χ^2 até ordem n^{-1} é

$$S_R^* = S_R\{1 - b(S_R, A_1, A_2, A_3, r)\}, \quad (2.15)$$

onde

$$b(x, A_1, A_2, A_3, r) = \frac{A_1 - A_2 + A_3}{12r} + \frac{A_2 - 2A_3}{12r(r+2)}x + \frac{A_3}{12r(r+2)(r+4)}x^2. \quad (2.16)$$

O fator $1 - b(S_r, A_1, A_2, A_3, r)$ é um tipo de correção de Bartlett envolvendo cumulantes de derivadas da log-verossimilhança e um polinômio de segundo grau na própria estatística S_R . Uma vez que os A 's possam ser calculados pelas fórmulas (2.6)-(2.8), um teste escore aperfeiçoado pode ser conduzido utilizando-se a estatística S_R^* e a distribuição χ_r^2 como referência.

Uma maneira alternativa de determinar a , b e c é através da expansão (1) do Teorema 1 de Cox e Reid (1987a; vide Apêndice 1). Notando que $S_R = O_p(1)$ o uso desta expansão conduz a

$$F_{S_R^*} = F_{S_R}(x) + f_{S_R}(x)(cx + bx^2 + ax^3)$$

até ordem n^{-1} . Substituindo f_{S_R} dada em (2.12) na expressão acima tem-se até esta ordem

$$F_{S_R^*}(x) = F_{S_R}(x) + g_r(x)(cx + bx^2 + ax^3). \quad (2.17)$$

Mas, das relações de recorrência

$$G_{m+2}(x) = G_m(x) - \frac{2x}{m}g_m(x), \quad g_{m+2}(x) = \frac{x}{m}g_m(x), \quad (2.18)$$

podemos reescrever (2.5) como

$$F_{S_R}(x) = G_r(x) - g_r(x)(B'_1x + B'_2x^2 + B'_3x^3),$$

onde $B'_1 = (A_1 - A_2 + A_3)/(12r)$, $B'_2 = (A_2 - 2A_3)/\{12r(r+2)\}$ e $B'_3 = A_3/\{12r(r+2)(r+4)\}$. Agora, substituindo F_{S_R} em (2.17) obtemos até ordem n^{-1}

$$F_{S_R^*}(x) = G_r(x) + g_r(x)\{(c - B'_1)x + (b - B'_2)x^2 + (a - B'_3)x^3\}.$$

Assim, $F_{S_R^*}(x) = G_r(x)$ até esta ordem se e somente se $a = B'_3$, $b = B'_2$ e $c = B'_1$, que coincide com o resultado obtido em (2.14).

Chandra e Mukerjee (1991) encontraram uma estatística escore aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} para testar hipóteses nulas simples. Não trataram, no entanto, de situações mais gerais como, por exemplo, aquela em que a hipótese nula é composta e há mais do que um parâmetro de perturbação. No final do artigo, eles comentam:

“In this paper, we have considered simple null hypotheses. In the presence of nuisance parameter(s) the position is as follows: Let θ be the parameter of interest and m be the nuisance parameter. If θ be one-dimensional then combining the methods in this paper with those in Mukerjee [23] it should be possible to derive appropriate modifications for Rao’s statistic ⁽¹⁾. The problem, however, becomes much more complex for multidimensional θ - in particular, if θ and m are both multidimensional then in general one cannot employ parametric orthogonality (Cox and Reid [17]) and tensor methods (McCullagh[19, 20]) should be useful. These aspects deserve further attention” ⁽²⁾.

É importante ressaltar que nossos resultados são bem abrangentes no sentido de que são válidos para hipóteses simples e compostas, para parâmetros uni e multidimensionais, ortogonais ou não-ortogonais, de forma que os resultados de Chandra e Mukerjee foram aqui generalizados.

Em geral, S_R^* é uma função de parâmetros desconhecidos. Neste caso, pode ser usada a estatística $\tilde{S}_R$ dada por (2.15) com os parâmetros substituídos por seus respectivos estimadores de máxima verossimilhança sob H_0 . É fácil mostrar que a diferença $\tilde{S}_R - S_R^*$ é $O_p(n^{-3/2})$. Da equivalência das expansões (A1.5) e (A1.6) (vide Apêndice 1) segue que, sob H_0 e até ordem n^{-1} , as distribuições de $\tilde{S}_R$ e S_R^* coincidem. Portanto, até esta ordem, $\tilde{S}_R$ também tem distribuição χ_r^2 .

A forma da estatística escore modificada em (2.15) está fortemente relacionada com a expansão assintótica dos quantis de S_R sob H_0 . Ao reescrever (2.11) de forma conveniente vemos que, desprezando termos de ordem inferior a n^{-1} , $P(S_R \leq z_\alpha) = G_r(x_\alpha) = 1 - \alpha$ com

$$z_\alpha = x_\alpha \{1 + b(x_\alpha, A_1, A_2, A_3, r)\}. \quad (2.19)$$

Comparando as formas funcionais dos fatores de correção de S_R e de seus quantis vemos

(1) Os autores denominam a estatística escore de “Rao’s statistic” pois esta foi originalmente proposta por Rao (1948).

(2) As referências citadas são: Mukerjee (1990a), Cox e Reid (1987b) e McCullagh (1984, 1987).

que existe uma estreita conexão entre a estatística modificada S_R^* e a expansão assintótica dos quantis de S_R .

Demonstramos então que, desprezando termos de ordem inferior a n^{-1} , $F_{S_R^*}(x) = F_{S_R}(z) = G_r(x)$ com S_R^* e z dados em (2.15) e (2.19) respectivamente. Este resultado pode ser obtido diretamente das expansões (1) e (2) do Teorema 1 de Cox e Reid (1987a; vide Apêndice 1), com a grande vantagem de não ser necessário o uso da fórmula de inversão de Hill e Davis (1968). Da expansão (2) de Cox e Reid temos

$$F_{S_R^*}(x) = F_{S_R}(x(1 + c + bx + ax^2))$$

ignorando termos de ordem inferior a n^{-1} . Esta igualdade é válida para quaisquer a , b e c de ordem n^{-1} . No entanto, a expressão no lado direito da igualdade só é igual a $G_r(x)$ se a , b e c são dados por (2.14) como já tínhamos demonstrado através do uso da expansão (1) de Cox e Reid.

No Capítulo 6, mostramos que, em geral, as estatísticas que têm, segundo a hipótese nula, distribuição assintótica χ^2 podem ser aperfeiçoadas por fatores de correção envolvendo um polinômio na própria estatística e encontramos expansões assintóticas para seus quantis. A demonstração é novamente baseada no uso conjunto das expansões (1) e (2) de Cox e Reid.

A estatística aperfeiçoada em (2.15) pode ser definida por $S'_R = S_R\{1 + b(S_R, A_1, A_2, A_3, r)\}^{-1}$ pois o resultado $F_{S'_R}(x) = F_{S_R}(z) = G_r(x)$ também é válido até ordem n^{-1} . A demonstração deste fato baseia-se na expansão da função geradora de momentos de S'_R sob H_0 .

Devemos ressaltar que, analogamente ao que ocorre para a estatística da razão de verossimilhança, a expansão (2.5) para a função distribuição de S_R não é válida se a distribuição dos dados for discreta. Desta forma, não se pode garantir que, nestes casos, a multiplicação da estatística score pelo fator de correção venha a melhorar a aproximação de sua distribuição por qui-quadrado. No entanto, nos capítulos seguintes encontramos

os fatores de correção para a estatística escore tanto para situações em que os dados têm distribuição contínua quanto discreta. Em trabalho futuro pretendemos desenvolver estudos de simulação para avaliar o desempenho dos testes modificadas no caso discreto.

2.5. O CASO UNIPARAMÉTRICO

$$\ell = n^{-1}L \quad L = n\ell$$

Nesta seção fazemos uma aplicação dos resultados da Seção 2.4 em um modelo uniparamétrico. O interesse é testar a hipótese nula simples $H_0 : \beta = \beta^{(0)}$ contra a alternativa composta $H : \beta \neq \beta^{(0)}$. De (2.6)-(2.8) é fácil ver que $A_1 = 0$, $A_2 = 3\kappa_4/\kappa_2^2$ e $A_3 = 5\kappa_3^2/\kappa_2^3$, onde $\kappa_2 = E\{[\partial \ell(\beta)/\partial \beta]^2\}$ é a informação total de Fisher para β , $\kappa_3 = E\{[\partial \ell(\beta)/\partial \beta]^3\}$ e $\kappa_4 = E\{[\partial \ell(\beta)/\partial \beta]^4\} - 3\kappa_2^2$ são o terceiro e o quarto cumulantes da função escore total $U(\beta)$, respectivamente. As quantidades A_2 e A_3 podem ser expressas em função de cumulantes padronizados, ou seja, $A_2 = 3\gamma_2$ e $A_3 = 5\gamma_1^2$, onde $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ e $\gamma_2 = \kappa_4/\kappa_2^2$ são as medidas usuais de assimetria e curtose da função escore total. Para funções escore normais, $A_2 = A_3 = 0$.

De (2.15), a estatística escore aperfeiçoada é

$$S_R^* = S_R \left\{ 1 - \frac{1}{36} \left[3(5\gamma_1^2 - 3\gamma_2) + (3\gamma_2 - 10\gamma_1^2)S_R + \gamma_1^2 S_R^2 \right] \right\}. \quad (2.20)$$

O primeiro coeficiente em (2.20), isto é, $(5\gamma_1^2 - 3\gamma_2)/12$, coincide com o primeiro termo da fórmula (13) de McCullagh e Cox (1986) para o fator de correção da estatística razão de verossimilhança e mede a não-normalidade ou não-normalidade inversa da função escore. Este é também o primeiro termo de ordem n^{-1} da expansão “saddlepoint” da densidade da média amostral. O terceiro coeficiente está relacionado com a assimetria da função escore e o segundo é uma combinação linear dos demais.

Exemplo 2.1

Sejam $y_1, \dots, y_n$ uma amostra aleatória de uma distribuição exponencial com média μ . De (2.2), temos que a estatística escore para testar $H_0 : \mu = \mu^{(0)}$ contra

$H : \mu \neq \mu^{(0)}$ é

$$S_R = n \frac{(\bar{y} - \mu^{(0)})^2}{\mu^{(0)2}}.$$

As propriedades da distribuição exponencial conduzem a $\kappa_2 = n/\mu^2$, $\kappa_3 = 2n/\mu^3$ e $\kappa_4 = 6n/\mu^4$ de onde segue que $A_2 = 18/n$ e $A_3 = 20/n$. Aqui $A_1 = 0$ pois a hipótese nula é simples. Uma maneira alternativa de se obter os A 's é através do cálculo direto dos três primeiros momentos de S_R e do uso das fórmulas em (2.10). Notando que $n\bar{y}$ tem distribuição gama com média $n\mu$ e índice igual a 1, obtêm-se $\mu'_1(S_R) = 1$, $\mu'_2(S_R) = 3 + 6/n$ e $\mu'_3(S_R) = 15 + 130/n$ até ordem n^{-1} . Substituindo em (2.10), chega-se aos A 's obtidos anteriormente. De (2.20) vem a estatística escore aperfeiçoada

$$S_R^* = S_R \left\{ 1 - \frac{1}{18n} (3 - 11S_R + 2S_R^2) \right\}.$$

2.6. UM COMENTÁRIO SOBRE O PODER DOS TESTES ESCORE A-PERFEIÇADOS

Harris e Peers (1980), seguindo o trabalho de Hayakawa (1975), desenvolveram uma expansão assintótica para o poder do teste escore para uma série de hipóteses alternativas de Pitman convergindo para a hipótese nula à taxa $n^{-1/2}$. Mais especificamente, consideram a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$, como definimos na Seção 2.1, e a série de alternativas contíguas $H : \beta_1 = \beta_1^{(0)} + \xi$, onde $\xi = (\xi_1, \dots, \xi_r)^T$ com $\xi_i = \beta_i - \beta_i^{(0)}$ suposto ser de ordem $n^{-1/2}$. As hipóteses alternativas locais não têm nenhum significado físico, mas refletem o interesse no estudo do poder de testes para valores do parâmetro próximos daquele estabelecido sob a hipótese nula, ou seja, na região em que o poder não é grande (Cox e Hinkley, 1974; pg 318). Eles mostraram que, sob certas condições de regularidade, a função de poder do teste escore pode ser expressa, localmente e até ordem $n^{-1/2}$, como $1 - P(S_R \leq x_\alpha)$, onde

$$P(S_R \leq x_\alpha) = G_{r,\lambda}(x_\alpha) + \sum_{l=0}^3 a_l G_{r+2l,\lambda}(x_\alpha), \quad (2.21)$$

sendo $G_{r,\lambda}(x) = P(\chi_{r,\lambda}^2 \leq x)$ com $\chi_{r,\lambda}^2$ representando uma variável aleatória qui-quadrado não-central com r graus de liberdade e parâmetro de não-centralidade λ . Aqui, x_α é tal que $G_{r,0}(x_\alpha) = 1 - \alpha$ e $\lambda = \delta^T K \delta$ com

$$\delta = \begin{pmatrix} I_r \\ -K_{22}^{-1} K_{21} \end{pmatrix} \xi \quad (2.22)$$

sendo que I_r denota uma matriz de dimensão $r \times r$ e as matrizes K , K_{22} e K_{21} estão definidas em (2.1). As quantidades $a_0, \dots, a_3$ são de ordem $n^{-1/2}$ e são dadas por

$$a_1 = \frac{1}{6} \left\{ \sum_{i,j,k=1}^p (\kappa_{ijk} - 2\kappa_{i,j,k}) \delta_i \delta_j \delta_k - 3 \sum_{i=1}^r \sum_{j,k=1}^p (\kappa_{ijk} + \kappa_{i,jk}) \xi_i \delta_j \delta_k \right. \\ \left. - 3 \sum_{i,j,k=1}^p \kappa_{i,j,k} m_{ij} \delta_k + 3 \sum_{i,j,k=1}^p (\kappa_{ijk} + 2\kappa_{i,jk}) a_{ij} \delta_k \right\}, \quad (2.23)$$

$$a_2 = \frac{1}{2} \sum_{i,j,k=1}^p \kappa_{i,j,k} m_{ij} \delta_k, \quad (2.24)$$

$$a_3 = \frac{1}{6} \sum_{i,j,k=1}^p \kappa_{i,j,k} \delta_i \delta_j \delta_k \quad (2.25)$$

e

$$a_0 = -(a_1 + a_2 + a_3). \quad (2.26)$$

Aqui δ_i é o i -ésimo elemento do vetor δ e a_{ij} e m_{ij} representam os elementos (i, j) das matrizes A e M dadas em (2.9) respectivamente. Nas expressões (2.21)–(2.26) todas as quantidades, exceto ξ , são avaliadas sob H_0 . De (2.21) deduz-se que os três primeiros cumulantes de S_R sob as alternativas contíguas e até ordem $n^{-1/2}$ são dados por

$$\begin{aligned} \mu'_{1,\lambda} &= r + \lambda + 2(a_1 + 2a_2 + 3a_3), \\ \mu_{2,\lambda} &= 2(r + 2\lambda) + 8(a_1 + 3a_2 + 6a_3), \\ \mu_{3,\lambda} &= 8(r + 3\lambda) + 48(a_1 + 4a_2 + 10a_3). \end{aligned} \quad (2.27)$$

As expansões para a função distribuição das estatísticas da razão de verossimilhança e de Wald sob alternativas contíguas têm formas iguais à da expressão (2.21) com

os a 's dados no artigo de Hayakawa (1975). Uma conclusão imediata deste fato é a de que os testes escore, da razão de verossimilhança e de Wald têm o mesmo poder até primeira ordem. Entretanto, comparando as parcelas de ordem $n^{-1/2}$ das três expansões, Harris e Peers (1980) notaram que os poderes são diferentes até esta ordem, mas que não é possível, em geral, afirmar que um teste é mais poderoso que os demais. Em algumas situações particulares, eles observaram que dois entre os três testes têm o mesmo poder.

Dos resultados da Seção 2.4, o teste baseado na estatística escore modificada S_R^* é preferível ao teste escore usual, baseado em S_R , no sentido de que, sob H_0 , a distribuição de S_R^* converge mais rapidamente para a distribuição χ^2 de referência do que a de S_R . Indo mais além, é interessante comparar os poderes dos dois testes. A comparação de poderes até ordem $n^{-1/2}$ faz sentido uma vez que os testes apresentam probabilidades de erro de tipo I iguais até esta ordem.

Uma consequência imediata da expansão (1) de Cox e Reid (1987a; vide Apêndice 1) é o fato de que os testes baseados em S_R e S_R^* têm poderes locais idênticos até ordem $n^{-1/2}$. Em outras palavras, a modificação na estatística não tem efeito sobre o poder do teste até esta ordem. Ainda da expansão (1) é possível concluir que, até ordem n^{-1} ,

$$P(S_R^* > x_\alpha) - P(S_R > x_\alpha) = -g_{r,\lambda}(x_\alpha)x_\alpha(c + bx_\alpha + ax_\alpha^2)$$

onde a , b e c são dados em (2.14) e $g_{r,\lambda}(\cdot)$ é a função densidade de uma distribuição χ^2 não-central com r graus de liberdade e parâmetro de não-centralidade λ . Daí se conclui que o teste escore modificado é mais poderoso que o usual para um valor fixado para α se e somente se $c + bx_\alpha + ax_\alpha^2 < 0$. No Capítulo 4 (Seção 4.4), mostraremos que $a \geq 0$, porém, isto pouco auxilia no estudo do sinal de $c + bx_\alpha + ax_\alpha^2$ pois a , b e c são, em geral, funções paramétricas complicadas. Somente em casos particulares é possível determinar qual teste é localmente mais poderoso para um determinado nível de significância nominal fixado.

Finalmente, é importante notar que, até ordem n^{-1} , o teste escore modificado, baseado na expansão em (2.19) para os quantis de S_R , tem o mesmo poder local que o

teste que se baseia na estatística S_R^* . Isto é consequência da expansão (2) de Cox e Reid (1987a; vide Apêndice 1) que implica que, desprezando termos de ordem inferior a n^{-1} , $P(S_R^* \leq x_\alpha) = P(S_R \leq z_\alpha)$ com z_α dado em (2.19).

CAPÍTULO 3

APERFEIÇOAMENTO DE TESTES ESCORE EM MODELOS LINEARES GENERALIZADOS

3.1. INTRODUÇÃO

Os modelos lineares generalizados (MLGs), definidos por Nelder e Wedderburn (1972), têm desempenhado um papel importante na análise de dados, pois englobam muitos métodos estatísticos e dispõem de um algoritmo único de ajuste de modelos. Entre os métodos abrangidos pelos MLGs destacam-se: (i) os modelos clássicos de regressão linear com erro normal; (ii) os modelos log-lineares para dados em forma de contagens dispostos em tabelas de contingência; (iii) os modelos logísticos lineares e modelos probito para análise de proporções. Estudos abrangentes sobre MLGs são encontrados nos livros de Cordeiro (1986b) e McCullagh e Nelder (1989).

O sistema GLIM (Payne, 1986) foi desenvolvido para a análise de dados via os MLGs e pode ser ainda utilizado para ajuste de outros modelos. O método usual de testes

de hipóteses em MLGs é baseado na razão de verossimilhança. No entanto, Pregibon (1982) mostra que os testes *score* podem facilmente ser computados no GLIM como a diferença entre as somas de quadrados de resíduos de dois modelos de regressão linear ponderada.

Os MLGs são definidos por um conjunto de variáveis aleatórias independentes tendo distribuição da família exponencial e uma estrutura linear nos parâmetros relacionada à média da variável resposta através de uma função de ligação. Estes modelos podem ser estendidos em algumas direções. Por exemplo, pode-se estabelecer uma família mais ampla de distribuições para os dados ou permitir uma estrutura não-linear ou parcialmente linear nos parâmetros.

Ultimamente, encontram-se na literatura vários artigos sobre a utilização de testes *score* em MLGs e suas extensões, em especial, em testes de diagnóstico. Davison e Tsai (1990) fazem uma revisão bibliográfica sobre testes de diagnóstico em MLGs e estendem os testes a uma classe mais ampla de modelos. Várias formas de falta de ajuste são colocadas em um mesmo contexto e os autores mostram que podem ser testadas com base na estatística *score*. Em geral, os testes *score* de diagnóstico são utilizados em conjunto com métodos gráficos. Neste sentido, destacam-se os artigos de Cook e Weisberg (1983) sobre testes de heterocedasticidade no modelo usual de regressão e o trabalho de Wang (1985) sobre o teste de significância de uma variável auxiliar. ^{em MLGs} Métodos de diagnóstico e testes de transformação de ^{ódoos} (variáveis) baseados na estatística *score* são encontrados ainda nos trabalhos de Atkinson (1982), Hinkley (1985, 1988), Lawrance (1987), Atkinson e Lawrance (1989), Tsai (1988) e outras referências citadas nestes artigos.

Aspectos de robustez do teste *score* são abordados por Chen (1983). O autor mostra que o teste da hipótese nula de média constante em MLGs é robusto no sentido de que não depende da forma funcional escolhida para a função de ligação. Em seu artigo de 1985, Chen demonstra uma propriedade semelhante em modelos de regressão parcialmente não-lineares.

Tosteson e Tsiats (1988) e Stefansky e Carrol (1990) estudam alguns testes es-

core em MLGs com erro de medida na variável auxiliar. Breslow (1989) discute duas versões do teste ^{sob as hipóteses}escore em MLGs com superdispersão. Day e Byar (1979) mostram que o teste de Mantel-Haenzel é equivalente a um teste ⁵escore em um modelo logístico. Um estudo de simulação sobre testes ⁵escore em modelos logísticos é encontrado no artigo de Brown (1982). Gart e Tarone (1983) obtiveram o mesmo resultado de Day e Byar de uma maneira mais simples utilizando uma verossimilhança condicional. Eles mostram também que, em modelos da família exponencial com um parâmetro sendo testado na presença de parâmetros de perturbação, o teste ⁵escore é a aproximação normal para o teste não viciado uniformemente mais poderoso. Várias referências sobre aplicações de testes ⁵escore em MLGs especiais são citadas no artigo de Gart e Tarone (1983). Referências mais recentes incluem os trabalhos de Gart (1985), Liang e Self (1985), Tarone (1986, 1988) e Tarone e Gart (1989).

Os testes ⁵escore aperfeiçoados descritos no Capítulo 2 envolvem as quantidades A_1 , A_2 e A_3 que são funções de cumulantes conjuntos de derivadas da log-verossimilhança. Dependendo do modelo associado aos dados, pode ser bastante difícil obter tais quantidades. Sendo assim, é importante expressar os A 's em forma fechada para diversas classes de modelos. Neste capítulo, exibimos os A 's em notação matricial para testes ⁵escore em MLGs e mostramos como podem ser computados no sistema GLIM. Extensões de ⁵estes resultados a modelos não-lineares da família exponencial serão feitas no Capítulo 5.

Na Seção 3.2, definimos os MLGs e apresentamos a estatística ⁵escore para testes sobre os parâmetros da regressão. Nas Seções 3.3 a 3.5 desenvolvemos expressões matriciais de fácil computação para as quantidades A_1 , A_2 e A_3 e mostramos como obtê-las através do GLIM. Na Seção 3.6, estudamos alguns casos especiais e nas Seções 3.7 e 3.8, fazemos várias aplicações e apresentamos dois estudos de simulação. Finalmente, na Seção 3.9, expressamos o poder dos testes ⁵escore aperfeiçoados, até ordem $n^{-1/2}$, em notação matricial.

3.2. TESTES ESCORE EM MODELOS LINEARES GENERALIZADOS

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$ sendo que cada y_l tem função densidade (ou de probabilidade) na família exponencial da forma

$$\pi(y; \theta_l, \phi) = \exp[\phi\{y\theta_l - b(\theta_l) + c(y)\} + a(y, \phi)], \quad (3.1)$$

onde $a(\cdot, \cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas. Assumimos que ϕ ($\phi > 0$) é conhecido e chamamos ϕ^{-1} de parâmetro de dispersão. Para o modelo em (3.1) valem as relações: $E(y_l) = \mu_l = b'(\theta_l)$ e $\text{Var}(y_l) = V_l/\phi$, onde $V = V(\mu) = d\mu/d\theta$ é chamada função de variância e $\theta = \int V^{-1}d\mu = q(\mu)$ sendo $q(\mu)$ uma função bijunívoca. Supomos que as variáveis auxiliares $x_1, \dots, x_p$, onde $x_j = (x_{1j}, \dots, x_{nj})^T$, produzem uma estrutura linear dada por $\eta = X\beta$ sendo $X = (x_1, \dots, x_p)$ a matriz modelo $n \times p$ ($p < n$) conhecida de posto completo e $\beta = (\beta_1, \dots, \beta_p)^T$ um conjunto de parâmetros desconhecidos.

Um MLG é definido por (3.1) e por uma função de ligação $d(\mu_l) = \eta_l$, $l = 1, \dots, n$, relacionando a média μ_l de uma observação com o preditor linear η_l do modelo. Assumimos que $d(\cdot)$ é uma função biunívoca.

As distribuições normal, Poisson, binomial, gama e normal inversa são casos especiais de (3.1). Diversas características destas distribuições são apresentadas na Tabela 3.1.

Como no Capítulo 2, consideramos a partição $\beta = (\beta_1^T, \beta_2^T)^T$ com $\beta_1 = (\beta_1, \dots, \beta_r)^T$ e $\beta_2 = (\beta_{r+1}, \dots, \beta_p)^T$. A matriz modelo particionada de maneira análoga é $X = (X_1 \ X_2)$. O interesse é testar $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$, onde $\beta_1^{(0)}$ é um vetor especificado de dimensão r .

A função escore total para β pode ser escrita como

$$U(\beta) = \phi X^T W^{1/2} V^{-1/2} (y - \mu)$$

e a matriz de informação total de Fisher é $\frac{1}{\phi}$

$$K = \phi X^T W X,$$

$$K^{-1}$$

Tabela 3.1. Características de algumas distribuições da família exponencial

	Normal $N(\mu, \phi^{-1})$	Poisson $P(\mu)$	Binomial $B(n, \mu)$	Gama $G(\mu, \phi)$	Normal inversa $N^{-1}(\mu, \phi)$
Suporte de y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)n}{n}$	$(0, \infty)$	$(0, \infty)$
$\phi c(y) + a(y, \phi)$	$-\frac{1}{2} \left(\phi y^2 + \log \frac{2\pi}{\phi} \right)$	$-\log y!$	$\log \binom{n}{ny}$	$(\phi - 1) \log(y\phi) + \log \phi - \log \Gamma(\phi)$	$\frac{1}{2} \left(\log \frac{\phi}{2\pi y^3} - \frac{\phi}{y} \right)$
$b(\theta)$	$\frac{\theta^2}{2}$	e^θ	$\log(1 + e^\theta)$	$(\phi - 1) \log y + \log \phi$ $-\log(-\theta)$	$-(-2\theta)^{1/2}$
$\mu = E(y)$	θ	e^θ	$\frac{e^\theta}{1 + e^\theta}$	$-\frac{1}{\theta}$	$(-2\theta)^{-1/2}$
$V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

Nota: No modelo binomial, a variável aleatória corresponde à proporção de sucessos em n ensaios de Bernoulli e $\phi = n$.

A parametrização do modelo gama é tal que a sua variância é μ^2/ϕ .

Fonte: McCullagh e Nelder (1989; Tabela 2.1, pg.30).

onde $V = \text{diag}\{V_1, \dots, V_n\}$, $W = \text{diag}\{w_1, \dots, w_n\}$ com $w_l = (d\mu_l/d\eta_l)^2/V_l$. A notação D^x , sendo D uma matriz diagonal e x um número real, é usada para indicar que cada elemento da diagonal de D está elevado à potência x . A matriz K é particionada segundo a partição de β (vide (2.1)) nas sub-matrizes $K_{11} = \phi X_1^T W X_1$, $K_{22} = \phi X_2^T W X_2$ e $K_{12} = K_{21}^T = \phi X_1^T W X_2$.

A estatística S_R (vide (2.2)) para testar H_0 versus H depende da matriz K^{-1} (vide (2.1)) que é obtida da inversa de K . Utilizando a fórmula de inversa de matrizes simétricas particionadas (Rao, 1973, pg. 33) e após alguma álgebra chega-se a

$$K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}^{-1} \quad (3.2)$$

com $K^{11} = \phi^{-1}(R^T W R)^{-1}$, $K^{12} = K^{21T} = -\phi^{-1}(R^T W R)^{-1} C^T$ e $K^{22} = \phi^{-1}\{(X_2^T W X_2)^{-1} + C(R^T W R)^{-1} C^T\}$, onde $R = X_1 - X_2 C$ com $C = (X_2^T W X_2)^{-1} X_2^T W X_1$ representando uma matriz $(p-r) \times r$ cujas colunas são vetores dos coeficientes da regressão normal linear das colunas de X_1 sobre a matriz modelo X_2 com matriz de pesos W . Assim, obtém-se

$$S_R = \tilde{S}^T \tilde{W}^{1/2} X_1 (\tilde{R}^T \tilde{W} \tilde{R})^{-1} X_1^T \tilde{W}^{1/2} \tilde{S}, \quad (3.3)$$

onde $S = (s_1, \dots, s_n)^T$, com $s_l = \phi^{1/2}(y_l - \mu_l)V_l^{-1/2}$, representa o vetor de resíduos de Pearson. Como no Capítulo 2, o til é utilizado para indicar que as funções de β são avaliadas no ponto $\tilde{\beta} = (\beta_1^{(0)T}, \tilde{\beta}_2^T)^T$ sendo $\tilde{\beta}_2$ o estimador de máxima verossimilhança de β_2 sob H_0 . O valor de $\tilde{\beta}_2$ é obtido diretamente no sistema GLIM das equações $X_2^T \tilde{W} X_2 \tilde{\beta}_2 = X_2^T \tilde{W} \tilde{y}^*$ onde $\tilde{y}^* = \tilde{\eta} + \tilde{W}^{1/2} \tilde{V}^{-1/2}(y - \tilde{\mu})$, utilizando o diretivo \$FIT X_2 e definindo o vetor $X_1 \beta_1^{(0)}$ como um offset (Payne, 1986).

Pregibon (1982) fornece uma maneira muito simples de calcular (3.3) no sistema GLIM como a diferença entre as somas de quadrados de resíduos de dois modelos de regressão linear ponderada. O cálculo é feito da seguinte forma:

1. Ajusta-se o modelo reduzido usando $\tau = X_1 \beta_1^{(0)}$ como offset (\$OFF τ seguido de \$FIT X_2) e calcula-se a estatística X^2 de Pearson (\$CAL %C=%X2).

2. Procede-se a primeira iteração do ajuste do modelo completo (\$REC 1 seguido de \$FIT +X₁) e calcula-se a diferença entre as estatísticas X^2 dos dois ajustes (\$CAL %C- %X2). Este é o valor da estatística escore (3.3). Pregibon observa que o valor fornecido pelo GLIM para a estatística de Pearson do modelo reduzido é baseado na estimativa de β_2 da penúltima iteração. Para calcular seu valor exato basta computar $\%C = \%CU(\%WT * (\%WV - \%LP)** 2)$.

Como no ajuste do modelo completo utiliza-se somente a primeira iteração, nota-se que a estatística escore tem vantagens computacionais em relação à da razão de verossimilhança já que esta envolve o ajuste dos dois modelos, em geral, em mais do que uma iteração. Num estudo particular de seleção "stepwise" de variáveis, Pregibon (1982) mostra a considerável redução de custo computacional gerada pela utilização da estatística escore em lugar da razão de verossimilhança. Além disto, nos exemplos examinados, ambas as estatísticas apresentam valores similares. Pregibon acaba por concluir: "For large data problems, the score test is the way to go".

3.3. TESTES ESCORE APERFEIÇADOS

Os testes escore aperfeiçoados apresentados no Capítulo 2 dependem das quantidades A_1 , A_2 e A_3 (expressões (2.6)–(2.8)) que são funções aparentemente complicadas dos cumulantes conjuntos κ 's de derivadas da log-verossimilhança. O objetivo desta seção é apresentar A_1 , A_2 e A_3 em forma matricial e de fácil computação para o teste discutido na seção anterior. Alguns destes cumulantes nos MLGs foram obtidos por Cordeiro (1983) que apresentou o fator de correção de Bartlett para a estatística da razão de verossimilhança em notação matricial. Uma propriedade fundamental dos MLGs é que todos os cumulantes envolvidos nos A 's são invariantes sob permutação de parâmetros. Isto conduz a uma simplificação considerável na obtenção dos A 's e, conseqüentemente, na da estatística escore aperfeiçoada. Outras classes de modelos mais amplas como a dos modelos de dispersão

(Jørgensen, 1987) também gozam desta propriedade.

Sejam $\sum'$ e $\sum$ os somatórios sobre todos os β 's e sobre os dados, respectivamente. Usamos a notação $\partial\eta_l/\partial\beta_r = x_{lr} = (r)_l$, $(\partial\eta_l/\partial\beta_r)(\partial\eta_l/\partial\beta_s) = x_{lr}x_{ls} = (rs)_l$, etc. Para os MLGs têm-se os seguintes cumulantes (Apêndice 2):

$$\begin{aligned}\kappa_{rst} &= -\phi \sum (f + 2g)_l (rst)_l, & \kappa_{r,st} &= \phi \sum g_l (rst)_l, \\ \kappa_{r,s,t} &= \phi \sum (f - g)_l (rst)_l, & \kappa_{r,s,t,u} &= \phi \sum b_l (rstu)_l \quad \text{e} \\ \kappa_{r,s,tu} &= \phi \sum \left\{ \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} - \frac{1}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4 \right\}_l (rstu)_l,\end{aligned}$$

onde f_l , g_l e b_l são escalares definidos por

$$f = \left(\frac{1}{V} \right) \left(\frac{d\mu}{d\eta} \right) \left(\frac{d^2\mu}{d\eta^2} \right)^2 \quad V, \quad \frac{d\mu}{d\eta} \quad (3.4)$$

$$g = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3 \quad 3 \quad (3.5)$$

e

$$b = \frac{1}{V^3} \left(\frac{d\mu}{d\eta} \right)^4 \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2V}{d\mu^2} \right\}. \quad (3.6)$$

Mostra-se facilmente que $\kappa_{r,s,t,u} + \kappa_{r,s,tu} = \phi \sum h_l (rstu)_l$, onde o escalar h_l é dado por

$$h = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} + \frac{1}{V^2} \frac{d^2V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4. \quad (3.7)$$

Definimos

$$Z = \{z_{lm}\} = X(X^T W X)^{-1} X^T, \quad Z_2 = \{z_{2lm}\} = X_2(X_2^T W X_2)^{-1} X_2^T, \quad (3.8)$$

ou seja, $Z = \phi X K^{-1} X^T$ e $Z_2 = \phi X_2 K_{22}^{-1} X_2^T$. As matrizes $\phi^{-1} Z$ e $\phi^{-1} Z_2$ são as estruturas de covariância assintótica de $X\hat{\beta}$ e $X_2\tilde{\beta}_2$ respectivamente.

Expressões simples para os A 's em notação matricial podem ser obtidas substituindo os κ 's nas expressões (2.6)–(2.8) e efetuando as somas sobre a amostra depois de avaliar as somas sobre os parâmetros. Fazendo isto, aparecerão termos da forma

$\sum' a_{ij}x_{li}x_{lj}$ e $\sum' m_{ij}x_{li}x_{mj}$, onde a_{ij} e m_{ij} são os elementos (i, j) das matrizes A e M dadas em (2.9). É fácil ver que estes somatórios representam os elementos de $\phi^{-1}Z_2$ e $\phi^{-1}(Z - Z_2)$ respectivamente.

Consideremos a obtenção da primeira parcela de A_1 , A_{11} digamos, em detalhes.

De $\kappa_{rst} + 2\kappa_{rs,t} = -\phi \sum f_l(rst)_l$, A_{11} reduz-se a

$$A_{11} = 3\phi^2 \sum' \sum f_l(rst)_l \sum f_m(ijk)_m a_{ij}a_{st}m_{kr}.$$

Invertendo a ordem das somas e rearrumando os termos tem-se

$$A_{11} = 3\phi^2 \sum f_l f_m \left(\sum' a_{ij}x_{mi}x_{mj} \right) \left(\sum' a_{st}x_{ls}x_{lt} \right) \left(\sum' m_{kr}x_{mk}x_{lr} \right).$$

Da definição de Z_2 e $Z - Z_2$ esta parcela se resume a

$$A_{11} = 3\phi^{-1} \sum_{l,m=1}^n f_l f_m z_{2mm} z_{2ll} (z_{l\bar{m}} - z_{2lm}).$$

A última expressão, em notação matricial, é dada por

$$A_{11} = 3\phi^{-1} 1^T F Z_{2d} (Z - Z_2) Z_{2d} F 1,$$

onde $Z_{2d} = \text{diag}\{z_{211}, \dots, z_{2nn}\}$ e 1 é um vetor $n \times 1$ de uns. As demais parcelas dos A 's são obtidas de forma análoga (vide Apêndice 2). As expressões resultantes são:

$$\begin{aligned} (A_1) &= \phi^{-1} [3 \ 1^T F Z_{2d} (Z - Z_2) Z_{2d} F 1 + 6 \ 1^T F Z_{2d} Z_2 (Z - Z_2)_d (F - G) 1 \\ &\quad - 6 \ 1^T F \{Z_2^{(2)} \otimes (Z - Z_2)\} (2G - F) 1 - 6 \ 1^T H (Z - Z_2)_d Z_{2d} 1], \end{aligned} \quad (3.9)$$

$$\begin{aligned} (A_2) &= \phi^{-1} [-3 \ 1^T (F - G) (Z - Z_2)_d Z_2 (Z - Z_2)_d (F - G) 1 \\ &\quad - 6 \ 1^T F Z_{2d} (Z - Z_2) (Z - Z_2)_d (F - G) 1 \\ &\quad - 6 \ 1^T (F - G) \{ (Z - Z_2)^{(2)} \otimes Z_2 \} (F - G) 1 + 3 \ 1^T B (Z - Z_2)_d^{(2)} 1] \end{aligned} \quad (3.10)$$

e

$$\begin{aligned} (A_3) &= \phi^{-1} \{ 3 \ 1^T (F - G) (Z - Z_2)_d (Z - Z_2) (Z - Z_2)_d (F - G) 1 \\ &\quad + 2 \ 1^T (F - G) (Z - Z_2)^{(3)} (F - G) 1 \}, \end{aligned} \quad (3.11)$$

onde $F = \text{diag}\{f_1, \dots, f_n\}$, $G = \text{diag}\{g_1, \dots, g_n\}$, $B = \text{diag}\{b_1, \dots, b_n\}$ e $H = \text{diag}\{h_1, \dots, h_n\}$ são matrizes diagonais de dimensão $n \times n$ com elementos definidos em (3.4)-(3.7) respectivamente, $\otimes$ denota um produto direto e usamos a seguinte notação: $Z^{(2)} = Z \otimes Z$, $Z^{(3)} = Z^{(2)} \otimes Z$ e $Z_d = \text{diag}\{z_{11}, \dots, z_{nn}\}$. Note-se que Z_d é uma matriz diagonal com os correspondentes elementos da diagonal de Z .

A estatística escore aperfeiçoada para o teste discutido na Seção 3.2 é obtida de (2.15) com os A 's dados em (3.9)-(3.11). É importante notar que estas quantidades são diretamente proporcionais ao parâmetro de dispersão ϕ^{-1} como era esperado. Em outras palavras, quanto mais dispersos forem os dados maior será a necessidade de se corrigir a estatística. Caso os A 's dependam de parâmetros desconhecidos estes devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob H_0 . Note-se que as expansões assintóticas da função distribuição e dos quantis da estatística escore sob H_0 são obtidas de (2.5) e (2.11) também substituindo os A 's por (3.9)-(3.11). Um teste escore aperfeiçoado pode ser conduzido de duas maneiras: utilizando a estatística escore aperfeiçoada e a distribuição χ_r^2 como referência, ou usando a própria estatística escore em conjunto com as expansões assintóticas citadas acima.

3.4. ALGUNS COMENTÁRIOS SOBRE OS COEFICIENTES

As expressões para A_1 , A_2 e A_3 são funções das matrizes Z_2 , $Z - Z_2$ e das médias desconhecidas e são ponderadas pelo parâmetro de dispersão. Os A 's dependem das funções de ligação e de variância com suas primeiras e segundas derivadas. O termo A_3 também depende das duas primeiras derivadas da função de ligação mas somente da primeira derivada da função de variância. Uma vez que as médias desconhecidas tenham sido estimadas sob H_0 e que as matrizes Z_2 e $Z - Z_2$ tenham sido obtidas, o cálculo de A_1 , A_2 e A_3 é conduzido sem dificuldade a partir das expressões (3.9)-(3.11) que só envolvem operações simples com matrizes e vetores. Note-se que se a hipótese nula é simples, não há necessidade de se fazer

nenhuma estimação. Para MLGs com formas fechadas para as matrizes Z_2 e $Z - Z_2$, é possível obter formas fechadas para as expressões dos A 's. Isto é o que torna as fórmulas (3.9)–(3.11) de grande interesse na prática.

A estimação de máxima verossimilhança, sob H_0 , das médias desconhecidas é conduzida no sistema GLIM ajustando o modelo reduzido (`$FIT X2`) aos dados y com o “offset” $\tau = X_1\beta_1^{(0)}$ (`$OFF  $\tau$`). Isto combinado com o uso de um sistema computadorizado que execute operações com matrizes (incluindo inversão), como é o caso do REDUCE (Rayna, 1987), produzirá a maneira mais fácil de se calcular os A 's. No entanto, estas quantidades podem também ser obtidas no GLIM (vide Seção 3.5) usando macros especiais para multiplicação de matrizes. Numericamente, elas também podem ser obtidas via o pacote estatístico GAUSS ou a linguagem APL.

Embora A_1 , A_2 e A_3 sejam calculados facilmente, eles são de difícil interpretação. A dificuldade fundamental é que os termos individuais nos A 's não são invariantes sob reparametrização e, portanto, não têm interpretação independente do sistema de coordenadas escolhido.

3.5. CÁLCULO DOS COEFICIENTES NO SISTEMA GLIM

Como visto na seção anterior, uma vez que as estimativas $\tilde{\mu}$ das médias desconhecidas tenham sido obtidas, as quantidades A_1 , A_2 e A_3 podem ser calculadas através de um sistema computacional que execute soma, produto e inversão de matrizes. Como estas estimativas são obtidas facilmente no GLIM, é interessante verificar como os A 's podem ser calculados diretamente neste sistema.

Inicialmente, damos uma forma simples para a matriz $Z - Z_2$ que facilita o seu cálculo numérico e que eventualmente pode reduzir a manipulação algébrica na obtenção de expressões em forma fechada para os A 's.

A matriz $Z - Z_2$ envolve as inversas das matrizes $X^T W X$, de dimensão $p \times p$, e

$X_2^T W X_2$, de dimensão $(p - r) \times (p - r)$. Nosso objetivo é escrever $Z - Z_2$ de uma forma conveniente tal que envolva somente a inversa de uma matriz de dimensão $r \times r$. Este resultado é obtido sem dificuldade substituindo a matriz $K^{-1} = \phi^{-1}(X^T W X)^{-1}$ dada em (3.2) na expressão (3.8) de Z . Usando propriedades triviais de matrizes simétricas, chega-se a

$$Z - Z_2 = R(R^T W R)^{-1} R^T, \quad (3.12)$$

onde R está definida na Seção 3.2 e representa a matriz dos resíduos das regressões normais lineares das colunas de X_1 sobre a matriz modelo X_2 com matriz de pesos W . Nota-se que $Z - Z_2$ contém somente a inversa de uma matriz de dimensão $r \times r$. Geralmente, é mais fácil encontrar $Z - Z_2$ através de (3.12) do que das definições das matrizes Z e Z_2 em (3.8). Expressões para os A 's deduzidas de (3.12) são dadas na Seção 3.8.

Os cálculos para obtenção dos A 's no GLIM são agora bastante simplificados. Como já vimos, o ajuste de y sobre a matriz modelo X_2 com o offset $X_1 \beta_1^{(0)}$ produz $\tilde{\mu}$. Os pesos iterativos $\tilde{W}$ e a estrutura de covariância $(X_2^T \tilde{W} X_2)^{-1}$ estarão nos vetores do sistema `%WT` e `%VC`, respectivamente. A matriz de resíduos R é obtida do ajuste de X_1 sobre X_2 com pesos $(\%PW) \tilde{W}$. A inversa de $R^T \tilde{W} R$ estará em `%VC` após qualquer regressão linear normal sobre R usando pesos $\tilde{W}$.

As matrizes $Z - Z_2$ e Z_2 , em (3.12) e (3.8) respectivamente, são construídas com base em macros para multiplicação de matrizes (Channon, 1981). As matrizes $Z_2^{(2)}$, $(Z - Z_2)^{(2)}$, $(Z - Z_2)^{(3)}$, $(Z - Z_2)_d$, os produtos diretos $Z_2^{(2)} \otimes (Z - Z_2)$ e $Z_2^{(2)} \otimes (Z - Z_2)^{(2)}$ e os vetores $F1$, $G1$, $F(Z - Z_2)_d1$ e $G(Z - Z_2)_d1$ são facilmente obtidos pelo uso do diretivo `$CALCULATE`. Finalmente, todos os termos necessários têm formas quadráticas simples que são calculadas com a mesma facilidade através deste diretivo.

3.6. ALGUNS CASOS ESPECIAIS

O objetivo desta seção é examinar alguns casos especiais que induzem simplificações nas fórmulas (3.9)–(3.11).

Para um MLG com função de ligação identidade, isto é, $\mu = \eta$, tem-se $F = 0$ e A_1 se reduz à última parcela em (3.9) com $H = \text{diag}\{V^{-2}d^2V/d\mu^2\}$ embora A_2 e A_3 não tenham redução substancial. Aqui $G = \text{diag}\{-V^{-2}dV/d\mu\}$ e $B = \text{diag}\{V^{-3}[(dV/d\mu)^2 + Vd^2V/d\mu^2]\}$.

Para modelos canônicos, para os quais $\theta = \eta$, $G = 0$ e os A 's pouco se simplificam. Para modelos normais ($V = 1$) com variância comum ϕ^{-1} conhecida, os termos A_2 e A_3 são nulos e

$$A_1 = 3\phi^{-1}1^T F[Z_{2d}(Z - Z_2)Z_{2d} - 2\{Z_2^{(2)} \otimes (Z - Z_2)\}]F1,$$

onde $F = \text{diag}\{(d\mu/d\eta)d^2\mu/d\eta^2\}$. Neste caso, a estatística escore aperfeiçoada é $S_R^* = S_R\{1 - A_1/(12r)\}$. Em particular, se a ligação for a identidade, $F = 0$ e, portanto, $A_1 = 0$ como era esperado já que S_R tem distribuição exata χ_r^2 .

Se o parâmetro β_1 especificado sob a hipótese nula é ortogonal ao parâmetro de perturbação β_2 , $Z - Z_2 = X_1(X_1^T W X_1)^{-1} X_1^T$, embora as expressões para os A 's não sofram grandes simplificações.

No caso de uma hipótese nula simples $H_0 : \beta = \beta^{(0)}$, tem-se $Z_2 = 0$ o que conduz a $A_1 = 0$ enquanto os outros A 's reduzem-se a

$$A_2 = 3\phi^{-1}1^T B Z_d^{(2)} 1$$

e

$$A_3 = \phi^{-1}1^T (F - G)(3Z_d Z_d^{(2)} + 2Z_d^{(3)})(F - G)1.$$

Em particular, para um parâmetro unidimensional, $Z = \{(\sum w_l x_l^2)^{-1} x_i x_j\}$ e a estatística escore se reduz a

$$S_R = \phi \left(\sum w_l^{(0)} x_l^2 \right)^{-1} \left\{ \sum (w_l^{(0)} / V_l^{(0)})^{1/2} (y_l - \mu_l^{(0)}) x_l \right\}^2,$$

$1/V_l$ $(y_l - \mu_l)^2$ $(\bar{y} - \mu^{(0)})^2$

onde as quantidades assinaladas com um (0) são avaliadas em $\beta^{(0)}$. Neste caso, A_2 e A_3 são dados por

$$A_2 = 3\phi^{-1} \left(\sum w_l x_l^2 \right)^{-2} \sum b_l x_l^4$$

e

$$A_3 = 5\phi^{-1} \left(\sum w_l x_l^2 \right)^{-3} \left\{ \sum (f - g)_l x_l^3 \right\}^2,$$

onde w , f , g e b devem também ser avaliados em $\beta^{(0)}$.

Se, além de β ser unidimensional, as variáveis forem identicamente distribuídas, sendo X um vetor de uns, tem-se

$$A_2 = 3(\phi n)^{-1} \left\{ V^{-1} \left(\frac{dV}{d\mu} \right)^2 + \frac{d^2 V}{d\mu^2} \right\} \quad (3.13)$$

e

$$A_3 = 5(\phi V n)^{-1} \left(\frac{dV}{d\mu} \right)^2. \quad (3.14)$$

Nota-se que A_2 e A_3 dependem somente do parâmetro de dispersão e da função de variância e não dependem, como esperado, da função de ligação. Aqui

$$S_R = n\phi(\bar{y} - \mu^{(0)})^2 / V^{(0)} \quad (3.15)$$

também não envolve a ligação. É fácil verificar que as fórmulas em (3.13)–(3.15) estão corretas calculando os três primeiros momentos de S_R sob H_0 e obtendo A_1 , A_2 e A_3 a partir das expressões em (2.10). Fazendo isto, obtêm-se $\mu'_1 = 1$, $\mu'_2 = 3 + (n\phi V)^{-1} \{ (dV/d\mu)^2 + V d^2 V / d\mu^2 \}$ e $\mu'_3 = 15 + 15(n\phi)^{-1} d^2 V / d\mu^2 + 25(n\phi V)(dV/d\mu)^2 + O(n^{-2})$. Para a obtenção de μ'_2 e μ'_3 foi utilizada a relação de recorrência $\kappa_{i+1} = \phi^{-1} d\kappa_i / d\mu$, onde aqui κ_j denota o j -ésimo cumulante de y . Igualando μ'_1 , μ'_2 e μ'_3 às correspondentes expressões em (2.10) chega-se a $A_1 = 0$ e a A_2 e A_3 dados em (3.13) e (3.14) respectivamente.

$\beta = \beta^{(0)}$

Exemplo 3.1

Fornecemos na Tabela 3.2 algumas aplicações das fórmulas (3.13)-(3.15) a diversos modelos. A estatística escore modificada para cada caso segue de (2.15). No modelo binomial, consideramos, na verdade, n variáveis aleatórias independentes tendo distribuição de Bernoulli com probabilidade de sucesso μ e, neste caso, $\phi = 1$. Para os modelos normal, gama e normal inverso assumimos conhecido o parâmetro de dispersão.

Nota-se que os resultados do modelo exponencial (gama com $\phi = 1$) coincidem com os do Exemplo 2.1. Quanto ao modelo normal, é evidente que $A_1 = A_2 = A_3 = 0$ e que, portanto, $S_R^* = S_R$ já que S_R tem distribuição χ_1^2 sob a hipótese nula.

3.7. MODELO DE CLASSIFICAÇÃO A UM FATOR

Consideremos $p \geq 2$ populações tendo densidade (3.1) com ϕ suposto conhecido e suponhamos que amostras aleatórias independentes de tamanhos $n_1, \dots, n_p$ ($n_i \geq 1$, $i = 1, \dots, p$) foram retiradas destas populações. Assumimos a estrutura linear $\eta_i = \beta + \beta_i$, $i = 1, \dots, p$ com $\beta_p = 0$. Aqui β_i é o efeito da i -ésima população na variável resposta e η está relacionado com $\mu = E(y)$ segundo alguma função de ligação. Denotamos as médias amostrais por $\bar{y}_1, \dots, \bar{y}_p$, o número total de observações por $n = \sum n_i$ e a média geral por $\bar{y} = \sum n_i \bar{y}_i / n$.

A estatística escore para testar a homogeneidade das médias das p populações, isto é, testar $H_0 : \beta_1 = \dots = \beta_{p-1} = 0$ contra a alternativa $H : \beta_i$ não é constante, é dada por

$$S_R = \phi \tilde{V}^{-1} \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2, \quad (3.16)$$

onde $\tilde{V}$ é o estimador de máxima verossimilhança de V sob H_0 . Assintoticamente, S_R tem distribuição χ_{p-1}^2 sob a hipótese nula. É conveniente notar que S_R e os A 's não dependem da ligação.

Tabela 3.2. Quantidades A_1 , A_2 e A_3 e estatísticas S_R e S_R^* para os testes do Exemplo 3.1

Modelo	A_2	A_3	S_R	S_R^*
Normal $N(\mu, \sigma^2)$	0	0	$\frac{n}{\sigma^2}(\bar{y} - \mu^{(0)})^2$	S_R
Poisson $P(\mu)$	$\frac{3}{n\mu}$	$\frac{5}{n\mu}$	$\frac{n}{\mu^{(0)}}(\bar{y} - \mu^{(0)})^2$	$S_R \left[1 - \frac{1}{36n\mu^{(0)}} \{6 - 7S_R + S_R^2\} \right]$
Binomial $B(1, \mu)$	$\frac{3}{n\mu(1-\mu)} - \frac{18}{n}$	$\frac{5}{n\mu(1-\mu)} - \frac{20}{n}$	$\frac{n(\bar{y} - \mu^{(0)})^2}{\mu^{(0)}(1-\mu^{(0)})}$	$S_R \left[1 - \frac{1}{36n\mu^{(0)}(1-\mu^{(0)})} [6 - 6\mu^{(0)}(1-\mu^{(0)}) - \{7 - 22\mu^{(0)}(1-\mu^{(0)})\} S_R + \{1 - 4\mu^{(0)}(1-\mu^{(0)})\} S_R^2] \right]$
Gama $G(\mu, \phi)$	$\frac{18}{n\phi} + \frac{1}{\mu}$	$\frac{20}{n\phi} + \frac{1}{\mu}$	$\frac{n\phi_{\infty}}{\mu^{(0)^2}(\bar{y} - \mu^{(0)})^2} + \dots$	$S_R \left[1 - \frac{1}{18n\phi} \{3 - 11S_R + 2S_R^2\} \right]$
Normal Inversa $N^-(\mu, \phi)$	$\frac{45\mu}{n\phi}$	$\frac{45\mu}{n\phi}$	$\frac{n\phi}{\mu^{(0)^3}(\bar{y} - \mu^{(0)})^2}$	$S_R \left[1 - \frac{\mu^{(0)}}{n\phi} \{-5S_R + S_R^2\} \right]$

Nota: $A_1 = 0$ para todos os modelos.

Para obter os A 's precisamos encontrar Z_2 e $Z - Z_2$ que são funções das matrizes X , X_2 e W de dimensões $n \times (p-1)$, $n \times 1$ e $n \times n$, respectivamente. Neste caso, $W = wI$, onde I é a matriz identidade $n \times n$. Da forma da matriz X para o modelo de classificação a um fator encontramos

$$X^T W X = w \begin{pmatrix} n_1 & 0 & \dots & 0 & n_1 \\ 0 & n_2 & \dots & 0 & n_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & n_{p-1} & n_{p-1} \\ n_1 & n_2 & \dots & n_{p-1} & n \end{pmatrix}$$

e $X_2^T W X_2 = nw$. A inversa de $X^T W X$ segue da fórmula de inversas da matrizes particionadas (Rao, 1973, pg. 33). Obtemos $Z_2 = (nw)^{-1} \mathbf{1} \mathbf{1}^T$

$$Z - Z_2 = \begin{pmatrix} (n - n_1)n_1^{-1}1_{n_1 n_1} & -1_{n_1 n_2} & \dots & -1_{n_1 n_p} \\ -1_{n_2 n_1} & (n - n_2)n_2^{-1}1_{n_2 n_2} & \dots & -1_{n_2 n_p} \\ \vdots & \vdots & \ddots & \vdots \\ -1_{n_p n_1} & -1_{n_p n_2} & \dots & (n - n_p)n_p^{-1}1_{n_p n_p} \end{pmatrix}$$

onde 1_{rs} é uma matriz $r \times s$ de uns. De (3.9)-(3.11) e das matrizes Z e $Z - Z_2$ acima, vem facilmente

$$A_1 = -6 \frac{(p-1)}{n\phi} \frac{d^2 V}{d\mu^2}, \quad (3.17)$$

$$A_2 = \frac{3}{n\phi} \left[n \left\{ \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 + \frac{d^2 V}{d\mu^2} \right\} \sum_{i=1}^p n_i^{-1} - (p^2 + 2p - 2) \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 - (2p - 1) \frac{d^2 V}{d\mu^2} \right] \quad (3.18)$$

e

$$A_3 = \frac{1}{n\phi} \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 \left\{ 5n \sum_{i=1}^p n_i^{-1} - (3p^2 + 6p - 4) \right\}. \quad (3.19)$$

Consideremos o caso especial $n_1 = \dots = n_p = 1$ e $n = p$ que corresponde a testar a homogeneidade de médias, isto é, testar $H_0 : \mu_1 = \dots = \mu_n (= \mu)$ contra $H : \mu_i$ não é constante. A estatística escore é $S_R = \phi \tilde{V}^{-1} \sum (y_i - \bar{y})^2$ com $\tilde{V} = V(\bar{y})$ e (3.17)-(3.19) conduzem a $A_1 = -6(n-1)(n\phi)^{-1}(d^2 V/d\mu^2)$, $A_2 = 3(n\phi)^{-1}\{(n^2 - 2n + 1)d^2 V/d\mu^2 - 2(n-1)V^{-1}(dV/d\mu)^2\}$ e $A_3 = 2(n-1)(n-2)(n\phi)^{-1}$.

Exemplo 3.2

Para o modelo Poisson ($\phi = 1$ e $V(\mu) = \mu$) as expressões acima reduzem-se a $S_R = \bar{y}^{-1} \sum (y_i - \bar{y})^2$, $A_1 = 0$, $A_2 = -6(n-1)(n\mu)^{-1}$ e $A_3 = 2(n-1)(n-2)(n\mu)^{-1}$. Das fórmulas em (2.10) obtemos os três primeiros momentos ordinários de S_R , até ordem n^{-1} , e sob H_0 dados por $n-1$, $(n-1)(n+1) - 2(n-1)/(n\mu)$ e $(n-1)(n+1)(n+3) - 2(n-1)(n+5)/(n\mu)$ respectivamente. Notamos que estes estão de acordo com os momentos exatos obtidos por Haldane (1937). Este exemplo foi estudado por Harris (1985) e serve como uma verificação parcial de nossos resultados.

Exemplo 3.3

Na comparação de médias de p populações normais com variância σ^2 comum e conhecida, os A 's são nulos já que $S_R = \sigma^{-2} \sum n_i(\bar{y}_i - \bar{y})^2$ tem distribuição χ^2_{p-1} .

Consideremos agora a comparação de médias de $p = 2$ populações das quais foram extraídas amostras de tamanhos $n_1 = n_2 = n/2$. A estatística score em (3.16) pode ser escrita como

$$S_R = \frac{n\phi}{4\tilde{V}}(\bar{y}_1 - \bar{y}_2)^2, \quad (3.20)$$

onde $\tilde{V} = V(\bar{y})$, e os A 's em (3.17)-(3.19) reduzem-se a

$$A_1 = -\frac{6}{n\phi} \frac{d^2 V}{d\mu^2}, \quad (3.21)$$

$$A_2 = -\frac{3}{n\phi} \left\{ \frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - \frac{d^2 V}{d\mu^2} \right\} \quad (3.22)$$

enquanto A_3 se anula.

Exemplo 3.4

Algumas aplicações das fórmulas (3.20)-(3.22) são apresentadas na Tabela 3.3. Note-se que nos modelos normal, gama e normal inverso supomos que o parâmetro de dispersão é conhecido e igual para as duas populações.

Tabela 3.3. Quantidades A_1 , A_2 e A_3 e estatísticas S_R e S_R^* para os testes do Exemplo 3.4.

Modelo	A_1	A_2	S_R	S_R^*
Normal	0	0	$\frac{n}{4\sigma^2}(\bar{y}_1 - \bar{y}_2)^2$	S_R
Poisson	0	$-\frac{6}{n\mu}$	$\frac{n}{4\bar{y}}(\bar{y}_1 - \bar{y}_2)^2$	$S_R \left\{ 1 - \frac{1}{6n\mu}(3 - S_R) \right\}$
Binomial	$\frac{12}{n}$	$-\frac{6}{n\mu(1-\mu)} + \frac{18}{n}$	$\frac{n(\bar{y}_1 - \bar{y}_2)^2}{4\bar{y}(1-\bar{y})}$	$S_R \left[1 - \frac{1}{6n} \left\{ -3 + \frac{1}{\mu(1-\mu)} - \left(\frac{1}{\mu(1-\mu)} - 3 \right) S_R \right\} \right]$
Gama	$-\frac{12}{n\phi}$	$-\frac{18}{n\phi}$	$\frac{n\phi}{4\bar{y}^2}(\bar{y}_1 - \bar{y}_2)^2$	$S_R \left\{ 1 - \frac{1}{2n\phi}(1 - S_R) \right\}$
Normal Inverso	$-\frac{36\mu}{n\phi}$	$-\frac{36\mu}{n\phi}$	$\frac{n\phi}{4\bar{y}^3}(\bar{y}_1 - \bar{y}_2)^2$	$S_R \left\{ 1 + \frac{\mu}{n\phi} S_R \right\}$

Nota: $A_3 = 0$ para todos os modelos.

Exemplo 3.5

Aplicamos as fórmulas (3.17)–(3.19) ao teste de homogeneidade de médias de p populações seguindo um modelo gama com índice comum ϕ suposto conhecido. Consideramos o caso balanceado, isto é, as amostras têm tamanhos iguais $n_i = n/p$, $i = 1, \dots, p$. Como $V = \mu^2$, vem de (3.17)–(3.19): $A_1 = -12(p-1)/(n\phi)$, $A_2 = 6(p-1)(p-6)/(n\phi)$ e $A_3 = 8(p-1)(p-2)/(n\phi)$. Testes escore aperfeiçoados podem ser baseados na estatística modificada S_R^* ou na própria estatística escore $S_R = n\phi p^{-1} \bar{y}^{-2} \sum_{i=1}^p (\bar{y}_i - \bar{y})^2$ em conjunto com a expansão de Harris para os seus quantis. Substituindo-se os A 's em (2.15) e (2.19), chega-se a

$$S_R^* = S_R \{1 - b(S_R, n, \phi, p)\}$$

e

$$z_\alpha = x_\alpha \{1 + b(x_\alpha, n, \phi, p)\}, \quad (3.23)$$

onde

$$b(x, n, \phi, p) = \frac{1}{6n\phi} \left\{ p + 1 - \frac{5p-1}{p+1}x + \frac{4(p-2)}{(p+1)(p+3)}x^2 \right\}$$

e x_α é tal que $P(\chi_{p-1}^2 \leq x_\alpha) = 1 - \alpha$.

Consideramos o Exemplo 3.5 para apresentar um estudo de simulação que visa comparar o teste escore usual com os testes modificados. Sob a hipótese nula, calculamos, via simulação (*), as taxas de rejeição de S_R e de S_R^* , isto é, as probabilidades $P(S_R \geq x_\alpha)$ e $P(S_R^* \geq x_\alpha)$, e $P(S_R \geq z_\alpha)$, onde z_α é dado em (3.23). Amostras foram geradas para diversas combinações variando os valores de p ($p = 2, 3, 4$), o tamanho das amostras ($n_i = 5, 10, 15$ e 20) e o índice ϕ ($\phi = 0,5, 1,0$ e $2,0$). As probabilidades citadas acima foram investigadas aos níveis nominais de 1,0, 2,5, 5,0 e 10,0%. A média comum das

(*) Os estudos de simulação apresentados nesta tese foram desenvolvidos no sistema GLIM processado no computador UNISYS B7900 da Universidade de São Paulo. O diretivo \$LSEED e a função %LR do GLIM foram utilizados para gerar as amostras.

observações para todas as simulações foi fixada em um. Para $p = 2$ populações, a Tabela 3.4 mostra a proporção de vezes em que S_R e S_R^* excederam os valores críticos obtidos da distribuição χ^2_{p-1} aos diversos níveis nominais e que S_R excedeu os correspondentes valores críticos modificados dados (3.23), em 10000 simulações. As Tabelas 3.5 e 3.6 fazem o mesmo para $p = 3$ e $p = 4$ populações.

Destas tabelas fica claro que as probabilidades $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ tendem a ser mais próximas dos correspondentes níveis nominais do que $P(S_R \geq x_\alpha)$. O teste escore usual se mostrou conservativo no sentido de que as taxas de rejeição de S_R estão sempre abaixo dos níveis nominais correspondentes. Como esperado, a estatística escore não tem uma boa aproximação por χ^2 quando o número total de observações não é alto. Isto se deteriora conforme ϕ decresce, isto é, quando o coeficiente de variação se torna grande. Assim, os testes modificados se tornam mais necessários para valores pequenos de ϕ . Evidentemente, conforme cresce o tamanho total da amostra, as probabilidades $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ vão se aproximando e pouca diferença faz em se optar por um teste em lugar de outro. Para a grande maioria dos 36 casos apresentados nas Tabelas 3.4, 3.5 e 3.6, há alguma evidência de superioridade da aproximação por χ^2 para os testes ajustados sobre o teste escore usual. Com base nisto, que confirma os resultados do Capítulo 2, a distribuição de S_R^* parece ser mais bem representada por χ^2 do que a distribuição de S_R . Além disto, como sugerido por Harris (1985), que o teste modificado baseado em (3.23) é melhor que aquele baseado na estatística escore S_R com sua aproximação de primeira ordem por χ^2 .

3.8. TESTE DA VARIÁVEL ADICIONADA

O objetivo desta seção é apresentar um teste escore aperfeiçoado para verificar a necessidade de inclusão de uma particular variável auxiliar em um MLG. Em outras palavras, estamos interessados em testar $H_0 : \beta_1 = 0$ versus $H : \beta_1 \neq 0$ onde a dimensão

Tabela 3.4. Valores simulados de $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ (em porcentagem) para o teste de homogeneidade de médias de duas populações com distribuição gama

n_i	Níveis nominais (%)	$\phi = 0,5$			$\phi = 1,0$			$\phi = 2,0$		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
5	1,0	0,0	0,0	4,6	0,2	0,7	1,4	0,5	1,0	1,2
	2,5	0,0	0,8	4,0	1,1	2,2	2,9	1,9	2,5	2,6
	5,0	1,0	3,1	5,6	3,3	4,8	5,4	4,3	5,0	5,2
	10,0	6,0	8,5	9,8	8,4	9,8	10,1	9,2	9,9	10,0
10	1,0	0,2	0,7	1,3	0,5	0,9	1,0	0,7	0,9	0,9
	2,5	0,9	2,0	2,7	1,9	2,5	2,7	2,1	2,4	2,4
	5,0	3,2	4,5	5,1	4,2	5,0	5,1	4,4	4,8	4,8
	10,0	8,5	9,8	10,2	9,4	10,3	10,4	9,7	10,2	10,2
15	1,0	0,3	0,7	1,0	0,7	1,0	1,1	0,7	0,8	0,8
	2,5	1,4	2,4	2,7	2,0	2,5	2,6	2,0	2,3	2,4
	5,0	3,7	4,5	4,7	4,7	5,2	5,2	4,6	4,8	4,8
	10,0	8,7	9,5	9,7	9,8	10,2	10,2	9,2	9,4	9,4
20	1,0	0,4	0,8	0,9	0,8	1,0	1,1	0,7	0,8	0,8
	2,5	1,5	2,0	2,2	2,2	2,5	2,6	2,1	2,3	2,3
	5,0	4,0	4,6	4,8	4,5	5,0	5,0	4,7	4,9	4,9
	10,0	8,7	9,5	9,5	9,6	9,9	9,9	9,8	10,0	10,0

Nota: Os números (1), (2) e (3) correspondem a $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.5. Valores simulados de $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ (em porcentagem) para o teste de homogeneidade de médias de três populações com distribuição gama

n_i	Níveis nominais (%)	$\phi = 0,5$			$\phi = 1,0$			$\phi = 2,0$		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
5	1,0	0,2	1,0	1,9	0,6	1,0	1,2	0,9	1,1	1,2
	2,5	0,8	2,5	3,8	1,5	2,3	2,6	2,3	2,8	2,8
	5,0	2,4	4,7	6,4	3,1	4,5	4,7	4,3	5,0	5,2
	10,0	6,1	9,0	10,6	7,1	8,9	9,4	9,1	10,2	10,3
10	1,0	0,6	1,2	1,4	0,8	1,1	1,2	1,0	1,1	1,1
	2,5	1,8	2,7	3,0	1,9	2,4	2,5	2,4	2,8	2,8
	5,0	3,6	5,1	5,3	4,1	4,9	5,0	4,5	5,1	5,1
	10,0	7,8	9,9	10,4	9,3	10,2	10,4	9,2	9,7	9,7
15	1,0	0,8	1,2	1,2	0,7	0,9	0,9	1,0	1,1	1,1
	2,5	1,7	2,3	2,4	2,0	2,4	2,4	2,3	2,6	2,6
	5,0	3,7	4,7	4,8	4,2	4,8	4,8	5,0	5,3	5,3
	10,0	8,1	9,6	9,8	8,8	9,5	9,5	9,9	10,2	10,2
20	1,0	0,8	1,0	1,1	0,7	0,9	0,9	1,0	1,2	1,2
	2,5	1,9	2,3	2,3	2,1	2,4	2,4	2,4	2,5	2,5
	5,0	3,8	4,6	4,7	4,3	4,7	4,8	4,8	5,0	5,0
	10,0	9,1	10,0	10,1	9,3	9,8	9,8	9,6	9,8	9,8

Nota: Os números (1), (2) e (3) correspondem a $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.6. Valores simulados de $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ (em porcentagem) para o teste de homogeneidade de médias de quatro populações com distribuição gama

n_i	Níveis nominais (%)	$\phi = 0,5$			$\phi = 1,0$			$\phi = 2,0$		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
5	1,0	0,5	1,2	1,1	0,9	1,4	1,3	1,0	1,2	1,1
	2,5	1,4	2,8	3,1	2,1	3,0	3,1	2,2	2,7	2,7
	5,0	3,0	5,5	6,3	4,2	5,4	5,6	4,6	5,4	5,4
	10,0	6,8	10,3	11,3	8,5	10,4	10,6	8,7	9,7	9,8
10	1,0	0,7	1,2	1,1	0,9	1,1	1,1	0,8	1,0	1,0
	2,5	2,0	2,8	2,8	2,0	2,4	2,5	2,4	2,6	2,6
	5,0	4,0	5,2	5,5	4,0	4,7	4,8	4,5	5,0	5,0
	10,0	8,3	10,3	10,6	9,0	10,0	10,1	9,3	9,8	9,9
15	1,0	0,9	1,0	1,0	1,0	1,1	1,1	0,9	1,0	1,0
	2,5	1,9	2,5	2,5	2,1	2,4	2,4	2,1	2,3	2,3
	5,0	4,0	5,0	5,1	4,5	4,9	4,9	4,8	5,1	5,1
	10,0	8,7	10,0	10,1	9,1	9,7	9,8	9,6	9,9	9,9
20	1,0	0,9	1,1	1,1	0,9	1,0	1,0	0,9	0,9	0,9
	2,5	2,2	2,6	2,6	2,1	2,3	2,3	2,3	2,4	2,4
	5,0	4,3	4,9	5,0	4,5	4,9	4,9	4,9	5,0	5,0
	10,0	8,9	10,1	10,2	9,5	10,2	10,3	9,6	9,9	9,9

Nota: Os números (1), (2) e (3) correspondem a $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

de β_1 é $r = 1$. Wang (1985) utilizou a estatística escore e sua versão gráfica para o teste destas hipóteses.

Notando que X_1 é um vetor $n \times 1$, segue de (3.3)

$$S_R = \phi \frac{\left\{ \sum (\tilde{w}_l / \tilde{V}_l)^{1/2} x_l (y_l - \tilde{\mu}_l) \right\}^2}{\sum \tilde{w}_l \tilde{r}_l^2}, \quad (3.24)$$

onde r_l é o resíduo correspondente ao l -ésimo elemento do vetor X_1 , após a regressão normal linear de X_1 sobre a matriz modelo X_2 com matriz de pesos W . Da forma de $Z - Z_2$ em (3.12) vemos que o elemento (k, l) desta matriz é $r_k r_l / (\sum w_l r_l^2)$ que substituído em (3.9)-(3.11) conduz a

$$A_1 = \frac{1}{\phi(\sum w_l r_l^2)} \left\{ 3 \left(\sum f_l z_{2ll} r_l \right)^2 + 6 \sum f_l z_{2ll} z_{2lk} r_k^2 (f - g)_k - 6 \sum f_l z_{2lk}^2 r_l r_k (2g - f)_k - 6 \sum h_l r_l^2 z_{2ll} \right\}, \quad (3.25)$$

$$A_2 = \frac{1}{\phi(\sum w_l r_l^2)^2} \left\{ -9 \sum (f - g)_l r_l^2 z_{2lk} r_k^2 (f - g)_k - 6 \left(\sum f_l z_{2ll} r_l \right) \left(\sum (f - g)_l r_l^3 \right) + 3 \sum b_l r_l^4 \right\} \quad (3.26)$$

e

$$A_3 = \frac{5}{\phi(\sum w_l r_l^2)^3} \left(\sum (f - g)_l r_l^3 \right)^2, \quad (3.27)$$

onde z_{2lk} é o elemento (l, k) da matriz Z_2 em (3.8).

É importante ressaltar que a estatística escore (3.24) e os A 's dados em (3.25)-(3.27) também se aplicam ao teste das hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$, onde o escalar $\beta_1^{(0)}$ pode não ser nulo.

3.8.1. MODELO DE REGRESSÃO COM DOIS PARÂMETROS

Consideremos um MLG com componente sistemática $\eta = \alpha u_l + \beta x_l$, $l = 1, \dots, n$, onde u_l e x_l denotam os valores de duas variáveis auxiliares associadas à l -ésima observação. A estatística escore para testar $H_0 : \beta = \beta^{(0)}$ contra $H : \beta \neq \beta^{(0)}$ pode ser obtida

de (3.24) notando que $r_l = x_l - u_l(\sum w_l x_l u_l)(\sum w_l u_l^2)^{-1}$ e $\sum w_l r_l^2 = (\sum w_l u_l^2)^{-1} \times \{(\sum w_l x_l^2)(\sum w_l u_l^2) - (\sum w_l x_l u_l)^2\}$, conduzindo a

$$S_R = \phi \frac{\left(\sum \tilde{w}_l u_l^2\right) \left\{ \sum (\tilde{w}_l / \tilde{V}_l)^{1/2} x_l (y_l - \tilde{\mu}_l) \right\}^2}{\left(\sum \tilde{w}_l x_l^2\right) \left(\sum \tilde{w}_l u_l^2\right) - \left(\sum \tilde{w}_l x_l u_l\right)^2}.$$

Para este problema o elemento (l, k) de Z_2 é $u_l u_k / \sum w_l u_l^2$. Substituindo a expressão de z_{2lk} em (3.25) e (3.26) chega-se a

$$A_1 = \frac{1}{\phi(\sum w_l r_l^2)(\sum w_l u_l^2)^2} \left[3 \left(\sum f_l u_l^2 r_l \right)^2 + 6 \left(\sum f_l u_l^3 \right) \left\{ \sum (f - g)_l u_l r_l^2 \right\} \right. \\ \left. - 6 \left(\sum f_l u_l^2 r_l \right) \left\{ \sum (2g - f)_l u_l^2 r_l \right\} - 6 \left(\sum h_l u_l^2 r_l^2 \right) \left(\sum w_l u_l^2 \right) \right]$$

e

$$A_2 = \frac{1}{\phi(\sum w_l r_l^2)^2 \sum w_l u_l^2} \left[-9 \left\{ \sum (f - g)_l u_l r_l^2 \right\}^2 - 6 \left(\sum f_l u_l^2 r_l \right) \left\{ \sum (f - g)_l r_l^3 \right\} \right. \\ \left. + 3 \left(\sum b_l r_l^4 \right) \left(\sum w_l u_l^2 \right) \right]$$

e A_3 é dado em (3.27).

As fórmulas para o modelo de regressão simples, que corresponde a considerar a componente sistemática $\eta_{\mathcal{L}} = \alpha + \beta x_l$, são deduzidas das expressões acima substituindo $u_l = 1$, $l = 1, \dots, n$. Em particular, para testar $H_0 : \beta = 0$ contra $H : \beta \neq 0$, a estatística escore e os A 's, que aqui não dependem da função de ligação, têm formas bastante simples dadas por

$$S_R = \frac{n\phi}{V} \frac{\bar{s}_2^2}{\bar{s}_2}, \quad V = V(\bar{y}) \quad (3.28)$$

$$A_1 = -\frac{6}{n\phi} \frac{d^2 V}{d\mu^2}, \quad (3.29)$$

$$A_2 = -\frac{3}{n\phi V} \left[3 \left(\frac{dV}{d\mu} \right)^2 - \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2 V}{d\mu^2} \right\} \frac{\bar{s}_4}{\bar{s}_2^2} \right] \quad (3.30)$$

e

$$A_3 = \frac{5}{n\phi V} \left(\frac{dV}{d\mu} \right)^2 \frac{\bar{s}_3^2}{\bar{s}_2^3}, \quad (3.31)$$

onde $\bar{s}_a = n^{-1} \sum (x_l - \bar{x})^a$, para $a = 2, 3, 4$ e $\bar{s} = n^{-1} (\sum x_l y_l - n \bar{x} \bar{y})$, sendo que as quantidades que dependem do parâmetro β devem ser avaliadas sob H_0 .

Usamos a notação γ_{1y} e γ_{2y} para o terceiro e quarto cumulantes padronizados da distribuição de y . Em outras palavras, γ_{1y} e γ_{2y} são as medidas usuais de assimetria e curtose da distribuição de y . Notação similar é utilizada para os cumulantes amostrais de x , ou seja, γ_{1x} e γ_{2x} . Os A 's podem ser escritos explicitamente em termos destas quantidades como: $A_1 = \frac{6}{n}(\gamma_{1y}^2 - \gamma_{2y})$, $A_2 = \frac{3}{n}\{\gamma_{2y}(\gamma_{2x} + 3) - 3\gamma_{1y}^2\}$ e $A_3 = \frac{5}{n}\gamma_{1y}^2\gamma_{1x}^2$. Evidentemente, se y é normalmente distribuído, $\gamma_{1y} = \gamma_{2y} = 0$ e os A 's são iguais a zero. Entretanto, se não for este o caso, a aproximação da distribuição da estatística escore por χ^2 é particularmente sensível a mudanças nas medidas amostrais de assimetria e curtose de x .

Apresentamos um pequeno estudo de simulação em um modelo de regressão simples com erro exponencial. Em outras palavras, supomos que y tem distribuição exponencial com média μ relacionada ao preditor linear $\eta = \alpha + \beta x$ através de uma função de ligação qualquer. A variável auxiliar x é gerada segundo diversas distribuições e, para cada situação estudada, uma única amostra de valores de x é mantida fixada. Claramente, os resultados são influenciados pelas medidas de assimetria e curtose da particular amostra obtida.

Dez mil amostras foram geradas para todas as combinações dos tamanhos da amostra ($n=5, 10, 15$ e 20) e das distribuições de x (normal com média 0 e variância 1, uniforme no intervalo $(0,1)$ e exponencial com média 1). As probabilidades $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ foram calculadas via simulação para $\alpha = 1, 2, 5, 5, 0$, e 10%. Aqui x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$ e z_α vem de (2.19) com os A 's dados em (3.29)–(3.31). Para todas as amostras, β foi fixado em 0, já que o interesse é testar $H_0 : \beta = 0$, e μ fixado em 1. Os resultados são apresentados na Tabela 3.7.

Notamos que o teste escore usual tende a ser conservativo apresentando, em geral, taxas de rejeição abaixo dos correspondentes níveis nominais. Observamos ainda que, para

Tabela 3.7. Valores simulados de $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ (em porcentagem) para o teste de $H_0 : \beta = 0$ em modelo de regressão simples com erro exponencial

n	Níveis nominais (%)	$N(0,1)$			$U(0,1)$			$\text{Exp}(1)$		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
5	1,0	0,5	0,8	0,7	0,0	0,6	3,4	0,3	0,8	1,0
	2,5	1,5	2,5	2,5	0,4	1,7	3,9	1,1	2,3	2,8
	5,0	3,3	5,2	5,4	1,7	3,5	5,7	3,0	5,0	5,6
	10,0	7,3	10,4	11,1	5,6	8,4	9,8	7,3	10,3	11,3
10	1,0	0,4	0,8	1,1	1,2	0,5	1,0	0,8	1,3	1,3
	2,5	1,5	2,4	2,7	2,4	2,4	2,5	1,9	2,7	2,9
	5,0	3,8	4,9	5,1	4,6	5,2	5,2	3,6	5,1	5,5
	10,0	8,5	9,8	10,0	8,6	10,5	10,6	7,5	9,7	10,2
15	1,0	0,8	0,7	0,8	0,8	1,0	1,1	1,1	0,5	1,0
	2,5	2,1	2,4	2,4	2,3	2,7	2,7	2,2	2,4	2,6
	5,0	4,2	4,9	4,9	4,3	5,0	5,1	3,8	4,7	5,0
	10,0	8,7	9,9	10,0	8,7	9,7	9,7	8,0	10,1	10,6
20	1,0	1,0	0,9	0,9	0,7	1,0	1,0	0,8	0,9	0,9
	2,5	2,3	2,4	2,4	2,0	2,4	2,4	2,2	2,5	2,5
	5,0	4,7	5,1	5,0	4,4	4,9	5,0	4,2	4,8	4,8
	10,0	9,0	10,0	10,0	9,4	10,2	10,2	8,9	9,9	10,0

Nota: 1. $N(0,1)$, $U(0,1)$ e $\text{Exp}(1)$ correspondem às distribuições a partir das quais foi gerada a variável auxiliar.

2. Os números (1), (2) e (3) correspondem a $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

a grande maioria dos casos estudados, há alguma evidência de que a distribuição de S_R^* está melhor representada por χ^2 do que a de S_R e, além disto, que o teste modificado, baseado na expansão dos quantis de S_R , tende a ser melhor que o teste escore usual.

Exemplo 3.6

Usamos as fórmulas (3.28)–(3.31) para apresentar uma ilustração numérica de um teste escore aperfeiçoado. Consideremos os dados da Tabela 3.8 analisados por Cox e Snell (1981; pgs. 148-150) relativos a tempos de sobrevivência y de 17 pacientes sofrendo de leucemia. Eles consideraram o modelo de regressão exponencial com estrutura log-linear $\log \mu_l = \alpha + \beta x_l$, onde μ_l é o tempo médio de sobrevivência desde o diagnóstico e x_l é o logaritmo na base 10 da contagem inicial de células brancas no sangue. McCullagh e Nelder (1989, pg.464) consideraram a estatística da razão de verossimilhança ω e a sua versão modificada $\omega^* = \omega(1 - b)$, onde $1 - b$ é o fator de correção de Bartlett, para o teste de $H_0 : \beta = 0$ versus $H : \beta \neq 0$. Os valores de ambas as estatísticas sugerem que H_0 é rejeitada ao nível de significância de 1%.

Tabela 3.8. Tempo de sobrevivência em semanas e $\log_{10}$ da contagem inicial de células brancas para 17 pacientes com leucemia

x	y	x	y	x	y
3,36	65	4,00	121	4,54	22
2,88	156	4,23	4	5,00	1
3,63	100	3,73	39	5,00	1
3,41	134	3,85	143	4,72	5
3,78	16	3,97	56	5,00	65
4,02	108	4,51	26		

De (3.28), a estatística escore para testar $H_0 : \beta = 0$ versus $H : \beta \neq 0$ é dada por $S_R = n\bar{s}^2(\bar{y}^2 \bar{s}_2)^{-1}$ conduzindo ao valor 5,681. O nível descritivo obtido da distribuição χ_1^2 vale 1,71% indicando que H_0 não é rejeitada ao nível de significância de 1%. Agora, de (3.29)-(3.31) vem $A_1 = -12n^{-1}$, $A_2 = -18n^{-1}(2 - \bar{s}_4 \bar{s}_2^{-2})$ e $A_3 = 20n^{-1} \bar{s}_3^2 \bar{s}_2^{-3}$ levando aos valores $A_1 = -0,7059$, $A_2 = 0,1803$ e $A_3 = 0,0008$.

Para proceder um teste escore aperfeiçoado podemos utilizar a estatística escore modificada em (2.15) e a distribuição χ_1^2 como referência, ou utilizar a própria estatística S_R e as expansões de Harris (1985) para sua distribuição e seus quantis sob H_0 (vide (2.5) e (2.11)). De (2.15) e dos valores de S_R e dos A 's dados acima vem $S_R^* = 5,939$ o que conduz ao nível descritivo de 1,48%. Por outro lado, da expansão em (2.5) para a função distribuição de S_R vem o nível descritivo de 1,46%. O valor crítico, correspondente ao quantil de ordem $1 - \alpha$ ($\alpha = 1\%$), iguala 6,635 enquanto que o valor crítico modificado obtido da expansão em (2.11) para os quantis de S_R vale 6,365. Notamos que, embora a decisão do teste não seja alterada ao nível de significância de 1%, os testes aperfeiçoados indicam mais evidência contra H_0 do que o teste escore usual.

Exemplo 3.7

Considere n observações independentes $y_1, \dots, y_n$ de uma distribuição de Poisson com a estrutura log linear $\log \mu_l = \alpha + \beta x_l$, $l = 1, \dots, n$. A distribuição assintótica do desvio no teste de $H_0 : \beta = \beta^{(0)}$ versus $H : \beta \neq \beta^{(0)}$ foi estudado por Maller (1986). A estatística escore e os A 's podem ser obtidos de (3.24)-(3.27) notando que $f = h = b = w = \mu$, $g = 0$ e que $\tilde{\alpha}$, o estimador de máxima verossimilhança de α sob H_0 , iguala

$$\tilde{\alpha} = \log \left[\sum y_l \left\{ \sum \exp(\beta^{(0)} x_l) \right\}^{-1} \right].$$

Em particular, para o teste de $H_0 : \beta = 0$ versus $H : \beta \neq 0$, $\tilde{\alpha} = \log \bar{y}$ e S_R e os A 's são escritos como $S_R = n\bar{s}^2(\bar{y} \bar{s}_2)^{-1}$, $A_1 = 0$, $A_2 = -3(3 - \bar{s}_4 \bar{s}_2^{-2})(n\mu)^{-2}$ e $A_3 = 5\bar{s}_3^2 \bar{s}_2^{-3}(n\mu)^{-1}$.

3.8.2. MODELO DE CLASSIFICAÇÃO A UM FATOR COM UMA VARIÁVEL ADICIONADA

Consideremos o modelo de classificação a um fator descrito na Seção 3.7 e coloquemos no modelo uma nova variável x , cuja significância queremos testar. Formalmente, consideremos o MLG com estrutura linear $\eta_{ij} = \beta_i + \beta x_{ij}$, $i = 1, \dots, p$, $j = 1, \dots, n_i$ e o teste das hipóteses $H_0 : \beta = 0$ contra $H : \beta \neq 0$. Da forma da matriz X_2 , pode-se ver que Z_2 tem elemento típico $\delta_{ij}(n_i \phi w_i)^{-1}$ onde $\delta_{ij} = 1$ se as observações i e j estão na mesma população e zero caso contrário (vide Cordeiro, 1987; Seção 4). A estatística escore segue de (3.24) com o resíduo r_{ij} substituído por $x_{ij} - \bar{x}_i$, onde $\bar{x}_i$ é a média da variável x na amostra i . Das fórmulas (3.25)-(3.27) chega-se a

$$A_1 = \frac{6}{\phi \sum_{i=1}^p w_i r_i^{(2)}} \sum_{i=1}^p \frac{f_i(f-g)_i - n_i w_i h_i}{n_i^2 w_i^2} r_i^{(2)},$$

$$A_2 = -\frac{3}{\phi \left(\sum_{i=1}^p w_i r_i^{(2)} \right)^2} \sum_{i=1}^p \frac{3(f-g)_i^2 (r_i^{(2)})^2 - n_i w_i b_i r_i^{(4)}}{n_i w_i}$$

e

$$A_3 = \frac{5}{\phi \left(\sum_{i=1}^p w_i r_i^{(2)} \right)^5} \left\{ \sum_{i=1}^p (f-g)_i r_i^{(3)} \right\}^2,$$

onde $r_i^{(a)} = \sum_{j=1}^{n_i} r_{ij}^a$, $a = 2, 3, 4$.

3.9. PODER DOS TESTES APERFEIÇADOS EM MODELOS LINEARES GENERALIZADOS

Na Seção 2.6 apresentamos uma expansão para a função distribuição da estatística escore sob hipóteses alternativas contíguas, isto é, da forma $H : \beta_1 = \beta_1^{(0)} + \xi$, onde $\beta_1 = (\beta_1, \dots, \beta_r)^T$, $\beta_1^{(0)} = (\beta_1^{(0)}, \dots, \beta_r^{(0)})$ e $\xi = (\xi_1, \dots, \xi_r)^T$ com $\xi_i = \beta_i - \beta_i^{(0)}$ suposto ser de ordem $n^{-1/2}$. Vimos naquela seção que as estatísticas S_R e S_R^* têm, sob H , uma

distribuição assintótica qui-quadrado não-central com r graus de liberdade e um certo parâmetro de não-centralidade λ . Vimos também que, até ordem $n^{-1/2}$, os testes escore usuais e aperfeiçoados têm o mesmo poder sob as alternativas contíguas. Nosso objetivo aqui é determinar o poder, até esta ordem, dos testes discutidos nas seções anteriores. Basicamente, será necessário encontrar as quantidades $a_0, \dots, a_3$ envolvidas na expressão (2.21) e o parâmetro λ .

Fórmulas matriciais para $a_0, \dots, a_3$ são desenvolvidas de forma semelhante à da obtenção de A_1, A_2 e A_3 como descrito na Seção 3.3. Ilustramos aqui somente o desenvolvimento da primeira parcela de a_1 que denotamos por a_{11} .

Como $\kappa_{ijk} - 2\kappa_{i,j,k} = -(\phi/3) \sum f_l(ijk)_l$, vem de (2.23)

$$a_{11} = -\frac{\phi}{2} \sum f_l(ijk)_l \delta_i \delta_j \delta_k,$$

onde δ_i é o i -ésimo elemento do vetor δ dado em (2.22). Rearranjando os termos tem-se

$$a_{11} = -\frac{\phi}{2} \sum f_l \sum' x_{li} \delta_i \sum' x_{lj} \delta_j \sum x_{lk} \delta_k.$$

Definimos o vetor $t = (t_1, \dots, t_n)^T = X\delta$ e a matriz diagonal T formada pelos elementos de t , isto é, $T = \text{diag}\{t_1, \dots, t_n\}$. É fácil mostrar que $t = R\xi$, onde R é uma matriz de resíduos definida na Seção 3.2. Agora, a_{11} se resume a

$$a_{11} = -\frac{\phi}{2} \sum f_l t_l^3$$

que, escrita em notação matricial, tem a forma

$$a_{11} = -\frac{\phi}{2} \mathbf{1}^T F T^{(3)} \mathbf{1}.$$

De modo semelhante obtemos as demais parcelas de a_1 e os termos a_0, a_2 e a_3 a partir de (2.23)–(2.26). Temos

$$a_1 = \frac{1}{2} [\phi \{-\mathbf{1}^T F T^{(3)} \mathbf{1} + \mathbf{1}^T (F+G) E T^{(2)} \mathbf{1}\} - \mathbf{1}^T (F-G)(Z-Z_2)_d T \mathbf{1} - \mathbf{1}^T F Z_{2d} T \mathbf{1}], \quad (3.32)$$

$$a_2 = \frac{1}{2} 1^T (F - G)(Z - Z_2)_d T 1, \quad (3.33)$$

$$a_3 = \frac{1}{6} 1^T (F - G) T^{(3)} 1 \quad (3.34)$$

e $a_0 = -(a_1 + a_2 + a_3)$. Aqui $E = \text{diag}\{e_1, \dots, e_n\}$, onde e_l , para $l = 1, \dots, n$, é o l -ésimo elemento do vetor $e = X_1 \xi$, e as demais matrizes foram definidas anteriormente.

Resta agora obter uma fórmula matricial para o parâmetro de não-centralidade $\lambda = \delta^T K \delta$. Da definição da matriz T vem

$$\lambda = \phi 1^T T W T 1. \quad (3.35)$$

Não é difícil mostrar que λ pode ser escrito como $\lambda = \phi \xi^T R^T W R \xi$.

A parcela de ordem $n^{-1/2}$ da expansão do poder dos testes escore usuais e aperfeiçoados é, portanto, função do parâmetro de dispersão, das médias desconhecidas, da matriz modelo X e das diagonais das matrizes Z_2 e $Z - Z_2$. Notamos que $a_0, \dots, a_3$ dependem das duas primeiras derivadas da função de ligação e somente da primeira derivada da função de variância.

As fórmulas (3.32)–(3.35) simplificam-se bastante em casos especiais. Consideremos, inicialmente, a situação em que as variáveis aleatórias $y_1, \dots, y_n$ são identicamente distribuídas com média μ . A estatística escore para testar $H_0 : \beta = \beta^{(0)}$ ou, equivalentemente, $H_0 : \mu = \mu^{(0)}$ é dada em (3.15). Neste caso, vem de (3.32)–(3.35)

$$a_1 = \frac{1}{2} \left[n\phi \left\{ \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3 \right\} \xi^3 - \frac{1}{V} \frac{dV}{d\mu} \frac{d\mu}{d\eta} \xi \right], \quad (3.36)$$

$$a_2 = \frac{1}{2} \frac{1}{V} \frac{dV}{d\mu} \frac{d\mu}{d\eta} \xi, \quad (3.37)$$

$$a_3 = \frac{n\phi}{6} \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3 \xi^3 \quad (3.38)$$

e

$$\lambda = n\phi \left(\frac{d\mu}{d\eta} \right)^2 \frac{1}{V} \xi^2, \quad (3.39)$$

onde $\xi = \beta - \beta^{(0)} = d(\mu) - d(\mu^{(0)})$.

As expressões em (2.27) fornecem expansões, até ordem $n^{-1/2}$, para os três primeiros cumulantes da estatística escore em função de $a_0, \dots, a_3$ e λ . Podemos, então, verificar se as fórmulas (3.36)–(3.39) estão corretas expandindo os correspondentes cumulantes da estatística S_R dada em (3.15) e comparando as expansões obtidas com as expressões em (2.27).

Escrevendo $S_R = n\phi\{(\bar{y} - \mu) + (\mu - \mu^{(0)})\}^2/V^{(0)}$ vem seu primeiro cumulante

$$\mu'_{1,\lambda} = \frac{V}{V^{(0)}} + \frac{n\phi}{V^{(0)}}(\mu - \mu^{(0)})^2. \quad (3.40)$$

Mas, sob a hipótese alternativa, $\mu = d^{-1}(\beta^{(0)} + \xi)$. Assim, expandindo $V(= V(\mu))$ e $\mu - \mu^{(0)}$ em série de Taylor em torno de $\xi = 0$ e substituindo em (3.40) vem, até ordem $n^{-1/2}$,

$$\mu'_{1,\lambda} = \frac{1}{V^{(0)}} \left(V^{(0)} + \xi \frac{dV}{d\mu} \frac{d\mu}{d\eta} \right) + \frac{n\phi}{V^{(0)}} \left(\xi \frac{d\mu}{d\eta} + \frac{\xi^2}{2} \frac{d^2\mu}{d\eta^2} \right)^2$$

que se reduz a

$$\mu'_{1,\lambda} = 1 + \frac{1}{V} \xi \frac{dV}{d\mu} \frac{d\mu}{d\eta} + \frac{n\phi}{V} \xi^2 \left(\frac{d\mu}{d\eta} \right)^2 + \frac{n\phi}{V} \xi^3 \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2}. \quad (3.41)$$

Na última expressão todas as quantidades, exceto ξ , são avaliadas sob H_0 . É fácil verificar que, substituindo as expressões de $a_0, \dots, a_3$ e λ dadas em (3.36)–(3.39) na expansão do primeiro cumulante de S_R dada em (2.27), chega-se a (3.41). Para assegurar que as fórmulas dos a 's e de λ estão corretas foi necessário ir mais além e expandir o segundo e terceiro cumulantes de S_R . Isto exigiu um desenvolvimento algébrico simples porém extremamente trabalhoso que é omitido aqui.

As expressões (3.36)–(3.39) sofrem simplificações quando a ligação é identidade ($\mu = \eta = \beta$). Neste caso,

$$a_1 = \frac{1}{2} \left\{ -\frac{n\phi}{V^2} \frac{dV}{d\mu} (\mu - \mu^{(0)})^3 - \frac{1}{V} \frac{dV}{d\mu} (\mu - \mu^{(0)}) \right\}, \quad (3.42)$$

$$a_2 = \frac{1}{2V} \frac{dV}{d\mu} (\mu - \mu^{(0)}), \quad (3.43)$$

$$a_3 = \frac{n\phi}{6} \frac{1}{V^2} \frac{dV}{d\mu} (\mu - \mu^{(0)})^3 \quad (3.44)$$

e

$$\lambda = \frac{n\phi}{V} (\mu - \mu^{(0)})^2 \quad (3.45)$$

Exemplo 3.8

Fornecemos aqui algumas aplicações das fórmulas (3.42)–(3.45). Para um modelo normal, para o qual $V = 1$, a_1 , a_2 e a_3 se anulam e $\lambda = n\phi(\mu - \mu^{(0)})^2$. Este resultado era esperado uma vez que S_R tem exatamente uma distribuição $\chi^2_{1,\lambda}$ sob a hipótese alternativa.

Para o modelo gama ($V = \mu^2$) vem de (3.42)–(3.45) $a_1 = -n\phi\Delta^3 - \Delta$, $a_2 = \Delta$, $a_3 = n\phi\Delta^3/3$ e $\lambda = n\phi\Delta^2$, onde $\Delta = (\mu - \mu^{(0)})/\mu^{(0)}$. Portanto, a função de poder do teste escore é dada, até ordem $n^{-1/2}$, por $1 - P(S_R \leq x_\alpha)$ com

$$\begin{aligned} P(S_R \leq x_\alpha) &= G_{1,\lambda}(x_\alpha) + \frac{2}{3}n\phi\Delta^3 G_{1,\lambda}(x_\alpha) - (n\phi\Delta^3 + \Delta)G_{3,\lambda}(x_\alpha) \\ &\quad + \Delta G_{5,\lambda}(x_\alpha) + \frac{n\phi}{3}\Delta^3 G_{7,\lambda}(x_\alpha). \end{aligned}$$

Como anteriormente, $G_{r,\lambda}$ denota a função distribuição de uma variável χ^2 não-central com r graus de liberdade e parâmetro de não-centralidade λ e x_α é tal que $P(\chi_r^2 \leq x_\alpha) = 1 - \alpha$.

No modelo normal inverso ($V = \mu^3$), temos

$$a_1 = -\frac{3}{2} \left(\frac{n\phi\Delta^3}{\mu^{(0)}} + \Delta \right), \quad a_2 = \frac{3}{2}\Delta, \quad a_3 = \frac{n\phi\Delta^3}{2\mu^{(0)}} \quad \text{e} \quad \lambda = \frac{n\Delta^2}{\mu^{(0)}}.$$

Aqui a função de poder do teste escore é dada, até ordem n^{-1} , por $1 - P(S_R \leq x_\alpha)$ com

$$\begin{aligned} P(S_R \leq x_\alpha) &= G_{1,\lambda}(x_\alpha) - \frac{n\phi\Delta^3}{\mu^{(0)}} G_{1,\lambda}(x_\alpha) - \frac{3}{2} \left(\frac{n\phi\Delta^3}{\mu^{(0)}} + \Delta \right) G_{3,\lambda}(x_\alpha) \\ &\quad + \frac{3}{2}\Delta G_{5,\lambda}(x_\alpha) + \frac{n\phi\Delta^3}{2\mu^{(0)}} G_{7,\lambda}(x_\alpha). \end{aligned}$$

Consideremos agora uma hipótese nula simples $H_0 : \beta = \beta^{(0)}$. Neste caso, $Z_2 = 0$, $\delta = \xi$, $E = T$, $t = X\xi$ e, portanto, a_1 e a_2 em (3.32) e (3.33) se reduzem a

$$a_1 = \frac{1}{2} \{ \phi 1^T G T^{(3)} 1 - 1^T (F - G) Z_d T 1 \}$$

e

$$a_2 = \frac{1}{2} 1^T (F - G) Z_d T 1$$

enquanto que a_3 e λ em (3.34) e (3.35) não sofrem simplificações. Em particular, no caso uniparamétrico

$$a_1 = -\frac{1}{2V} \frac{dV}{d\mu} \sum x_l^3 \left(\frac{1}{V} \xi^3 + \frac{\xi}{\sum x_l^2} \right), \quad a_2 = \frac{1}{2V} \frac{dV}{d\mu} \frac{\sum x_l^3}{\sum x_l^2} \xi,$$

$$a_3 = \frac{\phi}{6} \frac{1}{V^2} \frac{dV}{d\mu} \sum x_l^3 \xi^3 \quad \text{e} \quad \lambda = \frac{\phi}{V} \sum x_l^2 \xi^2.$$

Para finalizar, é importante fazer um comentário sobre a utilidade das fórmulas matriciais em (3.32)–(3.35) para $a_0, \dots, a_3$ e λ . Através destas fórmulas podemos obter expressões algébricas para o poder do teste escore até ordem $n^{-1/2}$, como fizemos em alguns casos particulares tratados acima. Fórmulas semelhantes podem ser desenvolvidas para o poder dos testes da razão de verossimilhança e de Wald a partir dos resultados de Hayakawa (1975). Da comparação dos poderes até ordem $n^{-1/2}$, talvez seja possível verificar se um dos testes é mais poderoso que os demais ao menos em alguns MLGs especiais. Esta pesquisa está sendo realizada em conjunto com G.M. Cordeiro e D.A. Botter.

CAPÍTULO 4

FÓRMULAS ESPECIAIS PARA O FATOR DE CORREÇÃO DA ESTATÍSTICA ESCORE

4.1. MODELOS COM PARÂMETROS GLOBALMENTE ORTOGONAIS

4.1.1. INTRODUÇÃO

Consideremos um problema regular cuja log-verossimilhança é função dos vetores de parâmetros $\phi = (\phi_1, \dots, \phi_p)^T$ e $\beta = (\beta_1, \dots, \beta_q)^T$. Diz-se que ϕ e β são globalmente ortogonais se $E(\partial^2 L / \partial \phi_i \partial \beta_{i'}) = 0$ para $i = 1, \dots, p$ e $i' = 1, \dots, q$, quaisquer que sejam ϕ e β no espaço paramétrico. Neste caso, a matriz de informação total de Fisher particionada em ϕ e β é bloco diagonal. Se $E(\partial^2 L / \partial \phi_i \partial \beta_{i'}) = 0$ somente quando avaliada em, digamos, $(\phi^T, \beta^T) = (\phi^{(0)T}, \beta^{(0)T})$, diz-se que a ortogonalidade é local em $(\phi^{(0)T}, \beta^{(0)T})$. Na Seção 4.2 discutiremos a construção de testes escore em problemas com parâmetros localmente ortogonais, porém, a nossa definição para tal ortogonalidade implica na diagonalização total da matriz de informação. Mais especificamente, diremos que os parâmetros

são localmente ortogonais se a matriz de informação total de Fisher avaliada em um determinado ponto do espaço paramétrico for igual a uma matriz identidade multiplicada por n , onde n é o tamanho da amostra. A ortogonalidade local pode sempre ser obtida por transformação linear de parâmetros, enquanto que a ortogonalidade global só é garantida quando um dos vetores paramétricos for unidimensional (Cox e Reid, 1987b; Huzurbazar, 1950; Mitchell, 1962).

Uma discussão bastante abrangente sobre inferência em modelos com parâmetros ortogonais é encontrada no artigo de Cox e Reid (1987b). Eles observam que, se ϕ e β são dois parâmetros unidimensionais globalmente ortogonais, então:

- i) os estimadores de máxima verossimilhança $\hat{\phi}$ e $\hat{\beta}$ são assintoticamente independentes;
- ii) o erro padrão do estimador de máxima verossimilhança de ϕ é o mesmo quer β seja tratado como conhecido ou desconhecido;
- iii) há simplificações na determinação numérica de $(\hat{\phi}, \hat{\beta})$;
- iv) sendo $\hat{\phi}_\beta = \hat{\phi}(\beta)$ o estimador de máxima verossimilhança de ϕ dado o valor de β , tem-se que $\hat{\phi}_\beta - \hat{\phi} = O_p(n^{-1})$. O resultado simétrico também é válido, isto é, $\hat{\beta}_\phi - \hat{\beta} = O_p(n^{-1})$, onde $\hat{\beta}_\phi$ é o estimador de máxima verossimilhança de β para ϕ fixado.

Propriedades análogas a (i)–(iv) são válidas para ϕ e β multidimensionais. É possível mostrar que se $\hat{\phi}_\beta = \hat{\phi}$ então ϕ e β são globalmente ortogonais.

Outra propriedade interessante que mostraremos nas seções seguintes é a de que a obtenção das quantidades envolvidas no fator de correção da estatística score é bastante simplificada se os parâmetros de interesse forem globalmente ortogonais aos de perturbação. O mesmo ocorre no desenvolvimento de fatores de Bartlett para a razão de verossimilhança (Cordeiro, 1986a; Kahle, 1991).

Exemplos de modelos com dois parâmetros globalmente ortogonais incluem o modelo normal com média μ e variância ϕ^{-1} , o modelo gama com média μ e índice ϕ e o modelo normal inverso com média μ e parâmetro de precisão ϕ . Estes exemplos são, na verdade, casos especiais do modelo tratado na Seção 5.1 do próximo capítulo.

Naquela seção considera-se um modelo linear generalizado no qual a distribuição para o erro pertence à família exponencial com parâmetros canônicos ϕ e $\phi\theta$ (vide (5.1)). Neste caso, os parâmetros da regressão são globalmente ortogonais a ϕ . Esta propriedade é válida também para os modelos próprios de dispersão (Barndorff-Nielsen e Jørgensen, 1990) tratados na Seção 5.4.

4.1.2. CONSTRUÇÃO DE PARÂMETROS GLOBALMENTE ORTOGONAIS

Sejam ϕ um parâmetro unidimensional e $\mu = (\mu_1, \dots, \mu_q)^T$ um vetor paramétrico de dimensão $q \geq 1$. O objetivo é determinar uma transformação dos parâmetros (ϕ, μ^T) para (ϕ, β^T) , onde $\beta = (\beta_1, \dots, \beta_q)^T$, de tal forma que ϕ e β sejam globalmente ortogonais. Seja $L^* = L^*(\phi, \mu)$ a log-verossimilhança total como função de ϕ e μ . Cox e Reid (1987b) mostram que o vetor paramétrico β , globalmente ortogonal a ϕ , é dado por qualquer solução do sistema de equações diferenciais parciais

$$\sum_{r=1}^q E\left(\frac{\partial^2 L^*}{\partial \mu_r \partial \mu_s}\right) \frac{\partial \mu_r}{\partial \phi} = -E\left(\frac{\partial^2 L^*}{\partial \phi \partial \mu_s}\right), \quad s = 1, \dots, q. \quad (4.1)$$

Um método alternativo de obtenção de parâmetros ortogonais à média de certas famílias de distribuições foi desenvolvido por Willmot (1990).

A seguir apresentamos aplicações de (4.1) a dois exemplos tratados por Cox e Reid (1987b) e na Seção 4.1.3 retomamos estes exemplos para ilustrar a construção de testes escore aperfeiçoados em problemas com parâmetros unidimensionais ortogonais.

Exemplo 4.1

Sejam $y_{11}, \dots, y_{1n_1}$ e $y_{21}, \dots, y_{2n_2}$ duas amostras aleatórias independentes de distribuições exponenciais com médias μ e $\phi\mu$ respectivamente. O parâmetro de interesse é o quociente entre as médias representado pelo parâmetro ϕ . Consideremos o caso balanceado, ou seja, $n_1 = n_2 = n/2$ onde $n \geq 2$ é um número par.

A log-verossimilhança total de ϕ e μ é

$$L^* = -n \log \mu - \frac{n}{2} \log \phi - \frac{n\bar{y}_1}{2\mu} - \frac{n\bar{y}_2}{2\phi\mu},$$

onde $\bar{y}_1$ e $\bar{y}_2$ representam as duas médias amostrais. É fácil obter $E(\partial^2 L^* / \partial \mu^2) = -n/\mu^2$ e $E(\partial^2 L^* / \partial \phi \partial \mu) = -n/(2\mu\phi)$. A equação em (4.1) reduz-se a

$$-\frac{1}{\mu} \frac{\partial \phi}{\partial \mu} = \frac{1}{2\phi},$$

cuja solução é $a(\beta) = \mu\phi^{1/2}$, onde $a(\cdot)$ é uma função arbitrária. Tomando $a(\beta) = \beta$, vem os parâmetros globalmente ortogonais ϕ e $\beta = \mu\phi^{1/2}$. Na nova parametrização as médias são $\beta\phi^{-1/2}$ e $\beta\phi^{1/2}$ respectivamente.

Exemplo 4.2

Consideremos uma amostra aleatória $y_1, \dots, y_n$ de uma distribuição Weibull com função densidade

$$\pi^*(y; \mu, \phi) = \frac{\phi}{\mu} \left(\frac{y}{\mu}\right)^{\phi-1} \exp\left\{-\left(\frac{y}{\mu}\right)^\phi\right\}, \quad y > 0. \quad (4.2)$$

A log-verossimilhança total de ϕ e μ é

$$L^* = n \log \phi - n\phi \log \mu + (\phi - 1) \sum \log y_i - \sum \left(\frac{y_i}{\mu}\right)^\phi.$$

Para a construção de parâmetros ortogonais neste modelo e para desenvolver testes escore aperfeiçoados para o parâmetro ϕ , as relações abaixo são bastante úteis:

$$\Gamma^{(i)}(j) = \int_0^\infty e^{-t} t^{j-1} (\log t)^i dt, \quad i = 0, \dots, 4, \quad j = 1, 2, 3, \quad (4.3)$$

onde $\Gamma^{(i)}(\cdot)$ denota a i -ésima derivada da função gama com $\Gamma^{(0)}(\cdot) = \Gamma(\cdot)$.

É fácil ver que

$$E\left(\frac{\partial^2 L^*}{\partial \mu \partial \phi}\right) = -\frac{n}{\mu} + \frac{n}{\mu} E\left\{\left(\frac{y}{\mu}\right)^\phi\right\} + E\left\{\left(\frac{y}{\mu}\right)^\phi \log\left(\frac{y}{\mu}\right)\right\},$$

onde y tem função densidade (4.2). Mas, pela transformação de variáveis $x = (y/\mu)^\phi$ chega-se a $E\{(y/\mu)^\phi\} = 1$ e $E\left\{\left(\frac{y}{\mu}\right)^\phi \log\left(\frac{y}{\mu}\right)^\phi\right\} = \Gamma'(2)$ (*). Logo, $E(\partial^2 L^*/\partial\mu\partial\phi) = n\Gamma'(2)/\mu$. Analogamente, mostra-se que $E(\partial^2 L^*/\partial\mu^2) = -n\phi^2/\mu^2$. Assim, a equação (4.1) reduz-se a

$$\frac{\phi^2}{\mu} \frac{\partial\mu}{\partial\phi} = \Gamma'(2)$$

cujas soluções são $a(\beta) = \mu \exp\{\Gamma'(2)/\phi\}$ com $a(\cdot)$ representando qualquer função arbitrária de β . Tomando $a(\beta) = \beta$ temos os parâmetros ortogonais ϕ e $\beta = \mu \exp\{\Gamma'(2)/\phi\}$.

4.1.3. TESTES ESCORE APERFEIÇOADOS

Consideremos a log-verossimilhança de um problema regular parametrizado em função dos parâmetros $\phi = (\phi_1, \dots, \phi_p)^T$ e $\beta = (\beta_1, \dots, \beta_q)^T$ globalmente ortogonais. Aqui o parâmetro ϕ é de interesse e β é um parâmetro de perturbação.

Como na Seção 2.2, usamos a seguinte notação para as derivadas da log-verossimilhança e seus cumulantes: $U_r = \partial L/\partial\beta_r$, $U_{rs} = \partial^2 L/\partial\beta_r\partial\beta_s$, etc e $\kappa_{rs} = E(U_{rs})$, $\kappa_{r,s} = E(U_r U_s)$, $\kappa_{rst} = E(U_{rst})$, $\kappa_{rs,t} = E(U_{rs} U_t)$, $\kappa_{rs,tu} = E(U_{rs} U_{tu}) - \kappa_{rs}\kappa_{tu}$, $\kappa_{r,s,tu} = E(U_r U_s U_{tu}) - \kappa_{r,s}\kappa_{tu}$ e $\kappa_{r,s,t,u} = E(U_r U_s U_t U_u) - \kappa_{r,s}\kappa_{t,u} - \kappa_{r,t}\kappa_{s,u} - \kappa_{r,u}\kappa_{s,t}$. Denotamos as derivadas dos cumulantes em relação aos componentes de β por $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{rs}^{(tu)} = \partial\kappa_{rs}/\partial\beta_t\partial\beta_u$, etc. Para denotar os correspondentes cumulantes em relação aos componentes de ϕ usamos notação do tipo $\kappa_{r's'} = E(U_{r's'}) = E(\partial^2 L/\partial\phi_{r'}\partial\phi_{s'})$. Em outras palavras, os índices i, j, k, r, s, t correspondem aos componentes de β e i', j', k', r', s', t' aos de ϕ . Assim, por exemplo, $\kappa_{rs}^{(r's')} = \partial^2 E(\partial^2 L/\partial\beta_r\partial\beta_s)/\partial\phi_{r'}\partial\phi_{s'}$. A estatística escore para testar $H_0: \phi = \phi^{(0)}$ versus $H: \phi \neq \phi^{(0)}$ tem a forma

$$S_R = \tilde{U}_\phi^T \tilde{K}_{\phi,\phi}^{-1} \tilde{U}_\phi^X, \quad (4.4)$$

onde $U_\phi = (\partial L/\partial\phi_1, \dots, \partial L/\partial\phi_p)^T$, $K_{\phi,\phi} = \{\kappa_{r's'}\}$ sendo as quantidades assinaladas

(*) $\Gamma'(2) = 1 - \gamma$, onde γ é a constante de Euler: $\gamma \cong -0,577$.

com um til avaliadas em $(\phi^T, \beta^T) = (\phi_1^{(0)}, \dots, \phi_p^{(0)}, \tilde{\beta}_1, \dots, \tilde{\beta}_q)$ com $\tilde{\beta}_i$ denotando o estimador de máxima verossimilhança de β_i sob H_0 .

Os cumulantes satisfazem relações (Bartlett, 1953a, b; Lawley, 1956) que, no caso de parâmetros globalmente ortogonais, facilitam enormemente a obtenção de formas simples para as quantidades envolvidas no fator de correção da estatística escore. Destas relações, introduzidas na Seção 2.2, e da ortogonalidade de β e ϕ vem

$$\begin{aligned}
 \kappa_{rr',s} &= -\kappa_{rsr'}, & \kappa_{rs,t} &= -\kappa_{rst} + \kappa_{rs}^{(t)} & \kappa_{r,r',s'} &= 2\kappa_{rr's'} - \kappa_{r's'}^{(r)} \\
 \kappa_{r,s,r'} &= 2\kappa_{rsr'} - \kappa_{rs}^{(r')}, & \kappa_{rs,r'} &= -\kappa_{rsr'} + \kappa_{rs}^{(r')}, \\
 \kappa_{r,s,r',s'} &= -3\kappa_{rsr's'} + 2 \sum_{(4)} \kappa_{rr's'}^{(s)} - \kappa_{rs}^{(r's')} - \kappa_{r's'}^{(rs)} + \sum_{(3)} \kappa_{rs,r's'}, \\
 \kappa_{rs,r',s'} &= \kappa_{rsr's'} - \kappa_{rsr'}^{(s')} - \kappa_{rsr'}^{(r')} + \kappa_{rs}^{(r's')} - \kappa_{rs,r's'}, \\
 \Rightarrow \kappa_{r,s,r',s'} + \kappa_{rs,r',s'} &= -2\kappa_{rsr's'} + 2 \sum_{(4)} \kappa_{rr's'}^{(s)} - \kappa_{rs}^{(r's')} + \sum_{(2)} \kappa_{rr',ss'},
 \end{aligned} \tag{4.5}$$

onde $\sum_{(k)}$ denota o somatório sobre as k permutações de índices.

As matrizes A e M definidas em (2.9) são dadas por

$$A = \begin{pmatrix} 0 & 0 \\ 0 & \kappa_{\beta,\beta}^{-1} \end{pmatrix} \quad \text{e} \quad M = K^{-1} - A = \begin{pmatrix} \kappa_{\phi,\phi}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{4.6}$$

onde $K_{\beta,\beta} = -K_{\beta\beta} = \{-\kappa_{rs}\}$ é uma matriz $q \times q$ e $K_{\phi,\phi} = -K_{\phi\phi} = \{\kappa_{r's'}\}$ é uma matriz $p \times p$. Os elementos de $K_{\beta,\beta}^{-1}$ e $K_{\phi,\phi}^{-1}$ são denotados por $\kappa^{r,s} = -\kappa^{rs}$ e $\kappa^{r',s'} = -\kappa^{r's'}$ respectivamente. As estruturas de A e M envolvendo várias matrizes nulas simplificam consideravelmente a obtenção das quantidades A_1 , A_2 e A_3 que determinam o fator de correção da estatística escore. No Apêndice 3, mostramos detalhadamente a obtenção dos A 's. Dos resultados deste apêndice temos

$$\begin{aligned}
 A_1 &= 6 \sum' (\kappa_{i'j'}^{(t)} - 2\kappa_{ti'j'}) (\kappa_{ijs} - 2\kappa_{ij}^{(s)}) \kappa^{ij} \kappa^{st} \kappa^{i'j'} \\
 &\quad - 3 \sum' \left(\sum_{i,j=1}^q \kappa_{ijj'} \kappa^{ij} \right) \left(\sum_{s,t=1}^q \kappa_{stj'} \kappa^{st} \right) \kappa^{i'j'} \\
 &\quad - 6 \sum' \kappa_{jtj'} (3\kappa_{isi'} - 2\kappa_{is}^{(i')}) \kappa^{ij} \kappa^{st} \kappa^{i'j'} \\
 &\quad + 6 \sum' (2\kappa_{iji'j'} - 4\kappa_{ii'j'}^{(j)} - 2\kappa_{ijj'}^{(j')} + \kappa_{ij}^{(i'j')} - 2\kappa_{ii',jj'}) \kappa^{ij} \kappa^{i'j'},
 \end{aligned} \tag{4.7}$$

$$\begin{aligned}
A_2 = \sum' \{ & 3(2\kappa_{ii'j'} - \kappa_{i'j'}^{(i)})(2\kappa_{jk'r'} - \kappa_{k'r'}^{(j)}) + 6\kappa_{ijr'}\kappa_{i',j',k'} \\
& + 6(2\kappa_{ii'k'} - \kappa_{i'k'}^{(i)})(2\kappa_{jj'r'} - \kappa_{j'r'}^{(j)})\} \kappa^{ij} \kappa^{i'j'} \kappa^{k'r'} \\
& + 3 \sum' \kappa_{i',j',k',r'} \kappa^{i'j'} \kappa^{k'r'}
\end{aligned} \quad (4.8)$$

e

$$\begin{aligned}
A_3 = & -3 \sum' \kappa_{i',j',k'} \kappa_{r',s',t'} \kappa^{i'j'} \kappa^{k'r} \kappa^{s't'} \\
& - 2 \sum' \kappa_{i',j',k'} \kappa_{r',s',t'} \kappa^{i'r'} \kappa^{j's'} \kappa^{k't'},
\end{aligned} \quad (4.9)$$

onde $\sum'$ denota o somatório sobre todos os índices envolvidos.

Em geral, é mais fácil calcular $\kappa_{i',j',k'}$ e $\kappa_{i',j',k',r'}$ usando as relações $\kappa_{i',j',k'} = 2\kappa_{i'j'k'} - \sum_{(3)} \kappa_{i'j'}^{(k')}$ e $\kappa_{i',j',k',r'} = -3\kappa_{i'j'k'r'} + 2 \sum_{(4)} \kappa_{i'j'k'}^{(r')} - \sum_{(6)} \kappa_{i'j'}^{(k'r')} + \sum_{(3)} \kappa_{i'j',k'r'}$. Vemos que as expressões em (4.7)–(4.9) dependem somente de esperanças de derivadas da log-verossimilhança até quarta ordem e das covariâncias de $\partial^2 L / \partial \beta_i \partial \phi_{i'}$ e $\partial^2 L / \partial \beta_j \partial \beta_{j'}$ e de $\partial^2 L / \partial \phi_{i'} \partial \phi_{j'}$ e $\partial^2 L / \partial \phi_{k'} \partial \phi_{r'}$. A estatística escore modificada S_R^* para testar $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ e a expansão para os quantis de S_R vêm, respectivamente, de (2.15) e (2.19) com os A 's dados em (4.7)–(4.9) e avaliados sob H_0 .

4.1.4. CASOS ESPECIAIS

4.1.4.1. DOIS PARÂMETROS UNIDIMENSIONAIS ORTOGONAIS

No caso em que os parâmetros ϕ e β são unidimensionais, as expressões para as quantidades A_1 , A_2 e A_3 sofrem grandes simplificações. Para denotarmos cumulantes de derivadas da log-verossimilhança em relação a ϕ e β usamos notação do tipo $\kappa_{\phi\phi} = E(\partial^2 L / \partial \phi^2)$, $\kappa_{\beta\phi\phi} = E(\partial^3 L / \partial \beta \partial \phi^2)$, etc. A estatística escore em (4.4) para testar $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ tem a forma simples

$$S_R = \tilde{u}^2 \tilde{K}_{\phi,\phi}^{-1}, \quad (4.10)$$

onde $u = \partial L / \partial \phi$ é a função escore ^{para ϕ} para ϕ .

Fazendo $i' = j' = k' = r' = s' = t' = \phi$ e $i = j = s = t = \beta$ em (4.7)–(4.9)

chega-se às expressões simples para os A 's

$$A_1 = \frac{6(\kappa_{\phi\phi}^{(\beta)} - 2\kappa_{\beta\phi\phi})(\kappa_{\beta\beta\beta} - 2\kappa_{\beta\beta}^{(\beta)}) - 3\kappa_{\beta\beta\phi}(7\kappa_{\beta\beta\phi} - 4\kappa_{\beta\beta}^{(\phi)})}{\kappa_{\phi\phi}\kappa_{\beta\beta}^2} + \frac{6(2\kappa_{\beta\beta\phi\phi} - 4\kappa_{\beta\phi\phi}^{(\beta)} - 2\kappa_{\beta\beta\phi}^{(\phi)} + \kappa_{\phi\phi}^{(\beta\beta)} - 2\kappa_{\beta\phi,\beta\phi})}{\kappa_{\phi\phi}\kappa_{\beta\beta}}, \quad (4.11)$$

$$A_2 = \frac{3\kappa_{\beta\beta}\kappa_{\phi,\phi,\phi,\phi} + 6\kappa_{\phi,\phi,\phi}\kappa_{\beta\beta\phi} + 9(2\kappa_{\beta\phi\phi} - \kappa_{\phi\phi}^{(\beta)})^2}{\kappa_{\beta\beta}\kappa_{\phi\phi}^2} \quad (4.12)$$

e

$$A_3 = -5 \frac{\kappa_{\phi,\phi,\phi}^2}{\kappa_{\phi\phi}^3}. \quad (4.13)$$

Os cumulantes $\kappa_{\phi,\phi,\phi}$ e $\kappa_{\phi,\phi,\phi,\phi}$ podem ser calculados através das relações $\kappa_{\phi,\phi,\phi} = 2\kappa_{\phi\phi\phi} - 3\kappa_{\phi\phi}^{(\phi)}$ e $\kappa_{\phi,\phi,\phi,\phi} = -3\kappa_{\phi\phi\phi\phi} + 8\kappa_{\phi\phi\phi}^{(\phi)} - 6\kappa_{\phi\phi\phi}^{(\phi\phi)} + 3\kappa_{\phi\phi,\phi\phi}$. Notamos que, enquanto A_1 e A_2 não parecem ter interpretações simples, A_3 é múltiplo do quadrado do coeficiente usual de assimetria da função escore para ϕ .

Apresentamos a seguir algumas aplicações das fórmulas (4.10)–(4.13).

Exemplo 4.1 (continuação)

Retomemos o problema tratado no Exemplo 4.1 e consideremos o teste da hipótese nula $H_0 : \phi = 1$, que equivale à igualdade das médias das duas populações, contra a alternativa $H : \phi \neq 1$.

A log-verossimilhança total em função dos parâmetros ortogonais ϕ e $\beta = \mu\phi^{1/2}$ é

$$L = -n \log \beta - \frac{n\bar{y}_1}{2\beta\phi^{-1/2}} - \frac{n\bar{y}_2}{2\beta\phi^{1/2}}$$

e suas duas primeiras derivadas parciais em relação a ϕ são

$$u_\phi = \frac{\partial L}{\partial \phi} = -\frac{1}{4} \frac{n\bar{y}_1}{\beta\phi^{1/2}} + \frac{1}{4} \frac{n\bar{y}_2}{\beta\phi^{3/2}}$$

e

$$\frac{\partial^2 L}{\partial \phi^2} = \frac{1}{8} \frac{n\bar{y}_1}{\beta\phi^{3/2}} - \frac{3}{8} \frac{n\bar{y}_2}{\beta\phi^{5/2}}.$$

Isto leva a $\kappa_{\phi\phi} = -n/(4\phi^2)$. Agora, de (4.10) vem a estatística escore para testar H_0 versus H

$$S_R = \frac{n(\bar{y}_1 - \bar{y}_2)^2}{4\bar{y}^2}.$$

Note-se que a estatística escore já tinha sido obtida anteriormente (vide Tabela 3.3; modelo gama com $\phi = 1$). Dos resultados daquela tabela sabemos que os A 's, que determinam o teste escore aperfeiçoado, são dados por $A_1 = -12/n$, $A_2 = -18/n$ e $A_3 = 0$. Podemos então, checar parcialmente as fórmulas (4.11)–(4.13) recalculando A_1 , A_2 e A_3 .

Das propriedades da distribuição exponencial obtêm-se sem dificuldade $\kappa_{\phi\phi\phi} = 3n/(4\phi^3)$, $\kappa_{\phi,\phi,\phi} = 0$, $\kappa_{\phi\phi\phi\phi} = -45n/(16\phi^4)$, $\kappa_{\phi\phi,\phi\phi} = 5n/(16\phi^4)$, $\kappa_{\beta\beta} = -n/\beta^2$, $\kappa_{\beta\beta\beta} = 4n/\beta^3$, $\kappa_{\beta\beta\phi} = 0$, $\kappa_{\phi\phi\beta} = n/(4\beta\phi^2)$, $\kappa_{\phi,\phi,\phi,\phi} = 3n/8\phi^4$, $\kappa_{\beta\beta\phi\phi} = -n/(2\beta^2\phi^2)$ e $\kappa_{\phi\phi,\beta\phi} = n/(4\beta^2\phi^2)$. A partir de (4.11)–(4.13) chega-se, após alguma álgebra, aos A 's dados acima.

Exemplo 4.2 (continuação)

A distribuição de Weibull com parâmetro de forma $\phi = 1$ corresponde à distribuição exponencial. Assim sendo, é interessante testar as hipóteses $H_0 : \phi = 1$ versus $H : \phi \neq 1$ no problema descrito no Exemplo 4.2. Inicialmente, encontraremos a estatística escore e os A 's para o teste de $H_0 : \phi = \phi^{(0)}$ e depois estudaremos o caso particular em $\phi^{(0)} = 1$.

A log-verossimilhança total em função dos parâmetros ortogonais ϕ e $\beta = \mu \exp\{\Gamma'(2)/\phi\}$ tem a forma

$$L = n \log \phi - \sum \log y_i + \sum \log z_i - \sum z_i,$$

onde $z_i = z_i(\phi, \beta) = \left(\frac{y_i}{\beta}\right)^\phi \exp\{-\Gamma'(2)\}$. De (4.10) e da função L acima vem

$$S_R = \frac{n\bar{w}^2}{\psi'(1)},$$

$\psi(x) = d \log \Gamma(x)/dx$ é a função digama, $\bar{w} = \sum \tilde{w}_i/n$, com $\tilde{w}_i = (1 + \psi(1))\tilde{z}_i - \tilde{z}_i \log \tilde{z}_i + \log \tilde{z}_i + \psi(1)$. As quantidades assinaladas com um til são avaliadas em $(\phi, \beta) = (\phi^{(0)}, \tilde{\beta})$,

onde $\tilde{\beta}$ é o estimador de máxima verossimilhança de β restrito a H_0 dado pela solução iterativa da equação $\sum \tilde{z}_i = n$.

A obtenção dos cumulantes da log-verossimilhança necessários à obtenção dos A 's é bastante trabalhosa e foi feita com o auxílio do programa REDUCE (Rayna, 1987). Alguns cumulantes foram obtidos por Kahle (1991) que desenvolveu fatores de Bartlett para testes da razão de verossimilhança no modelo Weibull. Obtemos^(*): $\kappa_{\beta\beta} = -n\phi^2/\beta^2$, $\kappa_{\phi\phi} = -n\psi'(1)/\phi^2$, $\kappa_{\phi\phi\beta} = n(\psi'(1) - 1)/(\beta\phi)$, $\kappa_{\beta\beta\phi} = -2n\phi/\beta^2$, $\kappa_{\beta\beta\beta} = n\phi^2(\phi + 3)/\beta^3$, $\kappa_{\phi\phi\beta\beta} = n\{(\psi'(1) - 1)\phi(\phi + 1) + 2\phi^2\}/(\beta^2\phi^2)$, $\kappa_{\phi\phi\phi} = -n\psi''(1)/\phi^3$, $\kappa_{\phi,\phi,\phi} = -2n(\psi''(1) + 3\psi'(1))/\phi^3$, $\kappa_{\phi\phi\phi\phi} = -n\{\psi'''(1) + 3(\psi'(1) - 1)^2\}/\phi^4$, $\kappa_{\phi\phi,\phi\phi} = n\{5\psi'(1)^2 - 10\psi'(1) + 4\psi''(1) + 2\psi'''(1) + 1\}/\phi^4$, $\kappa_{\phi,\phi,\phi,\phi} = 3n\{8\psi'(1)^2 - 4\psi'(1) + 12\psi''(1) + 3\psi'''(1) + 4\}/\phi^4$ e $\kappa_{\beta\phi,\beta\phi} = -n(8\psi(1) - 2\psi'(1) + 11)/\beta^2$. Agora, das fórmulas (4.11)–(4.13) vem

$$A_1 = \frac{24}{n\psi'(1)} \left\{ 4(2 + \psi(1)) - \frac{1}{\phi}(1 - \psi'(1)) \right\},$$

$$A_2 = \frac{3}{n(\psi'(1))^2} (21\psi'(1)^2 - 30\psi'(1) + 28\psi''(1) + 9\psi'''(1) + 9)$$

e

$$A_3 = \frac{20}{n(\psi'(1))^3} (3\psi'(1)^2 + \psi''(1))^2.$$

Note-se que os A 's não dependem do parâmetro de perturbação β .

Em particular, se as hipóteses a serem testadas são $H_0 : \phi = 1$ contra $H : \phi \neq 1$, os A 's acima e a expressão em (2.15) levam à estatística escore modificada

$$S_R^* \cong S_R(1 - b(S_R, n)),$$

onde

$$b(S_R, n) = \frac{1}{n} (4,537 - 1,365S_R + 0,160S_R^2).$$

(*) $\psi(1) = -\gamma$, onde $\gamma \cong -0,577$ é a constante de Euler, $\psi'(1) = \pi^2/6$ e $\psi'''(1) = \pi^4/15$.

As quantidades $\psi'(1)$, $\psi''(1)$ e $\psi'''(1)$ valem aproximadamente, 1,645, -2,404 e 6,494 (Abramowitz e Stegun, 1970).

Aqui as quantidades $\psi(1)$, $\psi'(1)$, $\psi''(1)$ e $\psi'''(1)$ foram substituídas por valores aproximados. O teste escore aperfeiçoado que desenvolvemos no Capítulo 1 baseia-se na estatística S_R^* e na distribuição χ_1^2 de referência. Já o teste escore aperfeiçoado proposto por Harris compara o valor obtido para a estatística S_R com o valor crítico modificado

$$z_\alpha \cong x_\alpha(1 + b(x_\alpha, n)),$$

obtido de (2.19), onde x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$. É interessante notar que, para o teste de $H_0 : \phi = 1$, o estimador de máxima verossimilhança de β sob H_0 é $\tilde{\beta} = \bar{y} \exp\{\Gamma'(2)\}$ e, conseqüentemente, o cálculo de S_R é bastante simples.

Exemplo 4.3

Damos agora aplicações de (4.10)–(4.13) a três distribuições. Consideremos que $y_1, \dots, y_n$ sejam variáveis aleatórias independentes e identicamente distribuídas seguindo um modelo uniparamétrico da família exponencial com média μ e parâmetro de dispersão ϕ . Isto inclui, como visto na Tabela 3.1, a distribuição normal $N(\mu, \phi^{-1})$ com variância ϕ^{-1} , a distribuição normal inversa $N^-(\mu, \phi)$ com parâmetro de precisão ϕ e a distribuição gama $G(\mu, \phi)$ com índice ϕ .

Consideremos primeiramente o teste de $H_0 : \mu = \mu^{(0)}$ versus $H : \mu \neq \mu^{(0)}$ quando a dispersão é desconhecida. A partir de (4.10)–(4.13) encontramos as estatísticas escore aperfeiçoadas da Tabela 4.1. Nesta tabela

$$\tilde{\phi} = \frac{n\mu^{(0)2}}{\sum_{i=1}^n (y_i - \mu^{(0)})^2 / y_i}$$

para a distribuição normal inversa e, para a distribuição gama, $\tilde{\phi}$ é a solução da equação

$$\log(\phi) - \psi(\phi) = \frac{\bar{y} - \mu^{(0)}}{\mu^{(0)}} + \log\left(\frac{\mu^{(0)}}{\tilde{y}}\right),$$

onde $\tilde{y}$ é a média geométrica das observações, isto é, $\tilde{y} = (\prod y_i)^{1/n}$. Para o teste de $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ quando a média é desconhecida as estatísticas escore aperfeiçoadas são exibidas na Tabela 4.2.

Tabela 4.1. Estatísticas escore aperfeiçoadas para o teste da média

$H_0: \mu = \mu^{(0)}$ com dispersão desconhecida

$N(\mu, \phi)$	$N^-(\mu, \phi)$	$G(\mu, \phi)$
<hr/>		
$S_R^* = S_R \left\{ 1 - \frac{1}{2n}(3 - S_R) \right\}$	$S_R^* = S_R \left[1 - \frac{1}{4n} \left\{ 6 - S_R \left(2 + \frac{5\mu^{(0)}}{\tilde{\phi}} \right) + S_R^2 \frac{\mu^{(0)}}{\tilde{\phi}} \right\} \right]$	S_R^* é dado por (2.15) com
onde	onde	$S_R = \frac{n\tilde{\phi}(\bar{y} - \mu^{(0)})^2}{\mu^{(0)2}},$
$S_R = \frac{n^2(\bar{y} - \mu^{(0)})^2}{\sum_{i=1}^n (y_i - \mu^{(0)})^2}$	$S_R = \frac{n(\bar{y} - \mu^{(0)})^2}{\mu^{(0)3}} \tilde{\phi}.$	$A_1 = \frac{6(1 - \tilde{\phi}^2 \tilde{\psi}'' - 2\tilde{\phi} \tilde{\psi}')}{n\tilde{\phi}(1 - \tilde{\phi} \tilde{\psi}')^2},$
		$A_2 = \frac{18}{n\tilde{\phi}} + \frac{9}{n\tilde{\phi}(1 - \tilde{\phi} \tilde{\psi}')},$
		$A_3 = \frac{20}{\tilde{\phi} \Gamma}.$

Nota: $\psi = \psi(\phi) = d \log \Gamma(\phi) / d\phi$ é a função digama e ψ' e ψ'' são suas derivadas. $\tilde{\psi}$, $\tilde{\psi}'$ e $\tilde{\psi}''$ denotam estas funções avaliadas em $\tilde{\phi}$.

Tabela 4.2. Estatísticas escore aperfeiçoadas para o teste da dispersão

$H_0: \phi = \phi^{(0)}$ com média desconhecida

$N(\mu, \phi)$	$N^-(\mu, \phi)$	$G(\mu, \phi)^{(*)}$
$S_R^* = S_R \left\{ 1 - \frac{1}{18n} (33 - 34S_R + 4S_R^2) \right\}$	$S_R^* = S_R \left\{ 1 - \frac{1}{18n} (33 - 34S_R + 4S_R^2) \right\}$	$S_R^* \text{ é dado por (2.15) com}$
<p>onde</p> $S_R = \frac{1}{2n} \left\{ n - \phi^{(0)} \sum_{i=1}^n (y_i - \bar{y})^2 \right\}^2$	<p>onde</p> $S_R = - \frac{n\phi^{(0)} \{ \log \phi^{(0)} - \psi_0 - \log(\bar{y}/\tilde{y}) \}^2}{1 - \phi^{(0)} \psi_0'}$	$A_1 = \frac{3}{n\phi^{(0)}(1 - \phi^{(0)}\psi_0')},$ $A_2 = \frac{3\phi^{(0)}(2\psi_0'' + \phi_0\psi_0''')}{n(1 - \phi^{(0)}\psi_0')^2},$ $A_3 = - \frac{5(1 + \phi^{(0)2}\psi_0'')^2}{n\phi^{(0)}(1 - \phi^{(0)}\psi_0')^3}$

Nota: (*) $\psi_0, \psi_0', \psi_0''$ e ψ_0''' denotam estas funções avaliadas em $\phi = \phi^{(0)}$.

Notamos que para os testes da média e da variância na distribuição normal e da dispersão nas distribuições normal inversa e gama, S_R^* não envolve parâmetros desconhecidos. Notamos também que as formas funcionais de S_R^* para testar a dispersão nos modelos normal e normal inverso são iguais.

O teste aperfeiçoado da média no modelo gama envolve a função digama $\psi(\phi)$ e suas duas primeiras derivadas, enquanto que o teste da dispersão envolve também a terceira derivada de $\psi(\phi)$. Tabelas que fornecem os valores de tais funções avaliadas em $1 \leq x \leq 2$ são encontradas no livro de Abramowitz e Stegun (1970; pgs.267-271). A relação de recorrência $\psi^{(n)}(x+1) = \psi^{(n)}(x) + (-1)^n n! x^{-n-1}$ pode ser usada para avaliar $\psi(x)$ e suas derivadas fora deste intervalo. Fórmulas aproximadas para os A 's são obtidas a partir da expansão (Abramowitz e Stegun, 1970; fórmula 6.4.12)

$$\psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \frac{1}{42x^7} - \frac{1}{30x^9} + \dots$$

que conduz a $A_1 \cong 4/(n\phi)$ e $A_2 \cong -6(3 - 4/\phi)/n$ para o teste da média (note-se que $A_3 = 20/(n\phi)$) e a $A_1 \cong -6\{1 - 1/(3\phi)\}/n$, $A_2 \cong 12\{1 + 1/(3\phi)\}/n$ e $A_3 \cong 40(1 - 3/\phi)/n$ para o teste da dispersão. Nestas aproximações termos de ordem inferior a ϕ^{-1} foram desprezados.

Podemos, em alguns casos, verificar que as expressões da estatística modificada S_R^* nas Tabelas 4.1 e 4.2 estão corretas encontrando A_1 , A_2 e A_3 a partir dos cálculos de momentos de S_R . Consideremos, inicialmente, o teste da média em um modelo normal. É fácil ver que

$$S_R = n \frac{T_1/T_2}{1 + T_1/T_2}$$

onde $T_1 = n\phi(\bar{y} - \mu^{(0)})^2$ e $T_2 = \phi \sum (y_i - \bar{y})^2$. Mas T_1 e T_2 são independentes e têm distribuições χ_1^2 e χ_{n-1}^2 respectivamente. Assim, nota-se que S_R tem distribuição beta com parâmetros $1/2$ e $(n-1)/2$, de onde seguem os três primeiros momentos de S_R até ordem n^{-1} : $\mu'_1 = 1$, $\mu'_2 = 3 - 6/n$ e $\mu'_3 = 15 - 90/n$. Substituindo estes momentos em (2.10)

vem $A_1 = 0$, $A_2 = -18/n$ e $A_3 = 0$. A partir dos A 's e de (2.15) chega-se à expressão de S_R^* dada na Tabela 4.1 para o modelo normal.

Para o teste de $H_0 : \phi = \phi^{(0)}$, ainda no modelo normal, a estatística S_R pode ser expressa como $S_R = (n - T_2)^2 / (2n)$ com T_2 definido acima, de onde seguem os três primeiros momentos, até ordem n^{-1} : $\mu'_1 = 1 - 1/(2n)$, $\mu'_2 = 3 + 1/n$ e $\mu'_3 = 15 + 235/(2n)$. Agora, substituindo μ'_1 , μ'_2 e μ'_3 em (2.10) chega-se a $A_1 = -6/n$, $A_2 = 12/n$ e $A_3 = 40/n$ que conduzem à expressão de S_R^* dada na Tabela 4.2.

Finalmente, para o teste de $H_0 : \phi = \phi^{(0)}$ no modelo gama, é possível verificar que a expressão de S_R^* está correta utilizando a função geradora de cumulantes de $\tilde{y}/\tilde{y}$ (Bartlett, 1937) $\log(\tilde{y}/\tilde{y})$

$$K(x) = -x \log n + n \log \Gamma(\phi - x/n) - \log \Gamma(n\phi - x) - n \log \Gamma(\phi) + \log \Gamma(n\phi).$$

A partir de $K(x)$ e expandindo a função digama e suas derivadas, avaliadas em $n\phi$, em série de Taylor pode-se mostrar que S_R^* está correta. Este procedimento envolve um trabalho algébrico enorme e, portanto, será omitido aqui.

4.1.4.2. PARÂMETRO DE INTERESSE UNIDIMENSIONAL

Nesta seção supomos que o parâmetro de interesse é unidimensional enquanto que o de perturbação tem dimensão $q \geq 1$. O fator de correção de Bartlett da estatística da razão de verossimilhança para este problema foi obtido por Cordeiro (1986a).

A notação usada para denotar os cumulantes envolvidos nas expressões dos A 's é a seguinte: $\kappa_{\phi ij} = E(\partial^3 L / \partial \phi \partial \beta_i \partial \beta_j)$, $\kappa_{\phi, ij} = E(\partial L / \partial \phi \partial^2 L / \partial \beta_i \partial \beta_j)$, etc.

Fazendo $i' = j' = k' = r' = s' = t' = \phi$ em (4.7)–(4.9) vem

$$\begin{aligned}
A_1 = \frac{1}{\kappa_{\phi\phi}} & \left\{ 6 \sum' (\kappa_{\phi\phi}^{(t)} - 2\kappa_{t\phi\phi})(\kappa_{ijs} - 2\kappa_{ij}^{(s)})\kappa^{ij}\kappa^{st} \right. \\
& - 3 \left(\sum' \kappa_{ij\phi}\kappa^{ij} \right)^2 - 6 \sum' \kappa_{jt\phi}(3\kappa_{is\phi} - 2\kappa_{is}^{(\phi)})\kappa^{ij}\kappa^{st} \\
& \left. + 6 \sum' (2\kappa_{ij\phi\phi} - 4\kappa_{i\phi\phi}^{(j)} - 2\kappa_{ij\phi}^{(\phi)} + \kappa_{\phi\phi}^{(ij)} - 2\kappa_{i\phi,j\phi})\kappa^{ij} \right\},
\end{aligned} \tag{4.14}$$

$$A_2 = \frac{1}{\kappa_{\phi\phi}^2} \left[\sum' \{ 9(2\kappa_{i\phi\phi} - \kappa_{\phi\phi}^{(i)})(2\kappa_{j\phi\phi} - \kappa_{\phi\phi}^{(j)}) + 6\kappa_{\phi\phi\phi}\kappa_{ij\phi} \} \kappa^{ij} + 3\kappa_{\phi,\phi,\phi,\phi} \right] \tag{4.15}$$

e

$$A_3 = -5 \frac{\kappa_{\phi,\phi,\phi}^2}{\kappa_{\phi\phi}^3} \tag{4.16}$$

Notamos que A_3 coincide com (4.13) enquanto que A_1 e A_2 igualam, respectivamente, (4.11) e (4.12) se tomarmos $i = j = s = t = \beta$.

No Capítulo 5, apresentamos uma aplicação importante das fórmulas acima a testes sobre o parâmetro de dispersão de modelos lineares generalizados com parâmetros canônicos ϕ e $\phi\theta$ bem como de modelos próprios de dispersão.

4.1.4.3. PARÂMETRO DE PERTURBAÇÃO UNIDIMENSIONAL

Supomos agora que o parâmetro de perturbação é unidimensional enquanto que o de interesse tem dimensão arbitrária $p \geq 1$. Os cumulantes κ 's são denotados por $\kappa_{\beta i' j'} = E(\partial^3 L / \partial \beta \partial \phi_{i'} \partial \phi_{j'})$, $\kappa_{\beta, i' j'} = E(\partial L / \partial \beta \partial^2 L / \partial \phi_{i'} \partial \phi_{j'})$, etc.

Fazendo $i = j = s = t = \beta$ em (4.7)–(4.9) segue-se que

$$\begin{aligned}
A_1 = \frac{1}{\kappa_{\beta\beta}^2} & \left[6 \sum' \left\{ (\kappa_{i' j'}^{(\beta)} - 2\kappa_{\beta i' j'}) (\kappa_{\beta\beta\beta} - 2\kappa_{\beta\beta}^{(\beta)}) \right. \right. \\
& \left. \left. - 3\kappa_{\beta\beta i'} \kappa_{\beta\beta j'} - 6\kappa_{\beta\beta j'} (3\kappa_{\beta\beta i'} - 2\kappa_{\beta\beta}^{(i')}) \right\} \kappa^{i' j'} \right] \\
& + \frac{6}{\kappa_{\beta\beta}} \sum' (2\kappa_{\beta\beta i' j'} - 4\kappa_{\beta i' j'}^{(\beta)} - 2\kappa_{\beta\beta i'}^{(j')} + \kappa_{\beta\beta}^{(i' j')} - 2\kappa_{\beta i', \beta j'}) \kappa^{i' j'},
\end{aligned} \tag{4.17}$$

$$A_2 = \frac{1}{\kappa_{\beta\beta}} \left[3 \sum_{i',j',k',r'} \{ (2\kappa_{\beta i'j'} - \kappa_{i'j'}^{(\beta)}) (2\kappa_{\beta k'r'} - \kappa_{k'r'}^{(\beta)}) + 6\kappa_{i',j',k'} \kappa_{\beta\beta r'} \right. \\ \left. + 6(2\kappa_{\beta i'k'} - \kappa_{i'k'}^{(\beta)}) (2\kappa_{\beta j'r'} - \kappa_{j'r'}^{(\beta)}) + 3\kappa_{i',j',k',r'} \} \kappa_{i'j'}^{K_P\beta} \kappa_{k'r'}^{K_P\beta} \right] \quad (4.18)$$

e A_3 coincide com (4.9). É fácil ver que se tomarmos $i' = j' = k' = r' = s' = t' = \phi$ as expressões de A_1 , A_2 e A_3 coincidem com (4.11), (4.12) e (4.13) respectivamente.

Como aplicação destas fórmulas desenvolvemos no Capítulo 5 testes escore aperfeiçoados para o conjunto completo dos parâmetros da regressão em modelos lineares generalizados e modelos próprios de dispersão, ambos com parâmetro de dispersão desconhecido.

4.2. PROBLEMAS COM PARÂMETROS LOCALMENTE ORTOGONAIS

Consideremos a log-verossimilhança de um problema regular dependendo do parâmetro $\beta = (\beta_1, \dots, \beta_p)^T$ desconhecido. Assumindo que a matriz de informação total de Fisher K é positiva definida, temos que existe uma matriz não singular $p \times p$ de constantes desconhecidas tal que $P^T K^{-1} P = n^{-1} I_p$, sendo K e P avaliadas nos verdadeiros valores dos parâmetros e onde I_p denota uma matriz identidade $p \times p$ (Cordeiro, 1987). Então, a matriz de informação do vetor paramétrico $\beta^* = P^T \beta$ calculada nos verdadeiros valores dos parâmetros é n vezes a matriz identidade. Dizemos, neste caso, que $\beta_1^*, \dots, \beta_p^*$ são localmente ortogonais nos verdadeiros valores dos parâmetros. Desta discussão vemos que é sempre possível encontrar uma transformação de parâmetros de forma a obter ortogonalização local. Técnicas para obter a matriz P que utilizam a decomposição de Choleski podem ser encontradas nos livros de Chambers (1977) e Healy (1986).

Particionamos o vetor β^* em $(\beta_1^{*T}, \beta_2^{*T})^T$ sendo $\beta_1^* = (\beta_1^*, \dots, \beta_r^*)^T$ e $\beta_2^* = (\beta_{r+1}^*, \dots, \beta_p^*)^T$ e consideramos o teste das hipóteses $H_0 : \beta_1^* = \beta_1^{*(0)}$ versus $H : \beta_1^* \neq \beta_1^{*(0)}$, onde $\beta_1^{*(0)}$ é um vetor de dimensão r de valores especificados para β_1^* . A notação utilizada aqui é semelhante à das seções anteriores. Assim, por exemplo, $\kappa_{r,st} = E(\partial L / \partial \beta_r^* \partial^2 L / \partial \beta_s^* \partial \beta_t^*)$.

Como $K = nI_p$ temos de (2.9)

$$A = \frac{1}{n} \begin{pmatrix} 0 & 0 \\ 0 & I_{p-r} \end{pmatrix} \quad \text{e} \quad M = K^{-1} - A = \frac{1}{n} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

Sejam $T_0 = \{1, \dots, r\}$ e $T = \{1, \dots, p\}$. De (2.6) e das matrizes acima vem

$$\begin{aligned} A_1 = \frac{1}{n^3} & \left\{ 3 \sum_{\substack{i,s \in T-T_0 \\ k \in T_0}} (\kappa_{iik} + 2\kappa_{i,ik})(\kappa_{kss} + 2\kappa_{k,s,s}) - 6 \sum_{\substack{i,k \in T-T_0 \\ s \in T_0}} (\kappa_{iik} + 2\kappa_{i,ik})\kappa_{k,s,s} \right. \\ & \left. + 6 \sum_{\substack{j,k \in T-T_0 \\ i \in T_0}} (\kappa_{i,jk} - \kappa_{i,j,k})(\kappa_{ijk} + 2\kappa_{i,j,k}) - 6 \sum_{\substack{k \in T-T_0 \\ i \in T_0}} (\kappa_{i,i,k,k} + \kappa_{i,i,kk}) \right\} \end{aligned}$$

Definimos '+' e '.' como somatório sobre T_0 e $T-T_0$ respectivamente. Assim, por exemplo,

$$\sum_{k \in T-T_0} \kappa_{k,k,i} = \kappa_{\cdot,\cdot,i} \quad \text{e} \quad \sum_{i \in T_0} \kappa_{k,i,i} = \kappa_{k,+,+}.$$

Desta forma, A_1 se reduz a

$$\begin{aligned} A_1 = \frac{1}{n^3} & \left\{ 3 \sum_{k \in T_0} (\kappa_{\cdot\cdot k} + 2\kappa_{\cdot,\cdot,k})^2 - 6 \sum_{j \in T-T_0} (\kappa_{\cdot\cdot j} + 2\kappa_{\cdot,\cdot,j})\kappa_{j,+,+} \right. \\ & \left. + 6 \sum_{\substack{i,j \in T-T_0 \\ k \in T_0}} (\kappa_{i,j,k} - \kappa_{i,j,k})(\kappa_{ijk} + 2\kappa_{i,j,k}) - 6 \sum_{\substack{k \in T-T_0 \\ i \in T_0}} (\kappa_{i,i,k,k} + \kappa_{i,i,kk}) \right\}. \end{aligned} \quad (4.19)$$

De maneira análoga, chega-se a

$$\begin{aligned} A_2 = \frac{1}{n^3} & \left\{ 3 \sum_{i,j \in T_0} \kappa_{i,i,j,j} - 6 \sum_{\substack{i,j \in T_0 \\ k \in T-T_0}} \kappa_{i,i,j,k}^2 \right. \\ & \left. - 3 \sum_{k \in T-T_0} \kappa_{+,+,k}^2 - 6 \sum_{j \in T_0} (\kappa_{\cdot,\cdot,j} + \kappa_{\cdot,\cdot,j})\kappa_{+,+,j} \right\} \end{aligned} \quad (4.20)$$

e

$$A_3 = \frac{1}{n^3} \left\{ 3 \sum_{k \in T_0} \kappa_{k,+,+}^2 + 2 \sum_{i,j,k \in T_0} \kappa_{i,i,j,k}^2 \right\}. \quad (4.21)$$

É importante ressaltar que estas fórmulas para os A 's são válidas somente no caso de ortogonalização local de parâmetros considerada aqui, ou seja, na qual a matriz de informação total de Fisher é n vezes a matriz identidade. É claro que estas não são válidas

no caso de parâmetros globalmente ortogonais tratado na Seção 4.1. Vejamos: naquele caso $\kappa_{\phi\beta}^{(\phi)} = \kappa_{\phi\beta}^{(\beta)} = 0$ mas estas igualdades não são necessariamente verdadeiras no contexto das fórmulas acima pois, sob a ortogonalização local considerada aqui, $\kappa_{\phi\beta} = 0$ somente nos verdadeiros valores dos parâmetros. A grande vantagem das fórmulas (4.19)–(4.21) é que a ortogonalização local é sempre possível de ser obtida enquanto que a global não, sendo garantida apenas quando pelo menos um dos parâmetros (de interesse ou de perturbação) for unidimensional (Cox e Reid, 1987b).

4.3. FÓRMULAS MATRICIAIS PARA O FATOR DE CORREÇÃO DA ESTATÍSTICA SCORE

4.3.1. INTRODUÇÃO

Nas seções seguintes, apresentamos fórmulas matriciais para calcular fatores de correção para a estatística score no teste de uma hipótese geral simples ou composta. Mostramos que as fórmulas matriciais têm vantagem computacional em relação à notação tensorial da expansão assintótica de Harris (1985) para a distribuição da estatística score. Combinando essas fórmulas com o uso de um sistema algébrico computadorizado como o REDUCE (Rayna, 1987) tem-se uma maneira simples de se calcular a expansão assintótica da estatística score. Numericamente, estas fórmulas também podem ser avaliadas via o programa GAUSS.

4.3.2. FÓRMULAS MATRICIAIS PARA OS COEFICIENTES

Consideremos o vetor $y = (y_1, \dots, y_n)^T$ de n observações independentes cuja distribuição depende de um vetor de parâmetros $\beta = (\beta_1^T, \beta_2^T)^T$ sendo $\beta_1 = (\beta_1, \dots, \beta_q)^T$ o parâmetro de interesse q -dimensional e $\beta_2 = (\beta_{q+1}, \dots, \beta_p)^T$ o parâmetro de perturbação $(p-r)$ -dimensional. O objetivo é testar a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$ contra a alternativa

$H : \beta_1 \neq \beta_1^{(0)}$. Os testes escore aperfeiçoados de H_0 versus H sugeridos no Capítulo 2 dependem das quantidades A_1 , A_2 e A_3 dadas por Harris (1985) e que são apresentadas, em notação tensorial, em (2.6)–(2.8). A seguir apresentaremos expressões matriciais simples para os A 's que podem ser facilmente implementados em sistemas computadorizados (como o REDUCE, Rayna, 1987), pois envolvem somente operações de multiplicação e soma de matrizes e vetores. O número total de matrizes envolvidas nos cálculos é $2p^2 + 3p + 13$ sendo $2p^2 + 3p + 6$ matrizes $p \times p$, 2 matrizes $p \times p^2$, 3 matrizes $p^2 \times p$ e 2 matrizes $p^2 \times p^2$. Note-se que não estamos fazendo nenhuma suposição sobre ortogonalização dos parâmetros β_1 e β_2 , de maneira que as fórmulas matriciais para os A 's desenvolvidas aqui são bastante gerais e podem, evidentemente, ser adaptadas para o caso de ortogonalização, local ou global, de parâmetros.

Denotamos por A_{lm} a m -ésima parcela de A_l . Definimos $b_{jk}^{(i)} = \kappa_{ijk} + 2\kappa_{i,jk}$ como o elemento (j, k) da matriz $B^{(i)}$. Nota-se que $B^{(i)}$ é uma matriz $p \times p$ simétrica. Assim, como $A_{11} = 3 \sum' (\kappa_{ijk} + 2\kappa_{i,jk})(\kappa_{rst} + 2\kappa_{rs,t}) a_{ij} a_{st} m_{kr}$ vem

$$A_{11} = 3 \sum' a_{ij} b_{jk}^{(i)} m_{kr} b_{rs}^{(t)} a_{st},$$

ou ainda,

$$A_{11} = 3 \sum_{i,t} a_i^T B^{(i)} M B^{(t)} a_t,$$

onde a_i^T é a i -ésima linha da matriz A definida em (2.9). Agora é fácil ver que

$$A_{11} = 3 \mathbf{1}^T F M F^T \mathbf{1},$$

onde F é uma matriz $p \times p$ cuja i -ésima linha é $a_i^T B^{(i)}$ e $\mathbf{1}$ é um vetor $p \times 1$ de uns.

Para a obtenção de A_{12} , definimos $c_{jk}^{(i)} = \kappa_{i,jk}$ como o elemento (j, k) da matriz $C^{(i)}$, simétrica e de dimensão $p \times p$. Já que $A_{12} = -6 \sum' (\kappa_{ijk} + 2\kappa_{i,jk}) \kappa_{rs,t} a_{ij} a_{kr} m_{st}$, vem

$$A_{12} = -6 \sum' a_{ij} b_{jk}^{(i)} a_{kr} c_{rs}^{(t)} m_{st},$$

ou seja,

$$A_{12} = -6 \sum_{i,t} a_i^T B^{(i)} A C^{(t)} m_t,$$

sendo m_t^T a t -ésima linha da matriz M definida em (2.9). A expressão acima conduz a

$$A_{12} = -6 \mathbf{1}^T F A G^T \mathbf{1},$$

onde G é uma matriz $p \times p$ cuja i -ésima linha é $m_i^T C^{(i)}$.

Definimos $d_{jk}^{(i)} = \kappa_{ijk} - \kappa_{i,j,k}$ como o elemento (j, k) da matriz $D^{(i)}$, simétrica e de dimensão $p \times p$. De $A_{13} = 6 \sum' (\kappa_{i,jk} - \kappa_{i,j,k})(\kappa_{rst} + 2\kappa_{rs,t}) a_{js} a_{kt} m_{ir}$ segue que

$$A_{13} = 6 \sum' m_{ir} b_{rs}^{(t)} a_{sj} d_{jk}^{(i)} a_{kt},$$

ou

$$A_{13} = 6 \sum_{i,t} m_i^T B^{(t)} A D^{(i)} a_t.$$

Sejam

$$J = \begin{pmatrix} m_1^T B^{(1)} & \dots & m_1^T B^{(p)} \\ \vdots & \ddots & \vdots \\ m_p^T B^{(1)} & \dots & m_p^T B^{(p)} \end{pmatrix}$$

e

$$H = \begin{pmatrix} D^{(1)} a_1 & \dots & D^{(p)} a_1 \\ \vdots & \ddots & \vdots \\ D^{(1)} a_p & \dots & D^{(p)} a_p \end{pmatrix}$$

matrizes $p \times p^2$ e $p^2 \times p$, respectivamente, e

$$\tilde{A} = \begin{pmatrix} A & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A \end{pmatrix}$$

uma matriz $p^2 \times p^2$ bloco diagonal. Observa-se que o i -ésimo elemento da diagonal da matriz $J \tilde{A} H$ é $\sum_t m_i^T B^{(t)} A D^{(i)} a_t$ e, portanto,

$$A_{13} = 6 \text{tr}(J \tilde{A} H),$$

onde tr indica o traço da matriz.

Para a obtenção da última parcela de A_1 , definimos o elemento (i, j) da matriz $E^{(k,r)}$, simétrica de dimensão $p \times p$ como $e_{ij}^{(k,r)} = \kappa_{i,j,k,r} + \kappa_{i,j,kr}$. Como $A_{14} = -6 \sum' (\kappa_{i,j,k,r} + \kappa_{i,j,kr}) a_{kr} m_{ij}$ temos

$$A_{14} = -6 \sum_{k,r} a_{kr} \sum_{i,j} m_{ij} e_{ij}^{(k,r)}$$

que conduz a

$$A_{14} = -6 \sum_{k,r} a_{kr} \text{tr}(M E^{(k,r)}).$$

Definindo $l_{ij} = a_{ij} \text{tr}(M E^{(i,j)})$ como o elemento (i, j) da matriz L , simétrica de dimensão $p \times p$, temos

$$A_{14} = -6 \mathbf{1}^T L \mathbf{1}.$$

Tem-se então,

$$A_1 = 3 \mathbf{1}^T F M F^T \mathbf{1} - 6 \mathbf{1}^T F A G^T \mathbf{1} + 6 \text{tr}(J \tilde{A} H) - 6 \mathbf{1}^T L \mathbf{1}, \quad (4.22)$$

onde as matrizes envolvidas foram definidas acima.

A obtenção de A_2 e A_3 em forma matricial a partir das expressões de A_2 e A_3 em (2.7) e (2.8) é semelhante à de A_1 . Temos

$$A_2 = -3 \mathbf{1}^T G A G^T \mathbf{1} + 6 \mathbf{1}^T F M G^T \mathbf{1} - 6 \text{tr}(P \tilde{M} S) + 3 \mathbf{1}^T R \mathbf{1}, \quad (4.23)$$

onde

$$P = \begin{pmatrix} m_1^T C^{(1)} & \dots & m_1^T C^{(p)} \\ \vdots & \ddots & \vdots \\ m_p^T C^{(1)} & \dots & m_p^T C^{(p)} \end{pmatrix}$$

e

$$S = \begin{pmatrix} C^{(1)} a_1 & \dots & C^{(p)} a_1 \\ \vdots & \ddots & \vdots \\ C^{(1)} a_p & \dots & C^{(p)} a_p \end{pmatrix}$$

são matrizes de ordens $p \times p^2$ e $p^2 \times p$, respectivamente,

$$\tilde{M} = \begin{pmatrix} M & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \end{pmatrix}$$

é uma matriz $p^2 \times p^2$ bloco diagonal e R é uma matriz simétrica de dimensão $p \times p$ cujo elemento (k, r) é $r_{kr} = m_{kr} \text{tr}(MQ^{(k,r)})$ com $Q^{(k,r)}$ sendo uma matriz também simétrica e de ordem p com elemento (i, j) igual a $q_{ij}^{(k,r)} = \kappa_{i,j,k,r}$. As demais matrizes envolvidas foram definidas anteriormente. Finalmente, temos

$$A_3 = 3 \mathbf{1}^T G M G^T \mathbf{1} + 2 \text{tr}(P \widetilde{M} T), \quad (4.24)$$

onde

$$T = \begin{pmatrix} C^{(1)}m_1 & \dots & C^{(p)}m_1 \\ \vdots & \ddots & \vdots \\ C^{(1)}m_p & \dots & C^{(p)}m_p \end{pmatrix}.$$

A seguir fazemos uma análise do custo computacional de obtenção de A_1 , A_2 e A_3 através das fórmulas matriciais. Definimos este custo como o número total de produtos envolvidos no cálculo destas quantidades.

Sejam A e B duas matrizes de dimensão $p \times p$. Note que o custo do produto de A por B é p^3 , o de seu traço é p^2 e que $\mathbf{1}^T A \mathbf{1} = \sum_{i,j} a_{ij}$, onde a_{ij} é o elemento (i, j) da matriz A , tem custo nulo.

Supomos que as parcelas envolvidas em A_1 , A_2 e A_3 são calculadas sequencialmente na ordem em que elas aparecem nas expressões (4.22)–(4.24). Desta forma, o cálculo do custo de uma determinada parcela levará em conta o fato de que as parcelas anteriores já foram calculadas e, portanto, que os produtos já efetuados podem ser aproveitados sem custo adicional.

Analisemos em detalhes o custo da primeira parcela de A_1 , ou seja, $A_{11} = 3 \mathbf{1}^T F M F^T \mathbf{1}$. O custo de F é p vezes o custo do produto de a_1^T por $B^{(1)}$, ou seja, é igual a p^3 . Agora, os custos dos produtos de F por M e de FM por F^T são também iguais a p^3 . No cálculo de A_{11} é envolvido ainda mais um produto, o de $\mathbf{1}^T F M F^T \mathbf{1}$ por 3. Logo, o custo de A_{11} é igual a $3p^3 + 1$.

O custo de $A_{12} = -6 \mathbf{1}^T F A G^T \mathbf{1}$ é calculado de forma análoga ao de A_{11} , lembrando que a matriz F já foi obtida e, portanto, que o custo de sua obtenção aqui é nulo. Assim, temos que o custo de A_{12} é $2p^3 + 1$.

Notemos agora que o i -ésimo elemento da diagonal de $J\tilde{A}H$ é $\sum_t m_i^T B^{(t)} A D^{(i)} a_t$ e que o custo de $A_{13} = 6\text{tr}(J\tilde{A}H)$ é dado por 1 mais p vezes o custo deste elemento, que envolve os custos dos produtos de m_i^T por $B^{(t)}$, de $m_i^T B^{(t)}$ por A , de $D^{(i)}$ por a_t e de $m_i^T B^{(t)} A$ por $D^{(i)} a_t$. Assim, temos que o custo de A_{13} é igual a $3p^4 + p^3 + 1$.

Finalmente, o custo de $A_{14} = -6 \mathbf{1}^T L \mathbf{1}$ é dado por 1 mais p^2 vezes o custo de $l_{ij} = a_{ij}\text{tr}(ME^{(i,j)})$, ou seja, é igual a $p^4 + p^2 + 1$. Temos então que o custo de A_1 é $4p^4 + 6p^3 + p^2 + 4$.

Os custos de A_2 e A_3 são obtidos de maneira semelhante ao de A_1 e são dados, respectivamente, por $3p^4 + 4p^3 + p^2 + 4$ e $3p^3 + 2$, de forma que o custo total da obtenção de A_1 , A_2 e A_3 através das fórmulas matriciais é $7p^4 + 13p^3 + 2p^2 + 10$, isto é, de ordem p^4 . O custo total a partir das expressões do Harris (1985) para os A 's é de ordem p^6 .

O uso das fórmulas matriciais em conexão com um sistema computadorizado algébrico como o REDUCE pode ser de grande utilidade na obtenção de expressões fechadas para as quantidades A_1 , A_2 e A_3 . Um programa desenvolvido no REDUCE montaria as matrizes a partir das seguintes informações fornecidas pelo usuário: (i) os valores de p e r que correspondem respectivamente, ao número total de parâmetros e daqueles especificados na hipótese nula e (ii) as expressões algébricas dos cumulantes envolvidos nas expressões dos A 's. Uma vez que o REDUCE permite operações com matrizes e vetores, o programa obteria expressões algébricas para os A 's a partir de (4.22)–(4.24). É claro que se o objetivo for simplesmente obter os valores numéricos para os A 's, o programa seria capaz de calculá-los desde que o usuário fornecesse as estimativas de máxima verossimilhança sob H_0 dos parâmetros que, eventualmente, sejam envolvidos nos cumulantes.

Jensen (1991) desenvolveu um programa no REDUCE para a obtenção do fator de correção de Bartlett para a estatística da razão de verossimilhança. Acreditamos ser possível desenvolver um programa mais simples a partir da fórmula matricial para este fator obtida por Cordeiro (1990).

Programas computacionais para obtenção de expressões algébricas para os A 's

serão desenvolvidos em trabalho futuro.

4.3.3. UMA APLICAÇÃO

Consideremos n variáveis aleatórias $y_i, i = 1, \dots, n$, com distribuição exponencial com média $\mu_i = \exp\{\alpha + \beta(x_i - \bar{x})\}$, onde x_i é uma variável auxiliar fixada e $\bar{x} = n^{-1} \sum x_i$. Digamos que o interesse seja testar a hipótese nula $H_0 : \beta = 0$ versus a alternativa $H : \beta \neq 0$. No Capítulo 3 (Seção 3.8.1) mostramos que a estatística escore e os A 's para este problema são dados, respectivamente, por $S_R = n\bar{s}^2/(\bar{y}^2\bar{s}_2)$, $A_1 = -12/n$, $A_2 = -18(2 - \bar{s}_4/\bar{s}_2^2)/n$ e $A_3 = 20\bar{s}_3^2/(n\bar{s}_2^3)$, onde $\bar{s}_a = n^{-1} \sum (x_i - \bar{x})^a$, $a = 2, 3, 4$ e $\bar{s} = n^{-1}(\sum x_i y_i - n\bar{x}\bar{y})$.

A seguir mostraremos que A_1, A_2 e A_3 podem ser obtidos das expressões matriciais (4.22)–(4.24) sem qualquer dificuldade.

O logaritmo da função de verossimilhança de α e β é dado por

$$L = -n\alpha - \exp(-\alpha) \sum y_i \exp\{-\beta(x_i - \bar{x})\}.$$

Devemos montar 21 matrizes: 14 de dimensão 2×2 , 2 de dimensão 2×4 , 3 de dimensão 4×2 e 2 de dimensão 4×4 . Através das propriedades da distribuição exponencial, os cumulantes do logaritmo da função de verossimilhança são calculados e estes conduzem a:

$$\begin{aligned} A &= \frac{1}{n} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad M = \frac{1}{n} \begin{pmatrix} \bar{s}_2^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad B^{(1)} = -n \begin{pmatrix} \bar{s}_3 & \bar{s}_2 \\ \bar{s}_2 & 0 \end{pmatrix}, \\ B^{(2)} &= -n \begin{pmatrix} \bar{s}_2 & 0 \\ 0 & 1 \end{pmatrix}, \quad F = - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad G = 2 \begin{pmatrix} \bar{s}_3/\bar{s}_2 & 1 \\ 0 & 0 \end{pmatrix}, \\ J &= - \begin{pmatrix} \bar{s}_3/\bar{s}_2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad E^{(2,2)} = 4n \begin{pmatrix} \bar{s}_2 & 0 \\ 0 & 1 \end{pmatrix}, \quad L = \frac{4}{n} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \\ P &= 2 \begin{pmatrix} \bar{s}_3/\bar{s}_2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q^{(1,1)} = 6n \begin{pmatrix} \bar{s}_4 & \bar{s}_3 \\ \bar{s}_3 & \bar{s}_2 \end{pmatrix}, \quad R = 6 \begin{pmatrix} \bar{s}_4/\bar{s}_2^2 & 0 \\ 0 & 0 \end{pmatrix}, \\ H &= -3 \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad S = 2 \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \bar{s}_2 & 0 \\ 0 & 1 \end{pmatrix}, \quad T = 2 \begin{pmatrix} \bar{s}_3/\bar{s}_2 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

$$\widetilde{M} = \frac{1}{n} \begin{pmatrix} \bar{s}_2^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{s}_2^{-1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \widetilde{A} = \frac{1}{n} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$C^{(i)} = -2B^{(i)} \quad \text{e} \quad D^{(i)} = 3B^{(i)}, \quad i = 1, 2.$$

Note-se que as matrizes $E^{(1,1)}$, $E^{(1,2)}$, $Q^{(1,2)}$, $Q^{(2,1)}$ e $Q^{(2,2)}$ não precisam ser calculadas, pois estas serão multiplicadas por elementos de M ou A que valem zero.

De (4.22)–(4.24) e das matrizes acima mostra-se facilmente que as expressões para A_1 , A_2 e A_3 coincidem com as dadas anteriormente.

4.4. ALGUMAS PROPRIEDADES DA ESTATÍSTICA ESCORE APERFEIÇOADA

A discussão sobre ortogonalização local de parâmetros apresentada na Seção 4.2 é bastante útil para analisarmos algumas propriedades da estatística escore aperfeiçoada S_R^* dada em (2.15). Da expressão (4.21) vemos que, se $K = nI_p$ quando avaliada nos verdadeiros valores dos parâmetros, tem-se $A_3 \geq 0$. Mas, como já mencionamos, é sempre possível obter uma matriz P não singular tal que a matriz de informação total de Fisher de $P^T\beta$ avaliada nos verdadeiros valores dos parâmetros seja $K := nI_p$. Logo, sem perda de generalidade, podemos assumir que os parâmetros foram escolhidos de tal forma que $K = nI_p$ e, portanto, $A_3 \geq 0$. Esta argumentação foi utilizada por Cordeiro (1987; pg.268) para mostrar que um certo coeficiente da expansão assintótica da distribuição da estatística da razão de verossimilhança (Hayakawa, 1977) vale zero sob condições gerais de regularidade. O fato de A_3 ser não negativa é importante para a análise que se segue.

Algumas propriedades da estatística escore modificada S_R^* em (2.15) merecem ser apontadas. Em primeiro lugar, notamos que, enquanto S_R é uma variável aleatória não negativa, S_R^* poderá assumir, em pequenas amostras, valores menores do que zero. Outro aspecto a ser mencionado é que S_R^* não é necessariamente uma função crescente de

S_R . Salientamos que a estatística da razão de verossimilhança modificada é uma função crescente da estatística sem correção.

Analisaremos o fato de que S_R^* pode assumir valores negativos. Lembremos que os A 's são de ordem n^{-1} e suponhamos, inicialmente, que $A_3 = 0$. Se $A_2 = 0$, temos, de (2.15) que $S_R^* \geq 0$ se e somente se $1 - A_1/(12q) \geq 0$. Se $A_1 \leq 0$, esta condição está satisfeita. O mesmo ocorre se $A_1 > 0$ e se n for suficientemente grande. Se $A_2 > 0$, é fácil de ver que $S_R^* \geq 0$ se e somente se $S_R \leq L_1$, onde

$$L_1 = \frac{12r(r+2)}{A_2} + (r+2)\left(1 - \frac{A_1}{A_2}\right).$$

Negligenciando a parcela de ordem 1, temos $L_1 \simeq 12r(r+2)/A_2$. Lembrando que S_R tem distribuição assintótica χ^2 com r graus de liberdade sob H_0 , notamos que, para n grande, se $S_R > 12r(r+2)/A_2$, temos forte evidência em favor da hipótese alternativa. Neste caso, não haveria necessidade prática de se calcular a estatística corrigida. A metodologia de correção só se faz necessária quando existe uma chance razoável de aceitar H_0 . Um forte indicativo de rejeição de H_0 torna a estratégia de correção inútil do ponto de vista prático. Em resumo, se $A_3 = 0$ e $A_2 > 0$ e se n for suficientemente grande, as amostras que não evidenciam forte rejeição de H_0 produzem valores de S_R^* não negativos.

Suponhamos agora que $A_3 = 0$ e $A_2 < 0$. De (2.15), temos que $S_R^* \geq 0$ se e somente se $S_R \geq L_1$. Mas $L_1 \leq 0$ para n suficientemente grande e, neste caso, a condição necessária e suficiente é satisfeita.

Vejamos o que ocorre se $A_3 > 0$. De (2.15), é fácil ver que $S_R^* \geq 0$ se e somente se $LI \leq S_R \leq LS$ onde $LI = a_1 - a_2$ e $LS = a_1 + a_2$ com

$$a_1 = (r+4)\left(1 - \frac{A_2}{2A_3}\right)$$

e

$$a_2 = \left\{ \frac{12r(r+2)(r+4)}{A_3} + (r+4)^2 \left(1 - \frac{A_2}{2A_3}\right)^2 + (r+2)(r+4) \left(\frac{A_2 - A_1}{A_3} - 1\right) \right\}^{1/2}.$$

Note que a expressão entre chaves em a_2 é positiva para n suficientemente grande. Negligenciando a_1 , que é de ordem 1 e as parcelas desta ordem entre as chaves de a_2 , temos que $LI \simeq -a_3$ e $LS \simeq a_3$, onde $a_3 = \{12r(r+2)(r+4)/A_3\}^{1/2}$. Assim, para n suficientemente grande, a condição necessária e suficiente para termos $S_R^* \geq 0$ se reduz a $S_R \leq a_3$. Mas, se S_R não satisfaz tal condição, há forte evidência contra H_0 de forma que não há necessidade de se calcular S_R^* .

Da análise acima concluímos que, do ponto de vista prático e desde que o tamanho da amostra não seja muito pequeno, o fato de S_R^* poder assumir valores negativos não apresenta nenhum problema pois só ocorre em situações em que a evidência contra H_0 é bastante forte, fazendo com que não seja necessária a utilização da estratégia de correção da estatística.

Passamos a analisar agora outra característica de S_R^* aparentemente indesejável, ou seja, o fato de S_R^* não ser necessariamente uma função crescente de S_R . Propomos estudar a expressão em (2.15) como função de S_R e determinar para que valores de S_R a primeira derivada desta função é positiva. Este estudo é análogo à análise de positividade de S_R^* e, portanto, será omitido. A conclusão a que se chega é novamente a de que, se o tamanho da amostra não for muito pequeno, S_R^* não será uma função crescente de S_R somente em um intervalo de valores para S_R que corresponde a forte evidência contra H_0 , ou seja, quando a estratégia de correção da estatística não é necessária do ponto de vista prático.

CAPÍTULO 5

TESTES ESCORE APERFEIÇADOS EM MODELOS LINEARES GENERALIZADOS COM DISPERSÃO DESCONHECIDA E OUTRAS EXTENSÕES

5.1. INTRODUÇÃO

No Capítulo 3, definimos a classe dos modelos lineares generalizados (MLGs) supondo o parâmetro de dispersão ϕ^{-1} conhecido e igual para todas as observações. Particionamos o vetor paramétrico β da componente sistemática em $(\beta_1^T, \beta_2^T)^T$ e apresentamos testes escore aperfeiçoados das hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H_0 : \beta_1 \neq \beta_1^{(0)}$, sendo $\beta_1^{(0)}$ um vetor pré-fixado. Neste capítulo, eliminamos a suposição de que ϕ é conhecido e derivamos testes escore aperfeiçoados de hipóteses referentes ao parâmetro ϕ bem como ao conjunto completo de parâmetros β . O aperfeiçoamento de testes escore relativos a um subconjunto de parâmetros β com ϕ desconhecido será objeto de pesquisa futura.

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$ com função de den-

sidade (3.1) com $a(y, \phi) = d_1(\phi) + d_2(y)$, isto é, *so vale para Gamma/N/NF*

$$\pi(y; \theta_l, \phi) = \exp[\phi\{y\theta_l - b(\theta_l) + \overset{=0}{c(y)}\} + d_1(\phi) + d_2(y)], \quad l = 1, \dots, n, \quad (5.1)$$

que corresponde a um modelo da família exponencial de distribuições na forma natural com parâmetros canônicos ϕ e $\phi\theta$. Supomos que $d_1(\cdot)$ tem as quatro primeiras derivadas. Seguindo a parametrização usada na Tabela 3.1, temos $d_1(\phi) = \phi \log \phi - \log \Gamma(\phi)$ para a distribuição gamma e $d_1(\phi) = \frac{1}{2} \log \phi$ para as distribuições normal e normal inversa. A média μ_l da observação y_l está relacionada com o vetor paramétrico $\beta = (\beta_1, \dots, \beta_p)^T$ através da função de ligação $d(\mu_l) = \eta_l$, onde η_l é a l -ésima componente do vetor $\eta = X\beta$ de dimensão p . A matriz X , suposta conhecida e de posto p ($p < n$), é formada pelos vetores de variáveis auxiliares $x_1, \dots, x_p$.

Sejam $\hat{L}_n$ e $\hat{L}_p$ as log-verossimilhanças maximizadas (com relação a β e não a ϕ) para o modelo saturado, isto é, com n parâmetros, e o modelo sob investigação respectivamente. Uma medida de qualidade do ajuste deste último modelo é definida por $S_p = 2(\hat{L}_n - \hat{L}_p) = \phi D_p(y, \hat{\mu})$ onde

$$D_p(y, \mu) = D_p(y, \hat{\mu}) = 2 \sum \left\{ v(y_l) - v(\mu_l) + (\mu_l - y_l)q(\mu_l) \right\}$$

é chamado de desvio do modelo e $\hat{\mu}$ é a *sob o modelo de investigação* estimativa de máxima verossimilhança de μ que não depende de ϕ . Aqui $v(\mu) = \mu q(\mu) - b(q(\mu))$ e $q(\mu) = \theta$. Para os modelos normal, normal inverso e gama, os desvios são escritos, respectivamente, como $\sum (y_l - \hat{\mu}_l)^2$, $\sum (y_l - \hat{\mu}_l)^2 / (\hat{\mu}_l^2 y_l)$ e $2 \sum \{-\log(y_l / \hat{\mu}_l) + (y_l - \hat{\mu}_l) / \hat{\mu}_l\}$. A estimativa de máxima verossimilhança de ϕ , cujo viés, de ordem n^{-1} , foi calculado por Cordeiro e McCullagh (1991), é dada pela solução da equação

$$2nd'_1(\phi) = D_p(y, \mu) - 2 \sum \{v(y_l) + c(y_l)\}. \quad (5.2)$$

Para os modelos normal e normal inverso, a equação (5.2) conduz a $\phi^{-1} = D_p(y, \mu) / n$. Entretanto, (5.2) não tem solução explícita para o modelo gama reduzindo-se a $\log \phi - \psi(\phi) = D_p(y, \mu) / (2n)$, onde $\psi(\cdot) = \Gamma'(\cdot) / \Gamma(\cdot)$ é a função digama.

A matriz de informação para (β^T, ϕ) é $K = \text{diag}\{\phi X^T W X, -nd_1''(\phi)\}$, implicando que os parâmetros β e ϕ são globalmente ortogonais (Cox e Reid, 1987b). A ortogonalidade de β e ϕ é fundamental para que a obtenção das quantidades A_1 , A_2 e A_3 , que determinam os testes escore aperfeiçoados, seja bastante simplificada podendo, inclusive, ser baseado diretamente nos resultados do Capítulo 4. A estrutura bloco diagonal de K e a normalidade assintótica dos estimadores de máxima verossimilhança de β e ϕ implicam que tais estimadores são assintoticamente independentes.

Este capítulo está estruturado da seguinte forma: na Seção 5.2, apresentamos testes escore aperfeiçoados para hipóteses referentes ao parâmetro ϕ , enquanto que, na Seção 5.3, são abordados testes sobre o parâmetro β . Na Seção 5.4, estendemos os resultados para os modelos próprios de dispersão recentemente desenvolvidos por Barndorff-Nielsen e Jørgensen (1990) e, finalmente, na Seção 5.5, desenvolvemos testes escore aperfeiçoados para os parâmetros da regressão em modelos não-lineares da família exponencial com dispersão conhecida.

5.2. TESTES DE HIPÓTESES SOBRE O PARÂMETRO ϕ

Consideremos a hipótese nula composta $H_0 : \phi = \phi^{(0)}$ e a alternativa também composta $H : \phi \neq \phi^{(0)}$ com β representando o parâmetro de perturbação e $\phi^{(0)}$ sendo um valor especificado para ϕ .

Antes de prosseguirmos é conveniente definir a seguinte notação:

$$\begin{aligned} d_{(1)} &= d_{(1)}(\phi) = \phi(d_1'(\phi) - 1), \\ d_{(2)} &= d_{(2)}(\phi) = \phi^2 d_1''(\phi), \\ d_{(3)} &= d_{(3)}(\phi) = \phi^3 d_1'''(\phi), \\ d_{(4)} &= d_{(4)}(\phi) = \phi^4 d_1^{iv}(\phi). \end{aligned} \tag{5.3}$$

Uma notação semelhante foi sugerida por Jensen e Kristensen (1990) no contexto da distribuição gama, ou seja, com $d_1(\phi) = \phi \log \phi - \log \Gamma(\phi)$.

A estatística escore para este teste de hipóteses pode ser escrita como

$$S_R = - \frac{\{\phi^{(0)} n^{-1} \sum (y_l \hat{\theta}_l - b(\hat{\theta}_l) + c(y_l)) + n d_{(1)} + n \phi^{(0)}\}^2}{n d_{(2)}}, \quad (5.4)$$

onde $d_{(1)}$ e $d_{(2)}$ são avaliados em $\phi^{(0)}$ e $\hat{\theta}_l$ é o estimador de máxima verossimilhança de θ_l e não depende de ϕ . Em grandes amostras S_R segue aproximadamente uma distribuição χ_1^2 .

Para desenvolver testes escore aperfeiçoados é necessário calcular as quantidades A_1 , A_2 e A_3 em (2.6)–(2.8) que envolvem cumulantes conjuntos de derivadas de log-verossimilhança. A obtenção de tais cumulantes, bem como dos A 's, é apresentada no Apêndice 4. Temos

$$A_1 = - \frac{3p(p-2)}{n d_{(2)}}, \quad (5.5)$$

$$A_2 = - \frac{3}{n d_{(2)}^2} \{2p d_{(3)} + d_{(4)}\} \quad (5.6)$$

e

$$A_3 = - \frac{5d_{(3)}^2}{n d_{(2)}^3}. \quad (5.7)$$

A estatística escore aperfeiçoada S_R^* é obtida de (2.15) com os A 's dados em (5.5)–(5.7) e se reduz a

$$S_R^* = S_R \{1 - b(S_R, n, \phi^{(0)}, p)\}, \quad (5.8)$$

onde

$$\begin{aligned} b(x, n, \phi, p) = \frac{1}{n} \left\{ - \frac{p^2}{4d_{(2)}} + \frac{d_{(2)} + d_{(3)}}{2d_{(2)}^2} p + \frac{3d_{(2)}d_{(4)} - 5d_{(3)}^2}{12d_{(2)}^3} \right. \\ \left. + \left(- \frac{d_{(3)}}{6d_{(2)}^2} p - \frac{3d_{(2)}d_{(4)} - 10d_{(3)}^2}{36d_{(2)}^3} \right) x - \frac{d_{(3)}^2}{36d_{(2)}^3} x^2 \right\} \end{aligned} \quad (5.9)$$

com $d_{(2)}$, $d_{(3)}$ e $d_{(4)}$ em (5.3). Cordeiro (1987) e Cordeiro, Paula e Botter (1991) encontraram o fator de correção de Bartlett para a estatística da razão de verossimilhança para o teste das hipóteses em questão nos modelos lineares generalizados e modelos de dispersão respectivamente. Notamos que o fator de Bartlett nos modelos próprios de dispersão (expressão (40) de Cordeiro, Paula e Botter, 1991) coincide com o fator de correção em (5.8)

desde que as parcelas que dependem da própria estatística sejam eliminadas. As expansões assintóticas, sob a hipótese nula e até ordem n^{-1} , da função distribuição de S_R e de seus quantis são dadas, respectivamente, por (2.5) com os A 's substituídos pelas expressões em (5.5)–(5.7) e por

$$z_\alpha = x_\alpha \{1 + b(x_\alpha, n, \phi^{(0)}, p)\}, \quad (5.10)$$

onde $b(x, n, \phi, p)$ é dado em (5.9) e x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$.

Denotamos por $\gamma_{1\phi}$ e $\gamma_{2\phi}$ as medidas usuais de assimetria e curtose da função escore para o parâmetro ϕ , isto é, $\gamma_{1\phi} = \kappa_{\phi, \phi, \phi} / \kappa_{\phi, \phi}^{3/2}$ e $\gamma_{2\phi} = \kappa_{\phi, \phi, \phi, \phi} / \kappa_{\phi, \phi}^2$. A partir dos cumulantes encontrados no Apêndice 4, é fácil ver que $\gamma_{1\phi}^2 = -\frac{1}{n} \frac{d_{(3)}}{d_{(2)}^3}$ e $\gamma_{2\phi} = -\frac{1}{n} \frac{d_{(4)}}{d_{(2)}^2}$. Com esta notação podemos escrever $b(x, \phi, p) = b_1(x, n, \phi, p) + b_2(x, n, \phi)$ onde

$$b_1(x, n, \phi, p) = \frac{1}{4nd_{(2)}^2} \{-d_{(2)}p^2 + 2(d_{(2)} + d_{(3)})p\}$$

e

$$b_2(x, n, \phi) = \frac{1}{36} \{3(5\gamma_{1\phi}^2 - 3\gamma_{2\phi}) + (3\gamma_{2\phi} - 10\gamma_{1\phi}^2)x + \gamma_{1\phi}^2x^2\}.$$

Comparando a forma funcional de $1 - \{b_1(S_R, n, \phi, p) + b_2(S_R, n, \phi)\}$ com a do fator de correção para a estatística escore no caso uniparamétrico (vide expressão (2.20)), notamos que a presença dos parâmetros de perturbação β produz um efeito aditivo na parcela de ordem n^{-1} do fator de correção. Em outras palavras, se β fosse conhecido este fator se reduziria a $1 - b_2(S_R, n, \phi)$.

Um teste escore aperfeiçoado para as hipóteses $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ pode ser conduzido de duas maneiras distintas que são equivalentes exceto por termos de ordem inferior a n^{-1} : (i) utilizando a estatística S_R^* e a distribuição χ_1^2 como referência ou (ii) a estatística S_R e as expansões da função distribuição de S_R e de seus quantis.

Deve ser notado que as expressões (5.5)–(5.7) dependem do modelo somente através do posto p da matriz X e do parâmetro ϕ . Elas não dependem do parâmetro desconhecido β . Ao notar ainda que ϕ deve ser avaliado sob H_0 , fica claro que os testes

escore aperfeiçoados não requerem a estimação de ϕ . O teste da razão de verossimilhança, por outro lado, requer a estimação de ϕ sob a hipótese alternativa o que envolve um custo computacional a mais principalmente porque a equação de estimação de ϕ não tem, em geral, solução explícita.

Como aplicação, considere as variáveis aleatórias $y_1, \dots, y_n$ independentes e identicamente distribuídas com função densidade em (5.1). As estatísticas escore aperfeiçoadas para os modelos normal, normal inverso e gama foram dadas anteriormente na Tabela 4.1.² Evidentemente, estes resultados poderiam ser obtidos diretamente de (5.8).

Apresentamos agora um pequeno estudo de simulação para verificar se os testes escore aperfeiçoados rejeitam a hipótese nula com uma probabilidade mais próxima do nível de significância ^{nominal} fixado do que o teste escore usual em amostras pequenas e moderadas. Consideremos o teste das hipóteses $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ em um modelo gama simples. Fixamos a média geral igual a 1 e geramos 10000 amostras para cada combinação de $\phi^{(0)} = 0,5; 1,0; 2,0$ e $n = 5, 10, 15, 20$, sempre sob a suposição de que a hipótese nula é verdadeira. Com base nas amostras simuladas, calculamos as probabilidades $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ com níveis nominais $\alpha = 1,0; 2,5; 5,0; 10,0\%$. A Tabela 5.1 mostra os resultados obtidos.

Os desempenhos dos testes aperfeiçoados não parecem ser muito melhores do que o do teste escore usual para os níveis nominais de 1,0 e 2,5%. No entanto, o aperfeiçoamento é bastante acentuado para os níveis nominais de 5,0 e 10,0% mesmo em amostras bem pequenas. Notamos que o teste escore usual parece ser muito conservativo no sentido de que tende a rejeitar a hipótese nula com uma frequência bem menor do que a indicada pelo correspondente nível nominal. O efeito das correções para a estatística S_R e para seus quantis, pelo menos para $\alpha = 5,0$ e $10,0\%$, é de elevar as taxas de rejeição para valores bastante próximos dos níveis nominais. Evidentemente, as correções tem menor efeito quando o tamanho da amostra é grande.

Honda (1988) apresenta um pequeno estudo de simulação referente a testes escore

Tabela 5.1. Valores simulados de $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ (em percentagem) para testar $H_0: \phi = \phi^{(0)}$ em um modelo gama simples

n	Níveis	$\phi^{(0)} = 0,5$			$\phi^{(0)} = 1,0$			$\phi^{(0)} = 2,0$		
	Nominais (%)									
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
5	1,0	1,1	1,1	1,4	1,2	1,1	1,5	1,0	0,9	1,3
	2,5	1,9	2,3	3,3	1,9	2,1	3,3	1,6	2,0	3,2
	5,0	2,9	3,4	5,0	2,7	3,1	4,8	2,4	2,9	4,9
	10,0	6,0 4,4	8,5 6,5	9,8 10,0	8,4 8,0	19,8 5,7	10,1 10,0	9,2 3,8	19,9 5,1	10,0 10,2
10	1,0	1,0	0,9	1,2	1,0	0,7	1,1	1,2	0,9	1,5
	2,5	1,8	2,2	2,6	1,6	2,0	2,6	2,0	2,2	2,8
	5,0	3,4	4,5	5,3	3,0	3,9	4,9	3,2	4,2	5,3
	10,0	7,7	9,3	10,2	7,1	9,0	10,1	7,1	9,2	10,4
15	1,0	1,0	0,8	1,1	1,0	0,9	1,2	1,0	0,8	1,1
	2,5	2,1	2,5	2,7	1,8	2,2	2,4	1,8	2,0	2,4
	5,0	3,9	4,7	5,1	3,5	4,6	5,0	3,4	4,4	4,9
	10,0	8,4	9,8	10,2	7,9	9,4	9,8	7,9	9,2	9,8
20	1,0	1,2	1,1	1,3	1,1	0,9	1,1	0,9	0,8	1,0
	2,5	2,4	2,7	2,9	2,0	2,3	2,6	1,7	2,1	2,3
	5,0	4,1	4,9	5,2	4,0	4,8	5,1	3,6	4,3	4,6
	10,0	9,0	9,9	10,1	9,0	9,8	10,2	8,2	9,3	9,6

Nota: Os números (1), (2) e (3) correspondem a $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$ e

$P(S_R \geq z_\alpha)$ respectivamente.

de heterocedasticidade. As taxas de rejeição do teste escore usual e do teste baseado na expansão dos quantis da estatística, ao nível nominal de 5% , são comparadas. Ele nota que o teste usual tende a ser bastante conservativo enquanto que o modificado apresenta taxas de rejeição próximas do nível nominal. Além disto, num estudo de simulação sobre o poder destes testes, Honda observa que o teste modificado tende a ser mais poderoso que o usual.

Voltando à Tabela 5.1, temos uma indicação de que o teste aperfeiçoado de $H_0 : \phi = \phi^{(0)}$ baseado na expansão dos quantis de S_R tende a ter um desempenho melhor do que o teste baseado na estatística S_R^* . No entanto, isto não pode ser generalizado para outros testes escore uma vez que as simulações representam uma situação bastante particular.

Exemplo 5.1

Um teste de particular interesse é o das hipóteses $H_0 : \phi = 1$ versus $H : \phi \neq 1$ no modelo gama uma vez que, sob H_0 , o modelo corresponde à distribuição exponencial. A estatística escore e os A 's para este teste envolvem a função digama e suas três primeiras derivadas avaliadas em 1. De Abramowitz e Stegun (1972; pgs. 267-271) temos $\psi(1) \cong -0,577$, $\psi'(1) \cong 1,645$, $\psi''(1) \cong -2,404$ e $\psi'''(1) \cong 6,494$. No caso de um modelo gama simples, isto é, com $p = 1$ e $X = (1, \dots, 1)^T$, vem da Tabela 4.1

$$S_R \cong \frac{n\{0,577 - \log(\bar{y}/\tilde{y})\}^2}{0,645} \quad (5.11)$$

com $\bar{y}$ e $\tilde{y}$ representando as médias aritmética e geométrica das observações respectivamente. De (5.8) e (5.10) seguem a estatística escore aperfeiçoada

$$S_R^* = S_R\{1 - b(S_R, n)\} \quad (5.12)$$

e a expansão para o quantil de ordem $1 - \alpha$ de S_R

$$z_\alpha = x_\alpha\{1 + b(x_\alpha, n)\}, \quad (5.13)$$

onde $b(x, n) \cong \frac{1}{n}(1,660 - 1,703x + 0,204x^2)$ e x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$.

A seguir apresentamos algumas ilustrações numéricas.

Exemplo 5.2

Consideremos os dados apresentados por Lee (1980; pg. 262) relativos a 10 pacientes com melanoma após passarem por cirurgia e terapia. Os tempos de remissão da doença (em meses) são os seguintes: 5, 8, 10, 11, 15, 20, 21, 23, 30 e 40. Assumindo um modelo gama, Lee obtém a estimativa de máxima verossimilhança $\hat{\phi} = 3,024$ para ϕ , bem como a "corrigida por viés" $\hat{\phi}_c = \hat{\phi}(1 + 3/n)^{-1} = 2,326$. A estatística escore S_R para o teste de $H_0 : \phi = 1$ iguala 2,561 com 1 grau de liberdade e conduz a um nível descritivo de 11,0% indicando fraca consistência com H_0 . O fator de correção para S_R vale 1,14 levando a $S_R^* = 2,910$. O nível descritivo é então reduzido a 8,8%. O aumento de 14% no valor da estatística faz com que a decisão ao nível de significância de 10% seja alterada. Resultado similar é obtido se usarmos a própria estatística escore em conexão com as expansões até ordem n^{-1} de sua função distribuição e de seus quantis. Neste caso, como $A_1 = -0,4651$, $A_2 = 1,2158$ e $A_3 = 3,6730$, vem, de (2.3), o nível descritivo de 8,5%. Finalmente, de (5.13), obtém-se o valor crítico modificado $z_{0,10} = 2,312$. Notamos que os níveis descritivos dos testes aperfeiçoados são bastante próximos e que ambos indicam a rejeição de H_0 ao nível de significância de 10% ao contrário do que ocorre com o teste usual.

Exemplo 5.3

Lee (1980; pg. 295) apresenta dois conjuntos de 20 observações independentes geradas segundo uma distribuição gama com índice $\phi = 2$. Para efeito de ilustração consideremos o primeiro conjunto formado pelos valores 17, 28, 49, 98, 119, 133, 145, 146, 158, 160, 174, 211, 220, 231, 252, 256, 267, 322, 323, 327 e a hipótese nula $H_0 : \phi = 1$ que já sabemos ser falsa. De (5.12) e (5.13) vem $S_R = 4,066$ e $S_R^* = 4,451$ levando a níveis descritivos, baseados na distribuição χ_1^2 , iguais a 0,044 e 0,035 respectivamente. O nível descritivo modificado baseado na expansão da função distribuição de S_R vale 0,033. Notamos que os testes modificados indicam maior evidência contra H_0 do que o teste usual.

Exemplo 5.4

Suponhamos agora que $p \geq 2$ populações tenham densidade gama e que amostras aleatórias independentes de tamanhos $n_1, \dots, n_p$ ($n_i \geq 1$, $i = 1, \dots, p$) sejam retiradas destas populações. Suponhamos ainda que todas as populações tenham o mesmo índice ϕ porém desconhecido. Admitimos a estrutura linear $\eta_i = \beta + \beta_i$, $i = 1, \dots, p$, com $\beta_p = 0$. O nosso interesse é testar $H_0 : \phi = 1$ versus $H : \phi \neq 1$, isto é, verificar se um modelo exponencial se ajusta bem aos dados. De (5.4) e de valores aproximados para a função gama e suas derivadas avaliadas no ponto 1, vem a estatística escore

$$S_R \cong \frac{\{0,577n - \sum_{i=1}^p n_i \log(\bar{y}_i/\tilde{y}_i)\}^2}{0,645n},$$

onde $\bar{y}_i$ e $\tilde{y}_i$ representam as médias aritmética e geométrica das observações da i -ésima amostra. A estatística escore aperfeiçoada S_R^* é obtida de (5.8) com $d_{(2)} \cong -0,645$, $d_{(3)} \cong 1,404$ e $d_{(4)} \cong -4,494$. Pode-se verificar que a expressão de S_R^* resultante coincide com a do Exemplo 5.1 se $p = 1$.

Exemplo 5.5

Como aplicação dos resultados do Exemplo 5.4, consideremos os dois conjuntos de 20 observações independentes geradas por Lee (1980; pg. 296) segundo uma distribuição gama com índice comum $\phi = 2$. O primeiro conjunto foi dado no Exemplo 5.3 e o segundo é formado pelos valores: 26, 34, 47, 59, 101, 112, 114, 136, 154, 154, 161, 186, 197, 226, 226, 243, 253, 269, 308, 465. As estatísticas escore S_R e escore modificada S_R^* para o teste de $H_0 : \phi = 1$ valem, respectivamente, 7,978 e 8,247. Embora estes valores sejam próximos e sejam bem maiores que os valores críticos, a níveis de significância usuais, obtidos da distribuição χ_1^2 de referência, notamos que S_R^* é a estatística que está indicando maior evidência contra H_0 que sabemos ser falsa.

Exemplo 5.6

Fornecemos agora uma verificação das expressões (5.5)–(5.7) para os modelos nor-

mal e normal inverso. Como, em ambos os casos, $d_1(\phi) = \frac{1}{2} \log \phi$, temos $A_1 = 6p(p-2)/n$, $A_2 = 12(3-2p)/n$ e $A_3 = 40/n$ para os dois modelos. Das expressões em (2.10) para os três primeiros momentos ordinários de S_R até ordem n^{-1} vem $\mu'_1(S_R) = 1 + p(p-2)/(2n)$, $\mu'_2(S_R) = 3 + (3p^2 - 14p + 12)/n$ e $\mu'_3(S_R) = 15 + 5(3p^2 - 66p + 104)/(2n)$. Para ambos os modelos obtém-se de (5.5) $S_R = \{n - \phi^{(0)} D_p(y, \hat{\mu})\}^2 / (2n)$ e, com ligação identidade, $\phi^{(0)} D_p(y, \hat{\mu})$ é distribuído, sob H_0 , exatamente como χ_{n-p}^2 . Pelo menos nestes casos, pode ser provado, através do cálculo direto dos três primeiros momentos de S_R , que estes coincidem com as expressões dadas acima e que, assim, os A 's estão corretos. Voltaremos a abordar os testes para o parâmetro ϕ nos modelos normal e normal inverso nos Exemplos 5.8 e 5.9.

5.3. TESTES DE HIPÓTESES SOBRE β COM ϕ DESCONHECIDO

Consideremos agora a hipótese nula composta $H_0 : \beta = \beta^{(0)}$ e a alternativa também composta $H : \beta \neq \beta^{(0)}$ com ϕ representando o parâmetro de perturbação e $\beta^{(0)}$ sendo um vetor de valores especificados para β . A estimativa de máxima verossimilhança de ϕ restrita a H_0 , $\tilde{\phi}$ digamos, é determinada por (5.2) com μ fixado em $\beta = \beta^{(0)}$. A estatística escore para o teste das hipóteses acima tem a forma

$$S_R = \tilde{S}^T W^{(0)1/2} X (X^T W^{(0)} X)^{-1} X^T W^{(0)1/2} \tilde{S}, \quad (5.14)$$

onde $\tilde{S} = (\tilde{s}_1, \dots, \tilde{s}_n)^T$, com $\tilde{s}_l = \tilde{\phi}^{1/2} (y_l - \mu_l^{(0)}) V_l^{(0)-1/2}$, é o vetor de resíduos de Pearson, W representa a matriz de pesos e V_l é a função variância para a l -ésima observação. As quantidades assinaladas com $^{(0)}$ são avaliadas em $\beta = \beta^{(0)}$. Em grandes amostras, S_R tem distribuição χ_p^2 aproximadamente. Vale a pena observar que a expressão de S_R para o teste de $H_0 : \beta = \beta^{(0)}$ no caso de ϕ ser conhecido coincide com (5.14) com $\tilde{\phi}$ substituído por ϕ .

Como na Seção 5.2, é possível obter expressões simples para os A 's devido à ortogonalidade de β e ϕ . A obtenção destas quantidades, baseada nos resultados do

Capítulo 4, é apresentada em detalhes no Apêndice 4. Obtemos

$$A_1 = \frac{6p(d_{(3)} + 2d_{(2)})}{nd_{(2)}^2}, \quad (5.15)$$

$$A_2 = \frac{3p(p+2)}{nd_{(2)}} + \frac{3}{\phi} 1^T B Z_d^{(2)} 1 \quad (5.16)$$

e

$$A_3 = \phi^{-1} 1^T (F - G)(3Z_d Z Z_d + 2Z^{(3)})(F - G)1, \quad (5.17)$$

onde $d_{(2)}$ e $d_{(3)}$ estão definidos em (5.3) e com a notação do Capítulo 3 para as matrizes envolvidas. As expressões (5.15) e (5.16) para A_1 e A_2 são iguais aos A 's correspondentes quando a hipótese nula é simples com o parâmetro ϕ sendo conhecido (vide Seção 3.6) mais termos adicionais que dependem somente de ϕ , n e p e representam efeitos provenientes do fato de ϕ ser desconhecido. Por outro lado, as expressões para A_3 para os casos de ϕ conhecido e desconhecido coincidem.

Como aplicação consideremos um modelo simples, isto é, com β unidimensional e X definido por um vetor de uns. A estatística score em (5.14) reduz-se a

$$S_R = n\tilde{\phi}(\bar{y} - \mu^{(0)})^2 / V^{(0)}$$

e os A 's resultantes são obtidos facilmente de (5.15)–(5.17) como

$$A_1 = \frac{6(d_{(3)} + 2d_{(2)})}{nd_{(2)}^2}, \quad \text{Tr} \left(\underbrace{W Z W Z}_{\phi} \right)$$

$$A_2 = \frac{9}{nd_{(2)}} + \frac{3}{n\phi V} \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2 V}{d\mu^2} \right\}$$

e

$$A_3 = \frac{5}{n\phi V} \left(\frac{dV}{d\mu} \right)^2.$$

Como esperado os A 's não dependem da função de ligação. As estatísticas score e score aperfeiçoada para o teste $H_0 : \mu = \mu^{(0)}$ nos modelos normal, normal inverso e gama foram apresentadas anteriormente na Tabela 4.2 mas poderiam, evidentemente, ser obtidas dos resultados acima.

Exemplo 5.7

As expressões (5.15)–(5.17) induzem outro caso especial interessante. Para o modelo normal, $A_1 = A_3 = 0$ e $A_2 = -6p(p+2)/n$. Podemos fazer uma verificação parcial das expressões dos A 's considerando o modelo normal com ligação identidade. Neste caso, $S_R = \tilde{\phi}(y - \mu^{(0)})^T X(X^T X)^{-1} X^T (y - \mu^{(0)})$ com $\tilde{\phi}^{-1} = \frac{1}{n}(y - X\beta^{(0)})^T (y - X\beta^{(0)})$. É fácil ver que S_R é igual a n vezes uma variável aleatória com distribuição beta com parâmetros $\frac{p}{2}$ e $\frac{n-p}{2}$ de onde seguem os três primeiros momentos ordinários: $\mu'_1(S_R) = p$, $\mu'_2(S_R) = np(p+2)/(n+2)$ e $\mu'_3(S_R) = n^2 p(p+2)(p+4)/\{(n+2)(n+4)\}$. Substituindo estes momentos nas expressões em (2.8) chega-se aos A 's dados acima, mostrando assim que eles estão corretos.

5.4. EXTENSÃO AOS MODELOS PRÓPRIOS DE DISPERSÃO**5.4.1. A CLASSE DOS MODELOS PRÓPRIOS DE DISPERSÃO**

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$ com função densidade

$$\pi(y; \theta_l, \phi) = \exp\{\phi t(y, \theta_l) + d_1(\phi) + d_2(y)\}, \quad (5.18)$$

onde $t(\cdot, \cdot)$, $d_1(\cdot)$ e $d_2(\cdot)$ são funções conhecidas e $\phi > 0$. Se o parâmetro ϕ é conhecido, (5.18) inclui todas as distribuições com um único parâmetro conhecido. Os modelos em (5.18) formam a classe dos modelos próprios de dispersão (Barndorff-Nielsen e Jørgensen, 1990) e são parte da classe mais ampla dos chamados modelos de dispersão (Jørgensen, 1987). Exemplos de modelos de dispersão podem ser vistos em Jørgensen (1983, 1987) e Cordeiro (1985). Em 1983 Jørgensen chamava os modelos de dispersão de classe estendida de modelos lineares generalizados.

Os modelos próprios de dispersão têm algumas propriedades importantes: (i) para ϕ conhecido, $T = t(y, \theta)$ é uma quantidade pivotal, isto é, sua distribuição não depende de θ ; (ii) os parâmetros θ e ϕ são globalmente ortogonais e (iii) para θ conhecido, (5.18) é

uma família exponencial com estatística canônica T .

Definimos a parte sistemática do modelo por

$$d(\theta_l) = \eta_l = f(x_l, \beta), \quad (5.19)$$

onde $d(\cdot)$ é uma função conhecida monótona e diferenciável, chamada de função de ligação, $\beta = (\beta_1, \dots, \beta_p)^T$, $p < n$, é um conjunto de parâmetros a ser estimado, $f(\cdot, \cdot)$ é diferenciável com respeito aos componentes de β e $x_l = (x_{l1}, \dots, x_{lp})^T$ é um vetor de valores fixados de p variáveis auxiliares para a l -ésima observação. Assumimos que a matriz de derivadas $X^* = X^*(\beta) = \partial\eta/\partial\beta$, chamada de matriz modelo local, tem posto p para todo β .

A classe dos modelos definidos por (5.18) com ϕ desconhecido e pela componente sistemática η em (5.19) engloba a classe dos MLGs definida na Seção 5.1. Neste caso, $t(y, \theta) = y\theta - b(\theta) + c(y)$, $\eta = X\beta$ e θ é expresso em função da média μ .

Barndorff-Nielsen e Jørgensen (1990) estudam a construção de modelos próprios de dispersão e apresentam diversas distribuições importantes nesta classe como, por exemplo, a distribuição de McCullagh (1989), von Mises, uma classe de distribuições simplex no intervalo unitário, etc.

É conveniente introduzir agora a seguinte notação:

$$D_i = D_i(\theta, \phi) = E\{\partial^i t(y, \theta)/\partial\theta^i\}, \quad i = 1, 2, 3, 4. \quad (5.20)$$

De condições de regularidade, tem-se $D_1 = 0$ e $D_2 = -\phi E\{(\partial t(y, \theta)/\partial\theta)^2\}$. A matriz de informação total de Fisher do parâmetro (β^T, ϕ) é $K = \text{diag}\{\phi X^{*T} W^* X^*, -n d_1''(\phi)\}$, onde $W^* = \text{diag}\{-D_2(d\theta/d\eta)^2\}$. A estrutura bloco diagonal da matriz K e a normalidade assintótica dos estimadores de máxima verossimilhança de β e ϕ , $\hat{\beta}$ e $\hat{\phi}$ digamos, implicam que estes estimadores são assintoticamente independentes. O fato de β e ϕ serem ortogonais facilita imensamente os cálculos envolvidos na obtenção de fórmulas simples para os fatores de correção das estatísticas escore dos testes de $H_0 : \phi = \phi^{(0)}$ e de $H_0 : \beta = \beta^{(0)}$.

A estimativa de máxima verossimilhança de β não depende do parâmetro ϕ e pode ser obtida numericamente pela solução iterativa das equações (Jørgensen, 1983, 1984

e Green, 1984)

$$X^{*(i)T} W^{*(i)} X^{*(i)} \beta^{(i+1)} = X^{*(i)T} W^{*(i)} y^{*(i)},$$

onde $y^* = M\dot{t} + X^*\beta$ com $M = \text{diag}\{-D_2^{-1} d\eta/d\theta\}$, $\dot{t} = (\partial t/\partial\theta_1, \dots, \partial t/\partial\theta_n)^T$ e W^* é definido acima. Uma maneira simples de encontrar a solução deste sistema de equações através do GLIM é sugerida por Cordeiro e Paula (1989a).

Por outro lado, a estimativa de máxima verossimilhança de ϕ é a solução da equação

$$\sum t(y_l, \theta_l) + n d'_1(\phi) = 0, \quad (5.21)$$

com θ avaliado em $\beta = \hat{\beta}$.

Inferências sobre os parâmetros β e ϕ em modelos de dispersão, incluindo análise de desvio, são consideradas por Cordeiro e Paula (1990) e Barndorff-Nielsen e Jørgensen (1990). Cordeiro, Paula e Botter (1991) desenvolvem testes da razão de verossimilhança aperfeiçoados para os parâmetros do modelo de dispersão e expressam os fatores de correção das estatísticas dos testes em forma matricial.

Nosso objetivo é estender os resultados das Seções 5.2 e 5.3 à classe dos modelos próprios de dispersão, ou seja, desenvolver testes escore aperfeiçoados para o parâmetro ϕ e para o conjunto completo de parâmetros β nesta classe. Evidentemente, outros testes são de grande interesse como, por exemplo, o de hipóteses relativas a somente alguns componentes de β . Estes testes não são tratados aqui e serão objeto de trabalho futuro.

5.4.2. TESTES DE HIPÓTESES SOBRE O PARÂMETRO ϕ

Consideremos a hipótese nula $H_0 : \phi = \phi^{(0)}$ e a alternativa $H : \phi \neq \phi^{(0)}$ com β representando um parâmetro de perturbação. A estatística escore para testar estas hipóteses tem a forma

$$S_R = -\frac{1}{n d_{(2)}} \left\{ \phi^{(0)} \sum t(y_l, \hat{\theta}_l) + n d_{(1)} + n \phi^{(0)} \right\}^2, \quad (5.22)$$

onde $d_{(1)}$ e $d_{(2)}$, definidos em (5.3), são avaliados em $\phi^{(0)}$. É fácil ver que, se $t(y, \theta) = y\theta - b(\theta) + c(y)$, S_R coincide com (5.4).

Na Seção 5.2, encontramos as quantidades A_1 , A_2 e A_3 relativas ao teste escore das hipóteses acima em MLGs. Notamos que estas dependem do modelo somente através do posto da matriz X e das funções $d_{(2)}(\cdot)$, $d_{(3)}(\cdot)$ e $d_{(4)}(\cdot)$ que envolvem derivadas da função $d_1(\cdot)$. Isto sugere que as expressões dos A 's sejam válidas para os modelos próprios de dispersão. Mostramos no Apêndice 4 que isto, de fato, é verdade e, conseqüentemente, a estatística escore aperfeiçoada e a expansão até ordem n^{-1} para os quantis de S_R têm as formas dadas em (5.8) e (5.10), respectivamente.

A seguir apresentamos diversas distribuições da classe dos modelos próprios de dispersão e os testes escore usuais e aperfeiçoados da hipótese nula $H_0 : \phi = \phi^{(0)}$ nestes modelos com componente sistemática dada em (5.19).

Exemplo 5.8

A distribuição normal com média μ e variância ϕ^{-1} é um modelo próprio de dispersão com $t(y, \mu) = -(y - \mu)^2/2$ e $d_1(\phi) = \frac{1}{2} \log \phi$. De (5.3) segue que $d_{(1)} = (1 - 2\phi)/2$, $d_{(2)} = -\frac{1}{2}$, $d_{(3)} = 1$ e $d_{(4)} = -3$. A estatística escore para testar $H_0 : \phi = \phi^{(0)}$ com β sendo um parâmetro de perturbação vem de (5.22) e tem a forma

$$S_R = \{n - \phi^{(0)} \sum (y_l - \hat{\mu}_l)^2\} / (2n).$$

Agora, de (5.8), (5.10) e das funções $d_{(i)}$, $i = 2, 3, 4$, dadas acima, vem a estatística escore aperfeiçoada

$$S_R^* = S_R \{1 - b(S_R, p, n)\} \quad (5.23)$$

e a expansão até ordem n^{-1} para os quantis de S_R

$$z_\alpha = x_\alpha \{1 + b(x_\alpha, p, n)\}, \quad (5.24)$$

onde x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$ e

$$b(x, p, n) = \frac{1}{18n} \{9p^2 + 18p + 6 - 2(6p + 11)x + 4x^2\}.$$

Se $p = 1$, S_R^* concorda com a expressão dada na Tabela 4.1.

A partir da distribuição normal, podemos obter, através de uma transformação de variável, a distribuição log-normal que é bastante utilizada em análise de sobrevivência. Esta pode ser definida como a distribuição de uma variável aleatória cujo logaritmo tem distribuição normal com média θ e variância ϕ^{-1} . A função de densidade correspondente à distribuição log-normal tem a forma

$$\pi(y; \theta, \phi) = \frac{1}{y} \sqrt{\frac{\phi}{2\pi}} \left\{ -\frac{\phi}{2} (\log y - \theta)^2 \right\}, \quad y > 0$$

com $E(y) = \exp(\theta + \phi/2)$ e $\text{Var}(y) = (\exp \phi^{-1} - 1) \exp(2\theta + \phi^{-1})$.

Note-se que este não é um modelo linear generalizado porém é um modelo próprio de dispersão com $t(y, \theta) = -\frac{1}{2}(\log y - \theta)^2$ e $d_1(\phi) = \frac{1}{2} \log \phi$. A função $d_1(\cdot)$ coincide com a da distribuição normal.

Exemplo 5.9

A distribuição normal inversa com função densidade

$$\pi(y; \theta, \phi) = \left(\frac{\phi}{2\pi} \right)^{\frac{1}{2}} y^{-\frac{3}{2}} \exp \left\{ -\phi \frac{(y - \mu)^2}{2\mu^2 y} \right\},$$

onde $\mu = -(-2\theta)^{-1/2}$ é a média, é um modelo próprio de dispersão com $t(y, \theta) = -(y - \mu)^2 / (2\mu^2 y)$ e $d_1(\phi) = \frac{1}{2} \log \phi$. Como $d_1(\cdot)$ coincide com a correspondente função do modelo normal, a forma da estatística escore aperfeiçoada e a expansão dos quantis de S_R coincidem com (5.23) e (5.24), porém aqui $S_R = \{n - \phi^{(0)} \sum (y_l - \hat{\mu}_l)^2 / (\hat{\mu}_l^2 y_l)\}^2 / (2n)$.

Exemplo 5.10

A distribuição gama com média μ e índice ϕ também faz parte da classe dos modelos próprios de dispersão com $t(y, \theta) = -y/\mu + \log(y/\mu)$, onde $\mu = -1/\theta$, e $d_1(\phi) = \phi(\log \phi - \psi(\phi))$. Aqui $d_{(1)} = \phi(\log \phi - \psi(\phi))$, $d_{(2)} = \phi^2(\phi^{-1} - \psi'(\phi))$, $d_{(3)} = \phi^3(-\phi^{-2} - \psi''(\phi))$ e $d_{(4)} = \phi^4(2\phi^{-3} - \psi'''(\phi))$, onde $\psi(\cdot)$ é a função digama. A estatística escore para testar $H_0 : \phi = \phi^{(0)}$ é dada por

$$S_R = -\frac{n}{\phi^{(0)}(1 - \phi^{(0)}\psi'(\phi^{(0)}))} \left\{ \phi^{(0)} \left(\frac{\bar{y}}{\hat{\mu}} - \log \phi^{(0)} + \psi(\phi^{(0)}) - 1 \right) + \log \left(\frac{\hat{\mu}}{\tilde{y}} \right) \right\}^2$$

e a estatística aperfeiçoada S_R^* e os quantis modificados z_α vêm, respectivamente, de (5.8) e (5.10).

A partir da expansão da função $\psi(\phi)$ dada no Exemplo 4.3 é possível escrever S_R^* e z_α como

$$S_R^* \cong S_R\{1 - b(S_R, n, \phi^{(0)}, p)\}$$

e

$$z_\alpha \cong z_\alpha\{1 + b(x_\alpha, n, \phi^{(0)}, p)\},$$

onde

$$b(x, n, \phi, p) = \frac{1}{6p} \left(3p^2 + 6p + 2 - \frac{p^2 + 60}{\phi} \right) - \frac{1}{3p(p+2)} \left(6p + 11 - \frac{p+60}{\phi} \right) x \\ + \frac{10}{3p(p+2)(p+4)} \left(1 - \frac{3}{\phi} \right) x^2.$$

e x_α é tal que $P(\chi_1^2 \leq x_\alpha) = 1 - \alpha$. Aqui, termos de ordem inferior a ϕ^{-1} foram desprezados, de forma que as aproximações são utilizadas para valores suficientemente grandes de $\phi^{(0)}$.^(*)

O logaritmo de uma variável aleatória com distribuição gama com média μ e índice ϕ tem distribuição log-gama com função densidade

$$\pi(y; \theta, \phi) = \frac{\phi^\phi}{\Gamma(\phi)} \exp\{\phi[y - \theta - \exp(y - \theta)]\}, \quad -\infty < y < \infty,$$

e com $E(y) = \theta + \psi(\phi) + \log \phi$ e $Var(y) = \psi''(\phi)$. Embora a distribuição gama faça parte da classe dos MLGs, a log-gama não faz. No entanto, é um modelo próprio de dispersão com $t(y, \theta) = y - \theta - \exp(y - \theta)$ e $d_1(\phi) = \phi(\log \phi - \psi(\phi))$. A função $d_1(\cdot)$ coincide com a da distribuição gama.

Exemplo 5.11

McCullagh (1989) define a função densidade

$$\pi(y; \theta, \phi) = \frac{y^{\phi-\frac{1}{2}}(1-y)^{\phi-\frac{1}{2}}2^{2\phi}}{\{(1+\theta)^2 - 4\theta y\}^\phi B(\phi + \frac{1}{2}, \frac{1}{2})}, \quad 0 < y < 1,$$

^(*) Neste exemplo e nos que se seguem, o desenvolvimento de fórmulas aproximadas para S_R^* e z_α envolve um trabalho algébrico extenso. Agradeço à colega Denise A. Botter por ter me cedido seus manuscritos com desenvolvimentos semelhantes que muito me auxiliaram na obtenção de tais fórmulas.

onde $B(\cdot, \cdot)$ é a função beta, $-1 \leq \theta \leq 1$ e $\phi > -\frac{1}{2}$, que é um modelo próprio de dispersão com

$$t(y, \theta) = \log\{y(1-y)\} - \log\{(1+\theta)^2 - 4\theta y\} + 2 \log 2$$

e

$$d_1(\phi) = -\log B\left(\phi + \frac{1}{2}, \frac{1}{2}\right)$$

desde que $\phi > 0$. Usamos a notação $y \sim \text{McC}(\theta, \phi)$ para indicar que a variável aleatória y tem a densidade acima com parâmetros θ e ϕ . Diversas propriedades importantes desta densidade foram estabelecidas por McCullagh (1989). Dentre elas, destacamos: (i) se $y \sim \text{McC}(\theta, \phi)$ com $\theta = -1, 0$ e 1 , então y tem distribuição beta com parâmetros $(\frac{1}{2}, \phi + \frac{1}{2})$, $(\phi + \frac{1}{2}, \phi + \frac{1}{2})$ e $(\phi + \frac{1}{2}, \frac{1}{2})$ respectivamente; (ii) se $\phi = 0$, a distribuição $\text{McC}(\theta, \phi)$ se reduz a uma distribuição beta com parâmetros $(1/2, 1/2)$ e não depende de θ ; (iii) para outros valores de ϕ , a distribuição $\text{McC}(\theta, \phi)$ não pertence à família beta exceto se $\theta = -1, 0$ e 1 ; (iv) se $y \sim \text{McC}(\theta, \phi)$, então $1 - y \sim \text{McC}(-\theta, \phi)$; (v) se $\theta = 0$ e $\phi = \frac{1}{2}$, a distribuição $\text{McC}(\theta, \phi)$ se reduz à uniforme no intervalo unitário; (vi) a média e a variância de y são dadas por

$$E(y) = \frac{\phi\theta}{2(\phi+1)} + \frac{1}{2} \quad \text{e} \quad \text{Var}(y) = \left\{1 - \frac{\phi(\phi-1)\theta^2}{(\phi+1)(\phi+2)}\right\} \frac{1}{8(\phi+1)}.$$

Outros cumulantes de y , bem como uma discussão sobre estimação e testes de hipóteses são encontradas no artigo de McCullagh.

A partir da função $d_1(\cdot)$ acima e de (5.3) obtemos

$$d_{(1)} = \phi(2\psi(\phi) - 2\psi(2\phi) + 2 \log 2 - 1) + 1,$$

$$d_{(2)} = 2\phi^2(\psi'(\phi) - 2\psi'(2\phi)) - 1,$$

$$d_{(3)} = 2\phi^3(\psi''(\phi) - 4\psi''(2\phi)) + 2$$

e

$$d_{(4)} = 2\phi^4(\psi'''(\phi) - 8\psi'''(2\phi)) - 6,$$

que dependem da função digama e suas três primeiras derivadas. Fórmulas aproximadas para os $d_{(i)}$'s para ϕ grande são obtidas a partir da expansão de $\psi(\phi)$ dada no Exemplo 4.3 e conduzem a fórmulas aproximadas para S_R^* e z_α como no exemplo anterior, porém com

$$b(x, n, \phi, p) = \frac{1}{n} \left[\frac{1}{12p} \left\{ 6p^2 + 12p + 148 + \frac{3(p^2 + 108)}{\phi} \right\} - \frac{1}{6p(p+2)} \left\{ 3(4p+3) + \frac{3(p+84)}{\phi} \right\} x + \frac{10}{3p(p+2)(p+4)} \left(1 + \frac{9}{2\phi} \right) x^2 \right].$$

Exemplo 5.12

Consideremos agora a família de distribuições simplex no intervalo unitário estudada por Barndorff-Nielsen e Jørgensen (1990). As funções densidade, indexadas por números reais conhecidos α_1 e α_2 , têm a forma

$$\pi(y; \theta, \phi) = a(\theta, \phi, \alpha_1, \alpha_2) y^{\alpha_1-1} (1-y)^{\alpha_2-1} \exp\{\phi t(y, \theta)\}, \quad (5.25)$$

onde

$$0 < y < 1$$

$$t(y, \theta) = -\frac{\theta^{2\alpha_1-1} (1-\theta)^{2\alpha_2-1} (y-\theta)^2}{2y(1-y)},$$

$0 < \theta < 1$ e $a(\theta, \phi, \alpha_1, \alpha_2)$ é uma constante normalizadora que, a princípio, pode ser obtida por integração. Para valores específicos de α_1 e α_2 , $a(\theta, \phi, \alpha_1, \alpha_2) = c(\phi, \alpha_1, \alpha_2)$ não depende de θ e, nestes casos, (5.25) corresponde a um modelo próprio de dispersão com $d_1(\phi) = \log c(\phi, \alpha_1, \alpha_2)$ indexado pelas constantes α_1 e α_2 . Barndorff-Nielsen e Jørgensen (1990) mostram que $c(\phi, \frac{1}{2}, \frac{1}{2}) = \pi^{-\frac{1}{2}} \{\Gamma(\frac{1}{2}, \frac{1}{2}\phi)\}^{-1} \exp(-\frac{1}{2}\phi)$, $c(\phi, -\frac{1}{2}, -\frac{1}{2}) = c(\phi, -\frac{1}{2}, \frac{1}{2}) = c(\phi, \frac{1}{2}, -\frac{1}{2}) = (2\pi\phi^{-1})^{-\frac{1}{2}}$ e $c(\phi, 0, 0) = \{2K_0(\phi)\exp\phi\}^{-1}$, onde $\Gamma(\cdot, \cdot)$ é a função gama incompleta e $K_\nu(\cdot)$ é a função modificada de Bessel de terceira espécie e de ordem ν (vide Abramowitz e Stegun, 1970, expressões 6.5.3 e 9.6.1). (Para $\alpha_1 = \alpha_2 = -\frac{1}{2}$, (5.25) corresponde à distribuição normal inversa generalizada.) Para ϕ grande, $c(\phi, \alpha_1, \alpha_2) \cong (\phi/2\pi)^{1/2}$ que independe dos valores de α_1 e α_2 . Isto implica que se ϕ for grande as

quantidades A_1 , A_2 e A_3 para o modelo normal podem ser usadas como uma primeira aproximação para os A 's da família (5.25).

Para $\alpha_1 = \alpha_2 = -\frac{1}{2}$, $\alpha_1 = -\alpha_2 = -\frac{1}{2}$ e $\alpha_1 = -\alpha_2 = \frac{1}{2}$, as funções $d_{(i)}(\cdot)$, $i = 1, \dots, 4$, coincidem com as correspondentes ao modelo normal e consequentemente a estatística escore aperfeiçoada e a expansão até ordem n^{-1} para os quantis de S_R vêm diretamente de (5.23) e (5.24).

No caso de $\alpha_1 = \alpha_2 = \frac{1}{2}$, temos

$$t(y, \theta) = -\frac{(y - \theta)^2}{2y(1 - y)} \quad \text{e} \quad d_1(\phi) = -\log \Gamma\left(\frac{1}{2}, \frac{1}{2}\phi\right) - \frac{1}{2}\phi - \frac{\pi}{2}$$

Definindo $M = M(\phi) = \{d\Gamma(\frac{1}{2}, \frac{1}{2}\phi)/d\phi\}/\Gamma(\frac{1}{2}, \frac{1}{2}\phi)\}^{-1}$ e usando a igualdade $d\Gamma(a, x)/dx = -x^{a-1}e^{-x}$ (Abramowitz e Stegun, 1970, relação 6.5.25), vem

$$d_{(1)} = -\phi\left(M + \frac{3}{2}\right), \quad d_{(2)} = \phi^2\left\{\frac{(\phi + 1)M}{2\phi} + M^2\right\},$$

$$d_{(3)} = -\phi^3\left\{\frac{(\phi^2 + 2\phi + 3)M}{4\phi^2} - \frac{(3\phi + 1)M^2}{2\phi} - M^3\right\}$$

e

$$d_{(4)} = \phi^4\left\{\frac{(\phi^3 + 3\phi^2 + 9\phi + 15)M}{8\phi^3} + \frac{(7\phi^2 + 14\phi + 15)M^2}{4\phi^2} + \frac{9(\phi + 1)M^3}{2\phi} + 3M^4\right\}.$$

Finalmente, se $\alpha_1 = \alpha_2 = 0$, temos

$$t(y, \theta) = -\frac{(y - \theta)^2}{2y(1 - y)\theta(1 - \theta)} \quad \text{e} \quad d_1(\phi) = -\log K_0(\phi) - \phi.$$

Agora, definindo $Q = Q(\phi) = K_1(\phi)/K_0(\phi)$ e usando as igualdades $K'_0(\phi) = -K_1(\phi)$ e $K'_1(\phi) = -K_0(\phi) - K_1(\phi)/\phi$ (Abramowitz e Stegun, 1970, relações 9.6.27 e 9.2.26), vem

$$d_{(1)} = \phi(Q - 2),$$

$$d_{(2)} = -\phi^2\left(1 + \frac{Q}{\phi} - Q^2\right),$$

$$d_{(3)} = -\phi^3\left\{\frac{2(\phi^2 - 1)Q}{\phi^2} + 3Q^2 - 2Q^3 - \frac{1}{\phi}\right\}$$

e

$$d_{(4)} = \phi^4 \left\{ \frac{2(4\phi^2 - 3)Q}{\phi^3} - \frac{8\phi^2 + 11}{\phi^2} - \frac{12Q^3}{\phi} \right\}.$$

Valores de $Q(\phi)$ para $0 \leq \phi \leq 20$ são obtidos da Tabela 9.8 de Abramowitz e Stegun (1970, pgs. 416–421).

Exemplo 5.13

Outro modelo próprio de dispersão de grande interesse é a distribuição de von Mises no círculo apropriada para estudar dados direcionais. A função de densidade tem a forma

$$\pi(y; \theta, \phi) = \frac{1}{2\pi I_0(\phi)} \exp\{\phi \cos(y - \theta)\}, \quad 0 < y < 2\pi,$$

onde $I_\nu(\cdot)$ é a função modificada de Bessel de primeira espécie e ordem ν (vide Abramowitz e Stegun, 1970, expressão 9.6.1), $0 < \theta < 2\pi$ e $\phi > 0$. Esta densidade é simétrica em torno de $y = \theta$ e, além disto, θ é a moda da distribuição. Por outro lado, ϕ é um parâmetro de precisão no sentido de que quanto maior for ϕ , maior é a concentração da densidade ao redor de θ e, se $\phi = 0$, a distribuição é uniforme. Propriedades e aplicações do modelo de von Mises, bem como um estudo sobre estimação e testes de hipóteses são encontrados no livro de Mardia (1972). Aqui

$$t(y, \theta) = \cos(y - \theta) \quad \text{e} \quad d_1(\phi) = -\log I_0(\phi).$$

Como $I'_0(\phi) = I_1(\phi)$ e $I'_1(\phi) = I_0(\phi) - I_1(\phi)/\phi$ (Abramowitz e Stegun; 1970, relações 9.6.27 e 9.6.26), vem

$$d_{(1)} = -\phi(r(\phi) + 1), \quad d_{(2)} = -\phi^2 r'(\phi) = \phi^2 \left(r(\phi)^2 + \frac{r(\phi)}{\phi} - 1 \right),$$

$$d_{(3)} = -\phi^3 r''(\phi) = \phi^3 \left(-2r(\phi)^3 - \frac{3}{\phi} r(\phi)^2 + \frac{2\phi^2 - 2}{\phi^2} r(\phi) + \frac{1}{\phi} \right)$$

e

$$d_{(4)} = -\phi^4 r'''(\phi) = \phi^4 \left(6r(\phi)^4 + \frac{12}{\phi} r(\phi)^3 - \frac{8\phi^2 - 11}{\phi^2} r(\phi)^2 - \frac{8\phi^2 - 6}{\phi^3} r(\phi) - \frac{3}{\phi^2} + 2 \right),$$

onde $r(\phi) = I_1(\phi)/I_0(\phi)$. Valores de $r(\phi)$, para $0 \leq \phi \leq 20$, são obtidos da Tabela 9.8 de Abramowitz e Stegun (1970, pgs. 416–421). Para ϕ grande^(*)

$$r(\phi) = 1 - \frac{1}{2\phi} - \frac{1}{8\phi^2} - \frac{1}{8\phi^3} - \frac{25}{128\phi^4} - \frac{13}{32\phi^5} + \dots$$

e, para ϕ pequeno (Mardia, 1972; pg.63),

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

Utilizando estas expansões, temos que fórmulas aproximadas para S_R^* e z_α têm as formas dadas no Exemplo 5.10 com

$$b(x, n, \phi, p) = \frac{1}{n} \left[\frac{1}{12p} \left\{ 2(3p^2 + 6p + 2) - 3 \frac{(p^2 + 60)}{\phi} \right\} - \frac{1}{6p(p+2)} \left\{ 2(6p + 11) - \frac{3(p+60)}{\phi} x + \frac{10}{3p(p+2)(p+4)} \left(1 - \frac{9}{2\phi} \right) x^2 \right\} \right]$$

para $\phi^{(0)}$ grande e

$$b(x, n, \phi, p) = \frac{1}{n} \left[\frac{1}{16p} \left\{ \frac{8}{\phi^2} p(p-2) + 3(p^2 + 2p + 4) \right\} - \frac{3(p+1)}{4p(p+2)} x \right]$$

para $\phi^{(0)}$ pequeno.

verto = O(1/2)

5.4.3. TESTES DE HIPÓTESES PARA β COM ϕ DESCONHECIDO

Consideremos as hipóteses $H_0 : \beta = \beta^{(0)}$ e $H : \beta \neq \beta^{(0)}$ onde $\beta^{(0)}$ é um vetor fixado. A estimativa de máxima verossimilhança de ϕ sob H_0 , $\tilde{\phi}$ digamos, vem da solução da equação (5.21) com θ_l avaliado em $\beta = \beta^{(0)}$. A estatística escore para este problema tem a forma

$$S_R = \tilde{S}^{*T} \tilde{W}^{*1/2} \tilde{X}^* (\tilde{X}^{*T} \tilde{W}^* \tilde{X}^*)^{-1} \tilde{X}^{*T} \tilde{W}^{*1/2} \tilde{S}^*, \quad (5.26)$$

(*) A expansão para $r(\phi)$ para ϕ grande é obtida das expansões de $I_1(\phi)$ e $I_0(\phi)$ dadas por Abramowitz e Stegun (1970, fórmula 9.7.1).

onde $S^* = (s_1^*, \dots, s_n^*)^T$ com $s_i^* = \phi^{1/2} t(y, \theta) \{-D_2(\theta, \phi)\}^{-\frac{1}{2}}$, e onde X^* e W^* foram definidos na Seção 5.4.1. As quantidades assinaladas com um til são avaliadas em $(\beta^T, \phi) = (\beta^{(0)T}, \tilde{\phi})$. Se $t(y, \theta) = y\theta - b(\theta) + c(y)$ e $\eta = X\beta$, (5.26) se reduz (5.14).

Na Seção 5.3, encontramos as quantidades A_1 , A_2 e A_3 referentes ao teste de $H_0 : \beta = \beta^{(0)}$ nos MLGs. Notamos que A_1 só depende do modelos através do posto p da matriz X e de derivadas da função $d_1(\cdot)$. O mesmo não ocorre com A_2 e A_3 que dependem fortemente do modelo assumido para os dados. Isto sugere que a expressão obtida para A_1 nos MLGs seja válida também para os modelos próprios de dispersão ao contrário das expressões de A_2 e A_3 . Mostramos, no Apêndice 4, que para os modelos próprios de dispersão

$$A_1 = \frac{6p(d_{(3)} + 2d_{(2)})}{nd_{(2)}^2}, \quad (5.27)$$

$$A_2 = \frac{3p(p+2)}{nd_{(2)}} + \frac{3}{\phi} \text{tr}(B^* Z_d^{*(2)}) \quad (5.28)$$

e

$$A_3 = \phi^{-1} 1^T C^* (3Z_d^* Z^* Z_d^* + 2Z^{*(3)}) C^* 1, \quad (5.29)$$

onde $B^* = \text{diag}\{b_1^*, \dots, b_n^*\}$, $C^* = \text{diag}\{c_1^*, \dots, c_n^*\}$ com

$$b^* = \left(\frac{d\theta}{d\eta} \right) (-3D_4 + 8D_3' - 6D_2'' + 3\phi D_2^{(2)} - 3\phi D_2^2), \quad c^* = \left(\frac{d\theta}{d\eta} \right)^3 (2D_3 - 3D_2')$$

e $Z^* = X^*(X^{*T}W^*X^*)^{-1}X^{*T}$ é, a menos da divisão pelo parâmetro ϕ , a estrutura de covariância assintótica de $\hat{\eta}$, isto é, do estimador de máxima verossimilhança de η . A notação introduzida no Capítulo 3 para matrizes do tipo Z_d e $Z^{(3)}$ é utilizada aqui. As quantidades D_i , $i = 2, 3, 4$, estão definidas em (5.20) e usamos ainda a notação: $D_i' = \partial D_i(\theta, \phi)/\partial \theta$, $i = 2, 3$, $D_2'' = \partial^2 D_2(\theta, \phi)/\partial \theta^2$ e $D_2^{(2)} = E\{(\partial^2 t(y, \theta)/\partial \theta^2)^2\}$.

Alguns comentários sobre os A 's merecem ser feitos. Primeiramente, nota-se que, de fato, a expressão para A_1 , que depende do modelo somente através do posto de X^* e de derivadas da função $d_1(\cdot)$, é idêntica àquela obtida para os MLGs. Por outro lado, as

expressões para A_2 e A_3 dependem fortemente do modelo assumido para os dados através das matrizes Z^* , B^* e C^* que são funções de alguns momentos de derivadas de $t(y, \theta)$ com respeito a θ , da função de ligação com sua primeira derivada e da matriz modelo local. É fácil observar que se $t(y, \theta) = y\theta - b(\theta) + c(y)$ e $\eta = X\beta$, A_2 e A_3 se reduzem a (5.16) e (5.17) respectivamente, pois, neste caso, $B^* = B$, $C^* = F - G$ e $Z^* = Z$. Dependendo do modelo associado aos dados, pode ser bastante difícil obter A_3 e a segunda parcela de A_2 pois estas dependem de momentos de derivadas de $t(y, \theta)$ com respeito a θ (quantidades D_2, D_3, D_4). É este o caso do modelo de McCullagh (1989) e dos modelos tratados no Exemplo 5.12. É claro que nas distribuições normal, normal inversa e gama este problema não ocorre. Mostramos no exemplo abaixo que, no modelo de von Mises, as quantidades D_2, D_3 e D_4 são obtidas sem dificuldade.

Exemplo 5.14

Retomamos o modelo de von Mises definido no Exemplo 5.13. Como $E(\sin(y - \theta)) = 0$, $E(\cos(y - \theta)) = r(\phi)$ e $E(\cos^2(y - \theta)) = 1 - r(\phi)/\phi$, temos $D_2 = -r(\phi)$, $D_3 = 0$, $D_4 = r(\phi)$ e $D_2^{(2)} = 1 - r(\phi)/\phi$. Logo, $b^* = 6r''(\phi) + 3\phi(1 - r(\phi)^2)$ e $c^* = 3r'(\phi)$. Mas, mostramos no Exemplo 5.13 que $r'(\phi)$ e $r''(\phi)$ podem ser escritas em função de ϕ e $r(\phi)$ através das funções $d_{(2)}$ e $d_{(3)}$. Assim, temos $b^* = -6d_{(3)}/\phi^3 + 3\phi(1 - r(\phi)^2)$ e $c^* = -3d_{(2)}/\phi^2$.

Para ilustrar a obtenção de A_3 e da segunda parcela de A_2 , isto é de $\text{tr}(B^* Z_d^{*(2)})$, tomemos um modelo simples, isto é, consideremos n variáveis aleatórias independentes e identicamente distribuídas e uma ligação identidade. Neste caso, a estatística escore para testar $H_0 : \theta = \theta^{(0)}$ versus $H : \theta \neq \theta^{(0)}$ é

$$S_R = \frac{\tilde{\phi} \{ \sum \cos(y_l - \theta^{(0)}) \}^2}{nr(\phi)},$$

onde $\tilde{\phi}$ é a solução de $r(\phi) = n^{-1} \sum \cos(y_l - \theta^{(0)})$. Agora, como $Z^* = 11^T(nw)^{-1}$ e $\omega = -D_2 = r(\phi)$, temos $\text{tr}(B^* Z_d^{*(2)}) = b^*/(nr(\phi)^2)$ e $A_3 = 5c^{*2}/(n\phi r(\phi)^3)$. Substituindo as expressões de b^* e c^* , chega-se a $\text{tr}(B^* Z_d^{*(2)}) = \{-6d_{(3)}/\phi^3 + 3\phi(1 - r(\phi)^2)\}/(nr(\phi)^2)$

e $A_3 = 45d_{(2)}^2/(n\phi^3r(\phi)^3)$. Fórmulas aproximadas para A_1 , A_2 e A_3 podem ser obtidas de maneira análoga ao que foi feito no Exemplo 5.12.

5.5. APERFEIÇOAMENTO DE TESTES ESCORE EM MODELOS NÃO-LINEARES DA FAMÍLIA EXPONENCIAL

5.5.1. INTRODUÇÃO

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$ com função densidade na família exponencial da forma dada em (3.1) com parâmetro de dispersão **conhecido** e igual para todas as observações. Fazemos aqui uma extensão dos resultados do Capítulo 3, referentes a aperfeiçoamento de testes sobre os parâmetros da regressão em MLGs, para uma classe mais ampla de modelos que chamamos de modelos não-lineares da família exponencial. Enquanto nos MLGs a função de ligação relaciona a média das observações a uma combinação **linear** dos parâmetros, consideramos aqui que a ligação tem a forma

$$d(\mu_l) = \eta_l = f(x_l, \beta),$$

onde $\mu = E(y)$, $d(\cdot)$ é uma função conhecida monótona e diferenciável, $\beta = (\beta_1, \dots, \beta_p)^T$ é um vetor de p ($p < n$) parâmetros desconhecidos, $f(\cdot, \cdot)$ é uma função diferenciável com respeito aos componentes de β e $x_l = (x_{l1}, \dots, x_{lp})^T$ é um vetor de constantes conhecidas associadas à l -ésima observação. A função $f(\cdot, \cdot)$ não é necessariamente linear em relação a $\beta_1, \dots, \beta_p$, porém, se $f(x_l, \beta) = x_l^T \beta$, os modelos definidos acima se reduzem aos MLGs. Para completar a definição dessa classe de modelos, assumimos que a matriz de derivadas $X^* = X^*(\beta) = \partial \eta / \partial \beta$ tem posto p para todo β .

Muitos modelos que aparecem nas aplicações são modelos não-lineares da família exponencial. O caso especial mais importante é o modelo normal não-linear. A literatura é extremamente vasta e os livros de Ratkowsky (1983) e Gallant (1987) apresentam várias referências. Um grande número de aplicações do modelo normal não-linear é proveniente

das áreas econômica e agrícola.

A função escore para β pode ser escrita como

$$U(\beta) = \phi X^{*T} W^{1/2} V^{-1/2} (y - \mu)$$

e a matriz de informação total de Fisher é

$$K = K(\beta) = \phi X^{*T} W X^*,$$

onde $W = \text{diag}\{w_1, \dots, w_n\}$, $V = \text{diag}\{V_1, \dots, V_n\}$ com $w_l = (d\mu/d\eta)_l^2/V_l$ e $V_l = (d\mu/d\theta)_l$ representando a função de variância. Utilizamos aqui a notação D^x para indicar que cada elemento da diagonal da matriz D está elevado à potência x . Particionamos o vetor β em $(\beta_1^T, \beta_2^T)^T$, onde $\beta_1 = (\beta_1, \dots, \beta_r)^T$ e $\beta_2 = (\beta_{r+1}, \dots, \beta_p)^T$ com β_1 e β_2 representando os vetores paramétricos de interesse e de perturbação, respectivamente. Particionando a matriz modelo local X^* segundo a partição de β , podemos escrever a inversa da matriz de informação total como

$$K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix},$$

onde $K^{11} = \phi^{-1}(R^{*T} W R^*)^{-1}$, $K^{12} = K^{21T} = -\phi^{-1}(R^{*T} W R^*)^{-1} C^{*T}$ e $K^{22} = \phi^{-1}\{(X_2^{*T} W X_2^*)^{-1} + C^*(R^{*T} W R^*)^{-1} C^{*T}\}$. Aqui, $R^* = X_1^* - X_2^* C^*$ com $C^* = (X_2^{*T} W X_2^*)^{-1} X_2^{*T} W X_1^*$ representando uma matriz $(p-r) \times r$ cujas colunas são vetores dos coeficientes da regressão normal linear das colunas de X_1^* sobre a matriz modelo X_2^* com matriz de pesos W .

A estimativa de máxima verossimilhança de β pode ser obtida numericamente pela solução iterativa das equações

$$X^{*(i)T} W^{(i)} X^{*(i)} \beta^{(i+1)} = X^{*(i)T} W^{(i)} y^{*(i)},$$

onde $y^* = X^* \beta + W^{1/2} V^{-1/2} (y - \mu)$. É fácil ver que este processo iterativo é um caso particular daquele descrito na Seção 5.4.1 para os modelos próprios de dispersão.

Nosso objetivo é desenvolver testes escore aperfeiçoados para as hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$. Uma fórmula, em notação matricial, para o fator de Bartlett da estatística da razão de verossimilhança para este problema foi determinada por Cordeiro e Paula (1989b) e vários casos especiais de interesse prático foram estudados (detalhes podem ser encontrados na tese de doutoramento de G.A.Paula, 1988).

A estatística escore para testar H_0 versus H tem a forma

$$S_R = \tilde{S}^T \tilde{W}^{1/2} \tilde{X}_1^* (\tilde{R}^{*T} \tilde{W} \tilde{R}^*)^{-1} \tilde{X}_1^{*T} \tilde{W}^{1/2} \tilde{S},$$

onde $S = (s_1, \dots, s_n)^T$, com $s_l = \phi^{1/2}(y_l - \mu_l)V_l^{-1/2}$, representa o vetor de resíduos de Pearson. As demais matrizes foram definidas anteriormente. Assinalamos com um til as quantidades avaliadas em $\beta = (\beta_1^{(0)T}, \tilde{\beta}_2^T)^T$, onde $\tilde{\beta}_2$ é o estimador de máxima verossimilhança de β_2 restrito a H_0 .

5.5.2. TESTES ESCORE APERFEIÇOADOS

Nesta seção desenvolvemos fórmulas matriciais para as quantidades A_1 , A_2 e A_3 que determinam os testes escore aperfeiçoados. As fórmulas podem ser de grande utilidade na obtenção de expressões algébricas simples para os A 's em modelos especiais. Ainda, apresentam vantagens na determinação numérica destas quantidades pois envolvem somente operações simples entre matrizes e vetores.

A notação utilizada aqui para cumulantes de derivadas da log-verossimilhança $L(\beta)$ segue a das seções e capítulos anteriores. Assim, por exemplo, $\kappa_{rs} = E(\partial^2 L / \partial \beta_r \partial \beta_s)$, $\kappa_{r,st} = E(\partial L / \partial \beta_r \partial^2 L / \partial \beta_s \partial \beta_t)$, etc. Entretanto, a notação introduzida na Seção 3.3 para certos produtos de derivadas de η em relação aos componentes de β precisa ser alterada. Adotamos, então, uma nova notação: $(r) = \partial \eta / \partial \beta_r$, $(r, s) = (\partial \eta / \partial \beta_r) \partial \eta / \partial \beta_s$, $(rs) = \partial^2 \eta / \partial \beta_r \partial \beta_s$, $(rs, t) = (\partial^2 \eta / \partial \beta_r \partial \beta_s) \partial \eta / \partial \beta_t$, etc.

O Apêndice 5 apresenta a obtenção dos cumulantes necessários ao cálculo dos A 's.

Temos,

$$\begin{aligned}
-\kappa_{r,s} &= \kappa_{rs} = -\phi \sum w_l(r,s)_l, \\
\kappa_{rst} &= -\phi \sum [(f+2g)(r,s,t) + w\{(r,st) + (s,rt) + (t,rs)\}]_l, \\
\kappa_{r,st} &= \phi \sum \{g(r,s,t) + w(r,st)\}_l, \\
\kappa_{r,s,t} &= \phi \sum (f-g)_l(r,s,t)_l, \\
\kappa_{r,s,t,u} &= \phi \sum b_l(r,s,t,u)_l, \\
\kappa_{r,s,tu} &= \phi \sum \left\{ \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} - \frac{1}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4 \right\}_l (r,s,t,u)_l \\
&\quad + \phi \sum (f-g)_l(r,s,tu)_l
\end{aligned}$$

e

$$\kappa_{r,s,t,u} + \kappa_{r,s,tu} = \phi \sum [h_l(r,s,t,u)_l + (f-g)_l(r,s,tu)_l],$$

onde f, g, h e b são dados em (3.4)–(3.7). Alguns destes cumulantes haviam sido obtidos anteriormente na tese de doutoramento de G.A.Paula (1988). O desenvolvimento de fórmulas matriciais para os A 's envolve grande trabalho algébrico cujos detalhes são apresentados no Apêndice 1. Dos resultados deste apêndice temos, $A_1 = L_1 + NL_1$, $A_2 = L_2 + NL_2$ com

$$\begin{aligned}
L_1 &= \phi^{-1} [3 \, 1^T F Z_{2d}^* (Z^* - Z_2^*) Z_{2d}^* F 1 + 6 \, 1^T F Z_{2d}^* Z_2^* (Z^* - Z_2^*)_d (F - G) 1 \\
&\quad - 6 \, 1^T (2G - F) \{Z_2^{*(2)} \otimes (Z^* - Z_2^*)\} F 1 - 6 \, 1^T H (Z^* - Z_2^*)_d Z_{2d}^* 1],
\end{aligned} \tag{5.30}$$

$$\begin{aligned}
NL_1 &= \phi^{-1} [6 \, 1^T W \{Q_2 (Z^* - Z_2^*) Z_{2d}^* - [(Z^* - Z_2^*) \otimes C_2]\} F 1 \\
&\quad + 3 \, 1^T W \{Q_2 (Z^* - Z_2^*) Q_2 + 2[(Z^* - Z_2^*) \otimes J_2]\} W 1 \\
&\quad + 6 \, 1^T W Q_2 Z_2^* (Z^* - Z_2^*)_d (F - G) 1 + 6 \, 1^T (2G - F) \{(Z^* - Z_2^*) \otimes C_2^T\} W 1 \\
&\quad - 6 \, 1^T (F - G) Q_2 (Z^* - Z_2^*)_d 1],
\end{aligned} \tag{5.31}$$

$$\begin{aligned}
L_2 &= \phi^{-1} [-3 \, 1^T (F - G) (Z^* - Z_2^*)_d Z_2^* (Z^* - Z_2^*)_d (F - G) 1 \\
&\quad - 6 \, 1^T F Z_{2d}^* (Z^* - Z_2^*) (Z^* - Z_2^*)_d (F - G) 1 \\
&\quad - 6 \, 1^T (F - G) \{(Z^* - Z_2^*)^{(2)} \otimes Z_2^*\} (F - G) 1 + 3 \, 1^T B (Z - Z_2^*)_d^{(2)} 1],
\end{aligned} \tag{5.32}$$

$$NL_2 = -6\phi^{-1} 1^T W Q_2 (Z^* - Z_2^*)(Z^* - Z_2^*)_d (F - G) 1 \quad (5.33)$$

e

$$\begin{aligned} A_3 = & \phi^{-1} \{ 3 1^T (F - G) (Z^* - Z_2^*)_d (Z^* - Z_2^*) (Z^* - Z_2^*)_d (F - G) 1 \\ & + 2 1^T (F - G) (Z^* - Z_2^*)^{(3)} (F - G) 1 \}, \end{aligned} \quad (5.34)$$

onde F , G , H e B são matrizes diagonais com elementos típicos dados em (3.4)–(3.7).

Aqui, $Z^* = X^*(X^{*T}WX^*)^{-1}X^{*T}$, $Z_2^* = X_2^*(X_2^{*T}WX_2^*)^{-1}X_2^{*T}$, $Q_2 = \text{diag}\{q_1, \dots, q_n\}$

com $q_l = \text{tr}((X_2^{*T}WX_2^*)^{-1}D_{22}^{(l)})$ sendo $D_{22}^{(l)}$ obtida da matriz

$$D^{(l)} = \left\{ \frac{\partial^2 \eta_l}{\partial \beta_i \partial \beta_j} \right\} = \begin{pmatrix} D_{11}^{(l)} & D_{12}^{(l)} \\ D_{21}^{(l)} & D_{22}^{(l)} \end{pmatrix}$$

particionada segundo a partição de β . Definimos ainda $C_2 = \{c_{lm}\}$ com $c_{lm} = x_{2m}^*(X_2^{*T}WX_2^*)^{-1}D_{22}^{(l)}(X_2^{*T}WX_2^*)^{-1}x_{2m}^{*T}$ sendo que x_{2m}^* é obtida da m -ésima linha da matriz X^* , isto é, do vetor linha $x_m^* = (x_{1m}^*, x_{2m}^*)$ também particionado segundo a partição de β . Finalmente, definimos $J_2 = \{j_{lm}\}$ com

$$j_{lm} = \text{tr}(D_{22}^{(l)}(X_2^{*T}WX_2^*)^{-1}D_{22}^{(m)}(X_2^{*T}WX_2^*)^{-1}).$$

A notação para o produto direto $\otimes$ e para matrizes do tipo $Z^{*(2)}$, $Z^{*(3)}$ e Z_d^* segue aquela associada às fórmulas dos A 's em (3.9)–(3.11) válidas para os MLGs. As quantidades L_1 , L_2 e A_3 são iguais, respectivamente, às fórmulas de A_1 , A_2 e A_3 desenvolvidas para os MLGs com Z e Z_2 substituídos por Z^* e Z_2^* . Por outro lado, NL_1 e NL_2 refletem a não-linearidade do modelo no sentido de que em modelos lineares estas quantidades se anulam.

O cálculo de $Z^* - Z_2^*$ envolve a inversa de uma matriz de dimensão $p \times p$. Entretanto, é possível escrever $Z^* - Z_2^*$ como $R^*(R^{*T}WR^*)^{-1}R^{*T}$ que depende somente da inversa de uma matriz de dimensão $r \times r$ com $r \leq p$ (vide Seção 3.5).

Uma vez que os A 's possam ser calculados a partir das fórmulas (5.30)–(5.34), um teste escore aperfeiçoado pode ser conduzido de duas maneiras: utilizando a estatística escore modificada S_R^* da forma (2.15) e a distribuição χ_r^2 como referência, ou usando a

própria estatística escore S_R em conexão com as expansões assintóticas em (2.5) e (2.19) para a função distribuição de S_R e de seus quantis.

5.5.3. ALGUNS CASOS ESPECIAIS

Nesta seção apresentamos alguns casos especiais que produzem simplificações nas fórmulas (5.30)–(5.34).

Inicialmente, consideremos uma estrutura linear $\eta_l = x_l^T \beta$, que corresponde a assumir um MLG para os dados. Neste caso, é fácil ver que $Z^* = Z$ e $Z_2^* = Z_2$, onde Z e Z_2 foram definidos na Seção 3.3, e que as matrizes D , Q_2 , C_2 e J_2 se anulam. Portanto, segue imediatamente que os A 's se reduzem às fórmulas (3.9)–(3.11) desenvolvidas para os MLGs.

Para hipóteses nulas simples, temos $Z_2^* = Q_2 = C_2 = J_2 = 0$. Logo, A_1 se anula e A_2 e A_3 se reduzem a

$$A_2 = 3\phi^{-1}1^T B Z_d^{*(2)} 1$$

e

$$A_3 = \phi^{-1}1^T (F - G) \{3Z_d^* Z^* Z_d^* + 2Z^{*(3)}\} (F - G) 1.$$

Para o importante caso dos modelos normais não-lineares obtemos: $H = B = 0$, $F = G = \text{diag}\{f_1, \dots, f_n\}$ com $f = (d\mu/d\eta)d^2\mu/d\eta^2$ e $W = \text{diag}\{w_1, \dots, w_n\}$ com $w = (d\mu/d\eta)^2$ levando a

$$\begin{aligned} A_1 = & \phi^{-1} [3 \ 1^T F Z_{2d}^* (Z^* - Z_2^*) Z_{2d}^* F 1 - 6 \ 1^T F \{Z^{*(2)} \otimes (Z^* - Z_2^*)\} F 1 \\ & + 6 \ 1^T W [Q_2 (Z^* - Z_2^*) Z_{2d}^* - \{(Z^* - Z_2^*) \otimes C_2\}] F 1 \\ & + 3 \ 1^T W [Q_2 (Z^* - Z_2^*) Q_2 + 2\{(Z^* - Z_2^*) \otimes J_2\}] W 1 \\ & + 6 \ 1^T F \{(Z^* - Z_2^*) \otimes C_2^T\} W 1] \end{aligned}$$

e $A_2 = A_3 = 0$. Em particular, se a ligação for identidade, $F = G = 0$, $W = I$ e A_1 se reduz a $A_1 = 3\phi^{-1}1^T [Q_2 (Z^* - Z_2^*) Q_2 + 2\{(Z^* - Z_2^*) \otimes J_2\}] 1$. É interessante notar que,

para a hipótese nula simples $H_0 : \beta = \beta^{(0)}$, $A_1 = A_2 = A_3 = 0$ pois, neste caso, S_R tem exatamente uma distribuição χ_p^2 sob H_0 .

Passamos agora a analisar os modelos parcialmente lineares. Suponhamos que $d(\mu) = f(x_1, \beta_1) + x_2^T \beta_2$, onde $x_1 = (x_1, \dots, x_r)^T$ e consideremos o teste de $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$. Aqui, $Z_2^* = Z_2$ e

$$D^{(l)} = \begin{pmatrix} D_{11}^{(l)} & 0 \\ 0 & 0 \end{pmatrix}.$$

A forma da matriz $D^{(l)}$ conduz a $Q_2 = C_2 = J_2 = 0$. Assim, os A 's podem ser obtidos a partir das fórmulas (3.9)–(3.11) referentes aos MLGs, isto é, como se a ligação fosse $d(\mu) = x_1^T \beta_1 + x_2^T \beta_2$, bastando substituir Z por Z^* . Finalmente, suponhamos que a função de ligação tenha a forma $d(\mu) = f(x_1^T \beta_1) + x_2^T \beta_2$. Já vimos que $Q_2 = C_2 = J_2 = 0$ e que $Z_2^* = Z_2$. Além disto, temos também

$$\frac{\partial \eta}{\partial \beta_i} = \begin{cases} x_i f'(x_1^T \beta_1^{(0)}), & i = 1, \dots, r \\ x_i, & i = r + 1, \dots, p. \end{cases}$$

Isto induz a várias simplificações: $X^* = (f'(x_1^T \beta_1^{(0)})X_1 \quad X_2)$, $R^* = f'(x_1^T \beta_1^{(0)})R$ e $Z_2^* = Z_2$, onde X_1 , X_2 , R e Z_2 foram definidos nas Seções 3.2 e 3.3. Portanto, os A 's vêm diretamente das fórmulas (3.9)–(3.11) válidas para os MLGs. Isto decorre do fato de que a estatística escore para este problema não depende da função $f(\cdot)$ (Chen, 1985).

Em trabalho futuro, pretendemos explorar mais os resultados da Seção 5.5.2 enfocando outros casos particulares como, por exemplo, os testes de diagnóstico sugeridos por Davison e Tsai (1990, Seção 3), analisando aspectos computacionais e realizando simulações para verificar o desempenho dos testes escore modificados em amostras pequenas e moderadas. Além disto, com base no trabalho de Cordeiro, Paula e Botter (1991), temos intenção de estender os resultados à classe dos modelos de dispersão.

CAPÍTULO 6

FATORES DE CORREÇÃO DE BARTLETT GENERALIZADOS

6.1. INTRODUÇÃO

No Capítulo 2, apresentamos a expansão assintótica até ordem n^{-1} da função distribuição da estatística escore sob a hipótese nula H_0 . Com base nesta expansão, derivamos uma estatística escore modificada por um fator de correção, cuja distribuição sob H_0 é qui-quadrado até esta ordem. Mostramos ainda que é possível determinar, diretamente da forma do fator de correção, a expansão sob H_0 e até ordem n^{-1} dos quantis da estatística escore.

O objetivo deste capítulo é estender estes resultados à classe de estatísticas cujas distribuições assintóticas são qui-quadrados. Em particular, estamos interessados em desenvolver o que chamamos de “fatores de correção de Bartlett generalizados” para as estatísticas desta classe. Tratamos o problema sob dois enfoques distintos. Primeiramente,

supomos conhecida a expansão assintótica da distribuição da estatística e combinamos as expansões (1) e (2) de Cox e Reid (1987a; vide Apêndice 1) para derivar estatísticas modificadas e obter diretamente expansões assintóticas para os quantis das correspondentes estatísticas não modificadas. Este enfoque é particularmente útil para as estatísticas da razão de verossimilhança, de Wald e escore ou quaisquer outras assintoticamente equivalentes a estas cujas expansões assintóticas sejam conhecidas (Hayakawa, 1977; Hayakawa e Puri, 1985 e Harris, 1985). A técnica usual para a obtenção destas expansões para uma estatística S tendo distribuição χ^2 em grandes amostras é baseada no trabalho desenvolvido por Peers (1971) e estendido por Hayakawa (1977). O termo de ordem n^{-1} das expansões são, em geral, funções complicadas de cumulantes conjuntos de derivadas da log-verossimilhança envolvendo parâmetros desconhecidos. Sob o segundo enfoque, supomos conhecidas as expansões dos $k_0(k_0 \geq 1)$ primeiros momentos ordinários da estatística e derivamos uma estatística modificada tendo seus k_0 primeiros momentos coincidindo, até ordem n^{-1} , com os correspondentes da distribuição χ^2 de referência. Mostramos ainda que, dependendo do valor escolhido para k_0 , a distribuição da estatística modificada coincide com a da χ^2 até esta ordem. Embora os dois enfoques estejam intimamente relacionados, um pode ser preferível ao outro em um particular problema. Em geral, a obtenção de expansões para distribuições é bastante trabalhosa enquanto que, em alguns casos especiais, expansões para os primeiros momentos de uma estatística são obtidas sem dificuldade.

6.2. ESTATÍSTICAS APERFEIÇOADAS TENDO DISTRIBUIÇÃO χ^2 ATÉ ORDEM n^{-1}

6.2.1. OBTENÇÃO DAS ESTATÍSTICAS APERFEIÇOADAS

Sob certas condições de regularidade, para uma estatística S com densidade contínua e cuja distribuição assintótica é qui-quadrado central com r graus de liberdade, podemos escrever sua função distribuição até ordem n^{-1} como uma combinação linear

finita de qui-quadrados centrais com graus de liberdade r , $r + 2$, $r + 4$, etc. Detalhes sobre as condições para que tal expansão seja válida são encontrados no artigo de Chandra (1985; vide também Chandra e Ghosh, 1979). A função distribuição de S , na qual termos de ordem inferior a n^{-1} são omitidos, pode ser escrita como

$$F_S(x) = G_r(x) + \sum_{l=0}^k a_l G_{r+2l}(x), \quad (6.1)$$

onde os a_l 's são $O(n^{-1})$ e, em geral, funções de parâmetros desconhecidos. Aqui usamos a mesma notação do Capítulo 2 em que G_m e g_m representam as funções distribuição e densidade da χ_m^2 respectivamente. Para que a função F_S em (6.1) seja uma função distribuição até esta ordem é necessário que a condição $\sum_{l=0}^k a_l = 0$ seja satisfeita.

Das relações de recorrência (2.18), podemos escrever (6.1) como

$$F_S(x) = G_r(x) - g_r(x) \sum_{i=1}^k C_i x^i, \quad (6.2)$$

onde $C_i = 2\mu_i'^{-1} \sum_{l=i}^k a_l$, para $i = 1, \dots, k$, são de ordem n^{-1} e $\mu_i' = 2^i \Gamma(\frac{r}{2} + i) / \Gamma(\frac{r}{2})$ é o i -ésimo momento ordinário da distribuição χ_r^2 .

A forma funcional da função distribuição de S envolvendo um polinômio de grau k sugere a estatística modificada

$$S^* = S \left(1 - \sum_{i=1}^k c_i S^{i-1} \right). \quad (6.3)$$

A idéia é determinar os c_i 's de ordem n^{-1} em função dos C_i 's de tal forma que S^* tenha distribuição χ_r^2 até esta ordem sob a hipótese nula.

Da expansão (1) de Cox e Reid (1987a; vide Apêndice 1) vem a função distribuição de S^* , até ordem n^{-1} ,

$$F_{S^*}(x) = F_S(x) + f_S(x) \sum_{i=1}^k c_i x^i.$$

Uma vez que os c_i 's são $O(n^{-1})$ e, até primeira ordem, S tem distribuição χ_r^2 , chega-se a

$$F_{S^*}(x) = F_S(x) + g_r(x) \sum_{i=1}^k c_i x^i \quad (6.4)$$

até ordem n^{-1} . Agora combinando (6.2) e (6.4) vem, até esta ordem,

$$F_{S^*}(x) = G_r(x) + g_r(x) \sum_{i=1}^k (c_i - C_i) x^i.$$

Portanto, $F_{S^*}(x) = G_r(x)$ até ordem n^{-1} se e somente se $c_i = C_i$, para $i = 1, \dots, k$.

Consequentemente, a estatística modificada

$$S^* = S \left\{ 1 - 2 \sum_{i=1}^k \left(\sum_{l=i}^k a_l \right) \mu_i'^{-1} S^{i-1} \right\} \quad (6.5)$$

tem distribuição χ_r^2 até ordem n^{-1} sob H_0 . O termo entre chaves é um tipo de correção de Bartlett envolvendo um polinômio na própria estatística. Portanto, acreditamos ser adequado denominá-lo **fator de correção de Bartlett generalizado**. É importante notar que para as estatísticas cujas distribuições assintóticas, até ordem n^{-1} , são conhecidas, os coeficientes do polinômio são construídos diretamente dos coeficiente envolvidos na própria distribuição assintótica.

Este resultado é uma ferramenta poderosa no aperfeiçoamento de amplas classes de estatísticas. Basta dizer que as estatísticas escore e da razão de verossimilhança aperfeiçoadas são obtidas diretamente de (6.5). Outras aplicações podem incluir a estatística de Wald e uma versão modificada desta estatística dada em Hayakawa e Puri (1985) e ainda diversas formas assintoticamente equivalentes à estatística escore (Bera e Mckenzie, 1986).

Agora, a expansão (2) de Cox e Reid (1987a; vide Apêndice 1) e o fato de S_R^* ter distribuição χ^2 até ordem n^{-1} fornecem a expansão assintótica para os quantis de S sob H_0 . Ignorando termos de ordem inferior a n^{-1} , chega-se a

$$z = x \left\{ 1 + 2 \sum_{i=1}^k \left(\sum_{l=i}^k a_l \right) \mu_i'^{-1} x^{i-1} \right\},$$

com $P(S \leq z) = G_r(x)$. Assim, estabelece-se uma estreita conexão entre o fator de correção da estatística S e o de seus quantis.

6.2.2. CASOS ESPECIAIS

Consideremos o problema descrito no Capítulo 2 em que o interesse é testar as hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$ sendo $\beta_1^{(0)}$ um vetor de dimensão r fixado. O vetor paramétrico β de dimensão p ($p \geq r$) é particionado na forma $\beta = (\beta_1^T, \beta_2^T)^T$ sendo β_2 um parâmetro de perturbação.

A expansão até ordem n^{-1} da função distribuição da estatística da razão de verossimilhança ω para o teste das hipóteses acima foi obtida por Hayakawa (1977; vide também Cordeiro, 1987) e é dada por (6.1) com $a_0 = -a_1 = b/2$, onde b é o termo de ordem n^{-1} da esperança de ω . De (6.5) com $k = 1$, vem a estatística aperfeiçoada $\omega^* = \omega(1 - b/r)$, sendo que a expressão entre parêntesis é o conhecido fator de correção de Bartlett da estatística da razão de verossimilhança.

Uma estatística score aperfeiçoada para este problema é também obtida diretamente dos resultados da seção anterior. Da expansão assintótica obtida por Harris (1985) e dada em (2.5) e da fórmula (6.5) com $k = 3$, chega-se à estatística S_R^* derivada no Capítulo 2.

Outras aplicações interessantes poderiam incluir a estatística Wald e sua versão modificada proposta por Hayakawa e Puri (1985). Estes autores encontraram as expansões assintóticas das distribuições destas estatísticas que, no entanto, parecem estar incorretas. Na Seção 6.5, outras aplicações serão apresentadas.

6.3. FATORES DE CORREÇÃO DE BARTLETT GENERALIZADOS DESENVOLVIDOS A PARTIR DE MOMENTOS DAS ESTATÍSTICAS NÃO MODIFICADAS

6.3.1. OBTENÇÃO DE FATORES DE CORREÇÃO DE BARTLETT GENERALIZADOS

Nesta seção, buscamos uma forma alternativa de determinar $c_1, \dots, c_k$ em (6.3) a

fim de obter uma estatística aperfeiçoada tendo distribuição χ_r^2 até ordem n^{-1} . Da forma de (6.3) e lembrando que os c 's são $O(n^{-1})$, vemos que

$$S^{*j} = S^j - j S^{j-1} \sum_{i=1}^k c_i S^i, \quad j = 1, 2, \dots,$$

onde termos de ordem estocástica inferior a n^{-1} são ignorados. A idéia é tomar a esperança em ambos os membros destas igualdades para obter um sistema de equações cujas incógnitas são os c 's e os coeficientes e os termos independentes são funções dos k primeiros momentos de S e da distribuição χ_r^2 .

Admitindo que os c 's sejam tais que S^* tenha distribuição χ_r^2 até ordem n^{-1} e que os momentos de S sejam $O(1)$, temos até esta ordem

$$\mu'_j = m'_j - j \sum_{i=1}^k c_i m'_{i+j-1}, \quad j = 1, 2, \dots,$$

onde μ'_j e m_j são os momentos ordinários de ordem j da distribuição χ_r^2 e de S respectivamente. Estas equações conduzem a

$$\sum_{i=1}^k c_i m'_{i+j-1} = \frac{m'_j - \mu'_j}{j}, \quad j = 1, 2, \dots.$$

Podemos ver de (6.1) que, até primeira ordem, $m'_{i+j-1} = \mu'_{i+j-1}$. Como, além disso, os c_i 's são $O(n^{-1})$ temos, até ordem n^{-1} ,

$$\sum_{i=1}^k c_i \mu'_{i+j-1} = \frac{b_j}{j}, \quad j = 1, 2, \dots, \quad (6.6)$$

onde $b_j = m'_j - \mu'_j$ é a parcela de ordem n^{-1} da expansão de j -ésimo momento ordinário de S . As k primeiras equações de (6.6) determinam $c_1, \dots, c_k$, ou seja, os c 's são a única solução de

$$M c = \bar{b}, \quad (6.7)$$

onde $M = \{m_{ij}\}$ é uma matriz $k \times k$ com elemento típico $m_{ij} = \mu'_{i+j-1}$, $i, j = 1, \dots, k$, e $\bar{b}^T = (b_1, b_2/2, \dots, b_k/k)$ é um vetor de dimensão k . Na realidade, isto é válido pois a matriz M é inversível. Este fato será demonstrado adiante.

É importante notar que, se expansões até ordem n^{-1} para os k primeiros momentos de uma estatística S forem conhecidas, a estatística S^* aperfeiçoada tendo esta distribuição até ordem n^{-1} é obtida diretamente de (6.3) com os c 's dados pela solução do sistema de equações em (6.7). Além disto, a própria expansão da função distribuição de S pode ser obtida pois os a_l 's são funções unívocas de $c_1, \dots, c_k$ como mostraremos na Seção 6.4.

Para evitar o custo computacional da inversão da matriz M , passamos a buscar uma solução explícita para este sistema. O momento ordinário de ordem $i + j - 1$ da distribuição χ_r^2 é dado por $\mu'_{i+j-1} = r(r+2) \cdots \{r+2(j+i-2)\}$ que, escrito de forma conveniente, torna-se

$$\mu'_{i+j-1} = \mu'_j (r+2j)^{(i-1)},$$

onde usamos a notação: $(a)^{(b)} = a(a+2) \cdots \{a+2(b-1)\}$, para $b \geq 1$, inteiro e $(a)^{(0)} = 1$.

Assim, o sistema de equações em (6.7) pode ser reescrito como

$$\sum_{i=1}^k c_i (r+2j)^{(i-1)} = \frac{b_j}{j \mu'_j}, \quad j = 1, \dots, k.$$

Seja $d_i = b_i / (i \mu'_i)$, isto é, $d_i = (m'_i - \mu'_i) / (i \mu'_i)$. O sistema de equações acima pode ser dado na forma matricial

$$P \underset{c = (c_1, \dots, c_k)^T}{c} = d$$

onde $P = \{p_{ij}\}$ com $p_{ij} = (r+2i)^{(j-1)}$, $i, j = 1, \dots, k$, e $d^T = (d_1, \dots, d_k)$.

Fazemos agora uma série de operações elementares entre as linhas da matriz e os correspondentes elementos do vetor d . Consideramos a matriz estendida $E = (P \ d)$ e procedemos às seguintes operações: (i) divide-se a i -ésima linha de E por $r+2(i-1)$, $i = 1, \dots, k$, de forma que a primeira coluna da matriz resultante será um vetor de uns; (ii) subtrai-se da primeira linha as demais, transformando a primeira coluna em $(1, 0, \dots, 0)^T$; (iii) agora a segunda coluna é $(r+2, 2, \dots, 2)^T$ e dividem-se os elementos de todas as linhas, com exceção da primeira, por 2; (iv) procedimento semelhante é repetido partindo-se da segunda linha e assim por diante. Ao final, chega-se ao sistema de equações equivalente

$$Q \ c = \bar{d},$$

onde $Q = \{q_{i,j}\}$ com

$$q_{i,j} = \begin{cases} \binom{j-1}{i-1}(r+2i)^{(j-i)}, & i \leq j \\ 0, & i > j \end{cases} \quad (6.8)$$

$i, j = 1, \dots, k$, e $\bar{d}^T = (\bar{d}_1, \dots, \bar{d}_k)$ com

$$\bar{d}_i = \frac{1}{2^{i-1}i!} \sum_{j=1}^k \binom{i}{j} (-1)^{i-j} \frac{\overbrace{m'_j - \mu'_j}^b}{\mu'_j}, \quad i = 1, \dots, k. \quad (6.9)$$

Assim, os c 's são dados por

$$c_i = \bar{d}_i - \sum_{j=i+1}^k c_j q_{i,j}, \quad i = 1, \dots, k-1$$

e

$$c_k = \bar{d}_k.$$

Após um longo desenvolvimento algébrico que omitimos aqui, chega-se a

$$c_i = \sum_{j=0}^{k-i} (-1)^j \bar{d}_{i+j} q_{i,i+j}, \quad (6.10)$$

para $i = 1, \dots, k$ onde $q_{i,j}$ e d_i são dados em (6.8) e (6.9) respectivamente. Esta é a única solução do sistema de equações (6.7). Logo, de fato, a matriz M do sistema em (6.7) é inversível e a sua inversa é encontrada, após alguma álgebra, como $M^{-1} = \{m^{ij}\}$, $i, j = 1, \dots, k$, com

$$m^{ij} = \frac{(-1)^{i-j}}{(i-1)!(j-1)!(r)^{(j)}} \sum_{t=\max\{i,j\}}^k \frac{(t-1)!}{(t-i)!(t-j)!} \frac{(r+2i)^{(t-i)}}{2^{t-1}}.$$

Em resumo, uma vez encontradas as expansões até ordem n^{-1} dos k primeiros momentos de S sob H_0 , a estatística aperfeiçoada S^* segue de (6.3) com os c 's em (6.10). Esta parece ser uma ferramenta poderosa na busca de aperfeiçoamento de estatísticas que têm distribuição assintótica χ^2 pois, em diversas situações, as expansões assintóticas de momentos de uma particular estatística são desenvolvidas sem dificuldade.

Na realidade, estamos agora face a um problema. O número k de momentos que devem ser usados no sistema de equações depende do número de parcelas da combinação

linear de qui-quadrados no termo de ordem n^{-1} da expansão da distribuição de S dada em (6.1). Se tal expansão não for conhecida como saber qual é o valor de k ? Claramente, para a razão de verossimilhança $k = 1$ e para a estatística escore $k = 3$. Quantos momentos de S devem ser obtidos para determinar os c 's corretamente?. Em outras palavras, quantas equações deve ter o sistema? Este, por enquanto, é um problema em aberto. Porém, algumas considerações podem ser feitas. Primeiramente, notamos que se tomarmos um sistema com, digamos, k_0 equações sendo $k_0 > k$, isto não causará problema algum uma vez que a solução do sistema (6.3) produzirá $c_{k+1} = \dots = c_{k_0} = 0$, pois $\bar{d}_j = 0$, para $j = k + 1, \dots, k_0$, e, além disto, $c_1, \dots, c_k$ determinados por este sistema coincidirão com aqueles do sistema com exatamente k equações. Segundo, se for tomado um valor $k_0 < k$, a estatística S^* obtida não terá necessariamente distribuição χ_r^2 até ordem n^{-1} , no entanto, terá seus k_0 primeiros momentos coincidindo com os desta distribuição. Isto deve produzir um aperfeiçoamento em relação à estatística não modificada. É importante observar que diversos autores procuram aperfeiçoar estatísticas através da correção de um ou dois momentos (vide, por exemplo, Cressie e Read, 1984 e Hayakawa, 1976). Segundo o método que propomos, é possível corrigir quantos momentos se desejar, bastando encontrar expansões para igual número de momentos da estatística.

As idéias desta seção podem ser estendidas a outras classes de estatísticas. Por exemplo, suponhamos que S tenha distribuição assintótica normal padrão $N(0, 1)$. Tipicamente, a expansão para a função distribuição de S é formada por parcelas de ordem 1, $n^{-1/2}$, n^{-1} , $n^{-3/2}$, etc. Propomos, então, uma estatística modificada S^* da forma (6.3) e buscamos encontrar $c_1, \dots, c_k$, de ordem $n^{-1/2}$, de tal forma que os k primeiros momentos de S^* coincidam com os correspondentes da distribuição $N(0, 1)$ até esta ordem. Procedendo de forma análoga ao desenvolvimento do início desta seção, chega-se ao sistema de equações (6.7) com μ'_l , $l = 1, \dots, 2k - 1$, representando o l -ésimo momento ordinário desta distribuição, ou seja, $\mu'_l = 0$ se l for ímpar e $\mu'_l = (l - 1)(l - 3) \dots 3 \cdot 1$, caso contrário. Assim, a matriz do sistema é $M = \{m_{ij}\}$ com $m_{ij} = 0$ se $i + j$ for par e

$m_{ij} = (i + j - 2)(i + j - 4) \cdots 3 \cdot 1$, caso contrário. Claramente, se k for ímpar, a matriz M não é inversível e, neste caso, não é possível obter uma estatística S^* da forma (6.3) com k momentos coincidindo com os da distribuição $N(0, 1)$ até ordem $n^{-1/2}$. Por outro lado, se k for par, é necessário verificar se M é inversível pois, neste caso, o sistema em (6.7) terá solução única e, portanto, a estatística modificada S^* poderá ser encontrada.

A viabilidade de se aperfeiçoar uma estatística da forma proposta aqui está intimamente relacionada com o fato da matriz M do sistema de equações ser inversível. Como M é uma matriz de Hankel ^(*) (Iochivdov, 1982), as propriedades deste tipo de matriz podem ser úteis na verificação da inversibilidade de M para uma particular classe de estatísticas.

6.3.2. CASOS ESPECIAIS

Consideremos novamente o problema de testes de hipóteses descrito no Capítulo 2. O primeiro momento da estatística da razão de verossimilhança ω para teste da hipótese nula H_0 é, até ordem n^{-1} , $r(1 + b/r)$ onde b é um termo desta ordem que é função, em geral, de parâmetros desconhecidos (Lawley, 1956). Do sistema de equações (6.7) com $k = 1$, isto é, com uma única equação, tem-se $c_1 = b/r$ levando à estatística aperfeiçoada $\omega^* = \omega(1 - b/r)$ como era esperado. Como a expansão da distribuição de ω até ordem n^{-1} é dada em (6.1) com $k = 1$, o aperfeiçoamento não só produz o ajuste do primeiro momento, como também faz com que ω^* tenha distribuição χ_r^2 até ordem n^{-1} . Já no caso da estatística escore S_R , deve-se tomar $k = 3$ para se obter uma estatística aperfeiçoada com esta propriedade. Substituindo os três primeiros momentos de S_R dados em (2.10) na expressão (6.9) tem-se $\bar{d}_1 = A_1/(12r)$, $\bar{d}_2 = A_2/\{12r(r+2)\}$ e $\bar{d}_3 = A_3/\{12r(r+2)(r+4)\}$. Agora, de (6.10) vem $c_1 = (A_1 - A_2 + A_3)/(12r)$, $c_2 = (A_2 - 2A_3)/\{12r(r+2)\}$ e $c_3 = A_3/\{12r(r+2)(r+4)\}$ produzindo a estatística escore aperfeiçoada desenvolvida no

(*) Uma matriz $A = \{a_{ij}\}$, $i, j = 1, \dots, n$, é dita ser de Hankel se $a_{ij} = b_{i+j}$, para quaisquer $b_1, \dots, b_{2n}$.

Capítulo 2.

Na Seção 6.5 outras aplicações serão apresentadas.

6.4. OBTENÇÃO DE EXPANSÃO ASSINTÓTICA DA DISTRIBUIÇÃO DE UMA ESTATÍSTICA A PARTIR DE SEUS MOMENTOS

Nas seções anteriores, propusemos dois métodos para aperfeiçoar estatísticas que têm distribuição χ^2 assintoticamente. O primeiro é baseado na expansão da distribuição da estatística, enquanto que o segundo parte de expansões de um certo número de seus momentos. Os dois métodos são equivalentes no sentido de que resultam na mesma estatística aperfeiçoada. Admite-se aqui que o número de expansões de momentos coincide com o número k de a_l 's desconhecidos em (6.1). Vejamos porque isto ocorre. Na Seção 6.2, mostramos que os coeficientes envolvidos na estatística aperfeiçoada S^* são

$$c_i = 2 \left(\sum_{l=i}^k a_l \right) \mu_i^{-1}, \quad i = 1, \dots, k, \quad (6.11)$$

onde os a 's determinam a expansão da distribuição (6.1) da estatística não modificada S . Substituindo os c 's nas k primeiras equações de (6.6) vem

$$\sum_{i=1}^k \sum_{l=i}^k \frac{a_l \mu'_{i+j-1}}{\mu'_i} = \frac{b_j}{2j}, \quad j = 1, \dots, k.$$

Agora, trocando a ordem dos somatórios, chega-se a

$$\sum_{l=1}^k a_l \left(\sum_{i=1}^l \frac{\mu'_{i+j-1}}{\mu'_i} \right) = \frac{b_j}{2j}, \quad j = 1, \dots, k$$

que, em notação matricial, torna-se

$$B a = e, \quad (6.12)$$

onde $B = \{b_{jl}\}$ com $b_{jl} = \sum_{i=1}^l \mu'_{i+j-1}/\mu'_i$, $j, l = 1, \dots, k$, $a = (a_1, \dots, a_k)^T$ e $e = (b_1/2, b_2/4, \dots, b_k/(2k))^T$ com b_j representando o termo de ordem n^{-1} da expansão de j -ésimo momento ordinário de S sob a hipótese nula H_0 . A matriz B é inversível como

mostraremos a seguir, de forma que os a 's são determinados como a única solução do sistema de equações em (6.12). Portanto, dados os k primeiros momentos de S até ordem n^{-1} , pode-se reproduzir a expansão da sua distribuição até esta ordem.

A recíproca é evidentemente verdadeira pois tem-se, diretamente de (6.1), que o j -ésimo momento ordinário de S é $m_j = \mu'_j + \sum_{l=0}^k a_l(r+2l)^{(j)}$, $j = 1, \dots, k$.

Para evitar a busca da solução do sistema de equações (6.11) através da inversão da matriz B , podemos exprimir os a 's diretamente em função dos c 's. Partindo de (6.11) chega-se, após alguma álgebra, a

$$a_l = \frac{1}{2}(c_l \mu'_l - c_{l+1} \mu'_{l+1}), \quad l = 1, \dots, k-1,$$

e

$$a_k = \frac{1}{2} c_k \mu'_k. \quad (6.13)$$

Como os a 's são determinados como a única solução de um sistema de equações, mostra-se que (6.12) também tem somente uma solução e que, portanto, B é inversível.

6.5. APLICAÇÕES

Apresentamos a seguir algumas aplicações das técnicas de obtenção de estatísticas aperfeiçoadas desenvolvidas neste capítulo. O primeiro exemplo já foi tratado anteriormente e é revisto aqui a título de ilustração. Como este trata de um teste escore, o sistema que produz os coeficientes da estatística aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} deve ter $k = 3$ equações implicando na necessidade de se derivar os três primeiros momentos da estatística não modificada até esta ordem.

Exemplo 6.1

Como uma primeira aplicação, consideremos uma amostra aleatória $y_1, \dots, y_n$ de uma distribuição exponencial com média μ . No Exemplo 2.1, exibimos a estatística escore para testar as hipóteses $H_0 : \mu = \mu^{(0)}$ versus $H : \mu \neq \mu^{(0)}$, onde $\mu^{(0)}$ é um valor

fixado, e demostramos que os seus três primeiros momentos sob H_0 são, respectivamente, 1 , $3 + 6/n$ e $15 + 130/n$ até ordem n^{-1} . Substituindo-os em (6.9) vem $\bar{d}_1 = 0$, $\bar{d}_2 = 1/(2n)$ e $\bar{d}_3 = 1/(9n)$ que, combinados com (6.10), levam a $c_1 = 1/(6n)$, $c_2 = -11/(18n)$ e $c_3 = 1/(9n)$ produzindo a estatística escore aperfeiçoada desenvolvida naquele exemplo.

Outros exemplos nos quais estatísticas escore aperfeiçoadas podem ser obtidas a partir do cálculo de momentos são encontradas ao longo dos capítulos anteriores. Este é o caso dos Exemplos 3.1, 3.2, 4.3 e 5.7 em que os três primeiros momentos das correspondentes estatísticas não modificadas foram obtidos.

Exemplo 6.2

Consideremos $p \geq 2$ populações das quais amostras aleatórias independentes $y_i = (y_{i1}, \dots, y_{in_i})^T$, $i = 1, \dots, p$, são retiradas. Suponhamos que y_{ij} , $i = 1, \dots, n_i$, $j = 1, \dots, p$, tenha densidade $\pi(y; \theta_i)$ que depende de um parâmetro desconhecido θ_i . O interesse é testar a hipótese nula $H_0 : \theta_1 = \dots = \theta_p (= \theta)$ versus a alternativa $H : \theta$ não é constante. Denotemos por $\hat{\theta}_i$ o estimador de máxima verossimilhança de θ_i sob H e $\tilde{\theta}$ o estimador de máxima verossimilhança do parâmetro comum θ sob H_0 . Hayakawa (1976) propôs, para este teste de homogeneidade de parâmetros, a estatística

$$T = \sigma^2(\tilde{\theta}) \sum_{i=1}^p n_i (\hat{\theta}_i - \tilde{\theta})^2, \quad (6.14)$$

onde

$$\sigma^2(\theta) = E \left[\left\{ \frac{d \log \pi(y; \theta)}{d\theta} \right\}^2 \right]$$

sendo a esperança calculada sob H_0 . Assintoticamente, T tem distribuição χ_{p-1}^2 sob a hipótese nula.

Supondo válidas certas condições de regularidade, Hayakawa (1976) encontrou a expansão assintótica até ordem n^{-1} da distribuição da estatística T sob H_0 . Nota-se que esta tem exatamente a forma (6.1) com $k = 3$ e com os a 's dados naquele artigo. Portanto,

de (6.5) temos que a estatística

$$T^* = T \left\{ 1 - 2 \sum_{i=1}^3 \left(\sum_{l=i}^3 a_l \right) \mu_i'^{-1} T^{i-1} \right\}$$

tem distribuição χ_r^2 até ordem n^{-1} sob H_0 .

Hayakawa propôs modificar a estatística T por um fator de correção $1-C$, onde C é de ordem n^{-1} , de tal forma que $T' = T(1-C)$ tem média que coincide, até esta ordem e sob H_0 , com a da distribuição χ^2 de referência. Assim, fica clara a importância dos resultados deste capítulo pois, enquanto T^* tem todos os momentos coincidindo, até ordem n^{-1} , com os correspondentes da distribuição χ_{p-1}^2 , a estatística T' satisfaz esta propriedade somente para o primeiro momento. Além disso, pode-se verificar, após alguma álgebra, que o fator de correção de Hayakawa é obtido do sistema (6.7) com uma única equação.

Exemplo 6.3

Este exemplo refere-se ao teste de homogeneidade de variâncias de populações normais com média zero baseado na estatística proposta por Hayakawa (1976). De (6.14) vem a estatística

$$T_\sigma = \frac{1}{2} \sum_{i=1}^p \frac{n_i (s_i - \bar{s})^2}{\bar{s}^2},$$

onde $s_i = \sum_{j=1}^{n_i} n_i^{-1} y_{ij}^2$, $\bar{s} = n^{-1} \sum_{i=1}^p n_i s_i$ sendo $n = \sum_{i=1}^p n_i$. Nota-se que T_σ é a estatística score para o teste em questão.

A expansão assintótica da distribuição de T_σ sob a hipótese nula tem a forma (6.1)

com

$$a_1 = \frac{1}{n}(-p^2 - p + 2\rho),$$

$$a_2 = \frac{1}{2n}(4p^2 + 6p - 3 - 7\rho)$$

e

$$a_3 = \frac{1}{3n}(-3p^2 - 6p + 4 + 5n\rho),$$

onde $\rho = n \sum_{i=1}^p n_i^{-1}$. Assim, de (6.5) vem

$$\begin{aligned} c_1 &= \frac{\rho - 1}{3n(p-1)} \\ c_2 &= \frac{6p^2 + 6p - 1 - 11\rho}{3n(p-1)(p+1)}, \\ c_3 &= \frac{2(-3p^2 - 6p + 4 + 5\rho)}{3n(p-1)(p+1)(p+3)}, \end{aligned} \quad (6.15)$$

que levam à estatística aperfeiçoada

$$T_\sigma^* = T_\sigma \left(1 - \sum_{i=1}^3 c_i T_\sigma^{i-1} \right)$$

tendo distribuição χ_{p-1}^2 até ordem n^{-1} sob H_0 .

Hayakawa notou que

$$T_\sigma = \frac{\eta}{\sum} \left(n \sum \frac{z_i^2}{n} - 1 \right)$$

$$T_\sigma = \frac{1}{2} \left(\sum_{i=1}^p \frac{z_i^2}{n_i} - 1 \right),$$

$$\begin{aligned} 0 &\leq z_i \leq 1 \\ \sum z_i &= 1 \end{aligned}$$

onde $\{z_1, \dots, z_p\}$ tem distribuição de Dirichlet. Baseado neste fato, encontrou alguns momentos de T_σ sob H_0 . Os três primeiros momentos ordinários de T_σ , em conexão com a expressão (6.10), fornecem uma maneira alternativa de se obter os c 's e, conseqüentemente, a estatística aperfeiçoada T_σ^* .

Exemplo 6.4

Consideremos o teste de igualdade de médias de distribuições exponenciais. A estatística (6.14) proposta por Hayakawa (1976) para este teste é

$$T_e = \sum_{i=1}^p \frac{n_i(y_i - \bar{y})^2}{\bar{y}^2}.$$

Notamos que T_e é uma estatística escore. A expansão assintótica da distribuição de T_e sob H_0 é da forma (6.1) com os a 's iguais aos do exemplo anterior divididos por dois. Portanto, a estatística

$$T_e^* = T_e \left(1 - \frac{1}{2} \sum_{i=1}^3 c_i T_e^{i-1} \right),$$

onde c_i são dados em (6.15), tem distribuição χ_{p-1}^2 até ordem n^{-1} sob H_0 .

Exemplo 6.5

Consideremos o vetor aleatório $y = (y_1, \dots, y_p)^T$ onde y_i , $i = 1, \dots, p$, é um inteiro não negativo e $\sum_{i=1}^p y_i = n$. Suponhamos que y tem distribuição multinomial com vetor de parâmetros $\pi = (\pi_1, \dots, \pi_p)$ sendo $\sum_{i=1}^p \pi_i = 1$. Na literatura estatística encontram-se inúmeros trabalhos a respeito de testes de hipóteses sobre o vetor π . Bishop, Fienberg e Holland (1975) apresentam um estudo bastante abrangente sobre estes testes. Neste exemplo estaremos enfocando, em particular, o teste da hipótese nula simples $H_0 : \pi = \pi^{(0)}$, versus a alternativa composta $H : \pi \neq \pi^{(0)}$ onde $\pi^{(0)} = (\pi_1^{(0)}, \dots, \pi_p^{(0)})$ é um vetor de valores fixados com $\sum_{i=1}^p \pi_i^{(0)} = 1$.

Cressie e Read (1984) propuseram uma família de estatísticas indexadas por um parâmetro real λ para o teste destas hipóteses. A família proposta tem a forma

$$T_\lambda = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^p y_i \left\{ \left(\frac{y_i}{n\pi_i^{(0)}} \right)^\lambda - 1 \right\}, \quad \lambda \in \mathbb{R},$$

onde, para $\lambda = 0$ e $\lambda = -1$, T_λ é definida por continuidade. Assim,

$$T_0 = \lim_{\lambda \rightarrow 0} T_\lambda = 2 \sum_{i=1}^p y_i \log \left(\frac{y_i}{n\pi_i^{(0)}} \right)$$

e

$$T_{-1} = \lim_{\lambda \rightarrow -1} T_\lambda = 2n \sum_{i=1}^p \pi_i^{(0)} \log \left(\frac{n\pi_i^{(0)}}{y_i} \right).$$

Uma característica muito interessante desta família é o fato de que ela engloba várias estatísticas conhecidas. A estatística da razão de verossimilhança, G^2 digamos, é igual a T_0 . A de Pearson, que corresponde à estatística score, é

$$X^2 = \sum_{i=1}^p \frac{(y_i - n\pi_i^{(0)})^2}{n\pi_i^{(0)}} \sim$$

e equivale a T_1 , enquanto que a estatística X^2 modificada de Neyman

$$NM^2 = 2 \sum_{i=1}^p \frac{(y_i - n\pi_i^{(0)})^2}{y_i}$$

é igual a T_{-2} . Esta última é, na verdade, a estatística de Wald para o teste em questão. Finalmente, a estatística de Freeman-Tukey

$$T^2 = 4 \sum_{i=1}^p \left\{ \sqrt{y_i} - \sqrt{n\pi_i^{(0)}} \right\}^2$$

vem de $T_{-1/2}$. Comparações entre as estatísticas X^2 , G^2 e uma versão modificada da estatística de Freeman-Tukey em pequenas amostras são encontradas em Larntz(1978; vide também Lawal, 1984). A partir de um estudo de simulação, Larntz conclui que a estatística de Pearson é a melhor no sentido de que apresenta probabilidades de erro de Tipo I mais próximas dos respectivos níveis nominais quando os testes são baseados na aproximação assintótica por χ^2 . Esta afirmação é sustentada pelos resultados de Lawal.

Cressie e Read demostram que, sob H_0 , T_λ tem distribuição assintótica χ_{p-1}^2 , qualquer que seja $\lambda \in \mathbb{R}$ desenvolvem uma expansão assintótica para sua função distribuição. Notamos que esta envolve um termo da forma de (6.1) mais uma parcela adicional que é proveniente do fato de T_λ não ter uma distribuição contínua. Além disto, encontram expansões assintóticas para os três primeiros momentos de T_λ até termos de ordem n^{-1} . Eles obtiveram, até esta ordem e sob H_0 ,

$$E(T_\lambda) = p - 1 + \frac{1}{n} \left\{ \frac{\lambda - 1}{3} (2 - 3p + s) + \frac{(\lambda - 1)(\lambda - 2)}{4} (1 - 2p + s) \right\},$$

$$E(T_\lambda^2) = (p - 1)(p + 1) + \frac{1}{n} \left[2 - 2p - p^2 + s + \frac{2(\lambda - 1)}{3} \{10 - 13p - 6p^2 + (p + 8)s\} + \frac{(\lambda - 1)^2}{3} (4 - 6p - 3p^2 + 5s) + \frac{(\lambda - 1)(\lambda - 2)}{2} \{3 - 5p - 2p^2 + (p + 3)s\} \right]$$

e

$$\begin{aligned}
E(T_\lambda^3) &= (p-1)(p+1)(p+3) \\
&+ \frac{1}{n} \left[26 - 24p - 21p^2 - 3p^3 + (19 + 3p)s \right. \\
&+ (\lambda - 1)\{70 - 81p - 64p^2 - 9p^3 + (65 + 18p + p^2)s\} \\
&+ (\lambda - 1)^2\{20 - 26p - 21p^2 - 3p^3 + (25 + 6p)s\} \\
&\left. + \frac{3(\lambda - 1)(\lambda - 2)}{4}\{15 - 22p - 15p^2 - 2p^3 + (15 + 8p + p^2)s\} \right],
\end{aligned}$$

onde $s = \sum_{i=1}^p (\pi_i^{(0)})^{-1}$.

A partir destes momentos, temos condições agora de obter uma família de estatísticas aperfeiçoadas T_λ^* , $\lambda \in \mathbb{R}$, cada qual tendo seus três primeiros momentos coincidindo com os da distribuição χ_{p-1}^2 de referência. Dos resultados da Seção 6.3, vêm as estatísticas

$$T_\lambda^* = T_\lambda \left(1 - \sum_{i=1}^3 c_i T_\lambda^{i-1}\right), \quad \lambda \in \mathbb{R},$$

onde

$$\begin{aligned}
c_1 &= -\frac{1-s}{6n(p-1)}, \\
c_2 &= \frac{\lambda^2(3p^2 - 1 - 2s) + \lambda(6p - 1 - 5s)}{12n(p-1)(p+1)}
\end{aligned}$$

e

$$c_3 = -\frac{\lambda^2(3p^2 + 6p - 4 - 5s)}{12n(p-1)(p+1)(p+3)}.$$

Substituindo λ por 0, 1, -2 e -1/2 são obtidos aperfeiçoamentos para as estatísticas usuais mencionadas anteriormente. Por exemplo, tomando $\lambda = 0$ vem a estatística da razão de verossimilhança modificada

$$G^{*2} = 2 \left\{ \sum_{i=1}^p y_i \log \left(\frac{y_i}{n\pi_i^{(0)}} \right) \right\} \left\{ 1 + \frac{1-s}{6n(p-1)} \right\},$$

onde o termo $1 + (1-s)/\{6n(p-1)\}$ é o fator de correção de Bartlett desenvolvido por Cordeiro (1982; Seção 6.6). Para o caso em que $\pi_i^{(0)} = \frac{1}{p}$, $i = 1, \dots, p$, G^{*2} equivale à estatística da razão de verossimilhança aperfeiçoada de Williams (1976; Seção 2.2).

Cressie e Read (1984) propuseram uma família de estatísticas modificadas tendo dois primeiros momentos iguais aos da distribuição χ^2_{p-1} . No entanto, a família que propusemos é “superior” a esta, no sentido de que produz um ajuste não só dos dois, mas sim dos três primeiros momentos.

CONCLUSÕES

Basicamente, os resultados desta tese são resumidos em três pontos principais: primeiro, na determinação de um fator de correção para a estatística *escore*; segundo, no desenvolvimento de fórmulas matriciais e algébricas que auxiliam no cálculo de tais fatores e, finalmente, na generalização dos fatores de Bartlett para uma ampla classe de estatísticas. Acreditamos que o primeiro ponto tenha sido buscado por vários pesquisadores. Esta crença se deve ao fato de que as pesquisas sobre fatores de Bartlett adquiriram grande impulso nos últimos 20 anos, bem como a vários comentários encontrados na literatura sobre a impossibilidade de se desenvolver um fator de correção simples para a estatística *escore*. O segundo ponto é de fundamental importância para a viabilidade de se utilizar as correções em problemas práticos. Já o terceiro mostra que os fatores de correção para as estatísticas *escore* e da razão de verossimilhança são casos especiais de um resultado mais geral e revela a possibilidade de se desenvolver correções para outras estatísticas.

Uma análise desta tese mostra inúmeras similaridades entre os fatores de correção para a estatística *escore* e os tradicionais fatores de Bartlett. Isto nos faz crer fortemente que as quatro direções principais que norteiam os trabalhos sobre fatores de Bartlett serão seguidas pelas pesquisas em correções para a estatística *escore*, de forma análoga ao que foi feito para a razão de verossimilhança. As quatro direções foram descritas no Capítulo 1.

Vários trabalhos poderão, a nosso ver, ser desenvolvidos a partir das idéias desta

tese. Entre estes, citamos:

- (i) cálculo de A_1 , A_2 e A_3 em vários modelos lineares generalizados especiais;
- (ii) extensão dos resultados da Seção 5.3 a testes sobre um ou mais componentes do vetor dos parâmetros de regressão;
- (iii) desenvolvimento de fórmulas para A_1 , A_2 e A_3 em certos modelos não-lineares da família exponencial de grande interesse prático;
- (iv) extensão dos resultados da Seção 5.5 para os modelos de dispersão;
- (v) extensão do resultado do Exemplo 4.2 a modelos de regressão utilizados em análise de sobrevivência, se possível, admitindo observações censuradas;
- (vi) aperfeiçoamento de testes escore na comparação de modelos ^{ARMA}ARIMA encaixados;
- (vii) estudos exaustivos de simulação, em particular, em modelos discretos;
- (viii) obtenção de expansão assintótica para o poder da estatística escore para os modelos de dispersão;
- (ix) estudos visando interpretar os A 's em situações menos triviais;
- (x) desenvolvimento de programas computacionais para o cálculo dos A 's em modelos lineares generalizados seguindo a Seção 3.5;
- (xi) desenvolvimento de programas para a obtenção de fórmulas algébricas para os A 's a partir das idéias da Seção 4.3;
- (xii) aperfeiçoamento de testes escore em modelos lineares generalizados com componente sistemática para o parâmetro de dispersão;
- (xiii) aperfeiçoamento de estatísticas que têm, aproximadamente, distribuição de Fisher em amostras grandes;
- (xiv) obtenção do fator de correção para a estatística de Wald a partir da determinação da expansão assintótica até ordem n^{-1} para sua distribuição.

Concluindo, esta tese é um trabalho inicial no desenvolvimento de fatores de correção para estatísticas alternativas à razão de verossimilhança e abre inúmeras linhas de pesquisa que merecem ser exploradas.

APÊNDICE 1

Cox e Reid (1987a) apresentam resultados importantes sobre relações entre expansões estocásticas de variáveis aleatórias e as correspondentes expansões assintóticas para suas funções distribuição. Estes resultados são utilizados nos Capítulos 2 e 6 para obtenção de estatísticas modificadas tendo distribuição χ^2 até ordem n^{-1} .

Enunciamos a seguir o Teorema 1 de Cox e Reid e seu corolário.

TEOREMA. Sejam X_0 e X_1 variáveis aleatórias próprias com densidades marginais contínuas tendo suporte em $(-\infty, \infty)$. Seja $Y = X_0 + \epsilon X_1$ para algum $\epsilon > 0$ fixado. Se a densidade conjunta de X_0 e X_1 satisfaz

$$(i) \quad \lim_{x_0 \rightarrow -\infty} \frac{\partial^2}{\partial x_0^2} f(x_0, x_1) = 0$$

$$(ii) \quad \sup_{x_0} \frac{\partial^2}{\partial x_0^2} |f(x_0, x_1)| < K g(x_1), \text{ onde } \int x_1^3 g(x_1) dx_1 < \infty,$$

$$(iii) \quad E(X_1^2/X_0) < \infty \text{ q.c.}$$

então

$$F_Y(y) = F_0(y) + \epsilon f_0(y)E(X_1/X_0 = y) + \frac{1}{2}\epsilon^2 \left\{ f_0(y)E(X_1^2/X_0 = y) \right\} + O(\epsilon^3) \quad (A1.1)$$

$$= F_0 \left(y - \epsilon E(X_1/X_0 = y) + \frac{1}{2}\epsilon^2 \left\{ \text{Var}(X_1/X_0 = y) \frac{f'_0(y)}{f_0(y)} + \frac{\partial}{\partial y} E(X_1^2/X_0 = y) \right\} \right) + O(\epsilon^3).^{(*)} \quad (A1.2)$$

COROLÁRIO. Se $E(X_1/X_0) = 0$ q.c., então $F_Y(y)$ difere de $F_0(y)$ por $O(\epsilon^2)$, para qualquer y fixado.

A igualdade em (A1.1) segue da expansão da função densidade de y retendo termos até ordem ϵ^2 e de sua posterior integração. A igualdade em (A1.2) é obtida igualando $F_Y(y)$ a $F_0(y + \epsilon\alpha_1(y) + \epsilon^2\alpha_2(y))$ e escolhendo $\alpha_1(y)$ e $\alpha_2(y)$ que satisfaçam a igualdade após expansão em série de Taylor. A demonstração deste teorema é encontrada no livro de Barndorff-Nielsen e Cox (1989; Seção 3.8).

Como Cox e Reid observam, as condições estabelecidas no teorema para a validade das expansões em (A1.1) e (A1.2) são mais fortes do que necessário e o teorema pode ser estendido em várias direções. Uma destas extensões é a de que, sob condições adequadas de regularidade, as expansões em (A1.1) e (A1.2) são válidas para variáveis aleatórias X_0 e X_1 que dependem de ϵ , mas que são $O_p(1)$. Discussões detalhadas sobre condições de regularidade em problemas relacionados a este são encontradas nos artigos de Bhattacharya e Ghosh (1978; 1988), Chandra e Ghosh (1979) e Chandra (1985). A extensão do teorema citado acima é utilizada por Cox e Reid (1987a, pg. 105; 1987b, pg. 15) e é usada nos Capítulos 2 e 6 para obtenção das estatísticas modificadas.

Consideremos sequências de variáveis aleatórias $\{y_n\}$ tais que

$$Y_n = X_0 + O_p(n^{-1/2}) \quad (A1.3a)$$

(*) As expressões (A1.1) e (A1.2) são chamadas nos Capítulos 2 e 6 de “expansões (1) e (2) do Teorema 1 de Cox e Reid (1987a)”.

$$Y_n = X_0 + n^{-1/2} X_1 + O_p(n^{-1}), \quad (A1.3b)$$

$$Y_n = X_0 + n^{-1/2} X_1 + n^{-1} X_2 + O_p(n^{-3/2}) \quad (A1.3c)$$

e sequências de funções distribuição $\{F_n\}$ tais que, para cada y fixado,

$$F_n(y) = F_0(y)\{1 + O(n^{-1/2})\}, \quad (A1.4a)$$

$$F_n(y) = F_0(y)\{1 + n^{-1/2} a_1(y) + O(n^{-1})\}, \quad (A1.4b)$$

$$F_n(y) = F_0(y)\{1 + n^{-1/2} a_1(y) + n^{-1} a_2(y) + O(n^{-3/2})\}. \quad (A1.4c)$$

As expressões em (A1.3a)–(A1.3c) e (A1.4a)–(A1.4c) são chamadas, respectivamente, de expansões estocásticas assintóticas para Y_n e expansões assintóticas para $F_n(y)$. Cox e Reid (1987a; pg. 107) mostram que, sob condições adequadas de regularidade, as expressões em (A1.3a)–(A1.3c) são equivalentes às respectivas expressões em (A1.4a)–(A1.4c) com $a_1(y)$ e $a_2(y)$ dados por

$$F_0(y) a_1(y) = -E(X_1/X_0 = y) f_0(y)$$

$$F_0(y) a_2(y) = -E(X_2/X_0 = y) f_0(y) + \frac{1}{2} \frac{\partial}{\partial y} \{E(X_1^2/X_0 = y) f_0(y)\}.$$

Nos Capítulos 2 e 6 utilizamos a equivalência das expansões

$$Y_n = X_0 + O_p(n^{-3/2}) \quad (A1.5)$$

e

$$F_n(y) = F_0(y)\{1 + O(n^{-3/2})\} \quad (A1.6)$$

que vem diretamente da equivalência de (A1.3c) e (A1.4c).

APÊNDICE 2

Neste apêndice apresentamos, em detalhes, a obtenção das quantidades A_1 , A_2 e A_3 referentes ao teste das hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$ em MLGs (vide Capítulo 3). Inicialmente, obtemos os cumulantes conjuntos de derivadas da log-verossimilhança envolvidos nas expressões dos A 's. Alguns cumulantes foram obtidos por Cordeiro (1983).

A2.1. OBTENÇÃO DE ALGUNS CUMULANTES CONJUNTOS DE DERIVADAS DA LOG-VEROSSIMILHANÇA EM MLGs

A log-verossimilhança total do parâmetro β do MLG descrito na Seção 3.1 tem a forma

$$L(\beta) = \sum [\phi \{ \underbrace{y_l \theta_l - b(\theta_l) + c(y_l)} \} + \overline{a(y_l, \phi)}],$$

onde $\sum$ indica o somatório sobre os dados. Por simples diferenciação chega-se a

$$\begin{aligned}\frac{\partial L(\beta)}{\partial \beta_r} &= \phi \sum \left\{ (y - \mu) w \frac{d\eta}{d\mu} \right\}_l (r)_l, \\ \frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} &= \phi \sum \left\{ (y - \mu) \frac{d^2 \theta}{d\eta^2} - w \right\}_l (rs)_l, \\ \frac{\partial^3 L(\beta)}{\partial \beta_r \partial \beta_s \partial \beta_t} &= \phi \sum \left\{ (y - \mu) \frac{d^3 \theta}{d\eta^3} - 2 \frac{d\mu}{d\eta} \frac{d^2 \theta}{d\eta^2} - \frac{d^2 \mu}{d\eta^2} \frac{d\theta}{d\eta} \right\}_l (rst)_l,\end{aligned}$$

onde $w = \left(\frac{d\mu}{d\eta} \right)^2 \frac{1}{V}$ com $V = \frac{d\mu}{d\theta}$. Tomando as esperanças das duas últimas expressões encontram-se os cumulantes

$$\kappa_{rs} = -\kappa_{r,s} = -\phi \sum w_l (rs)_l$$

e

$$\kappa_{rst} = -\phi \sum \left(2 \frac{d\mu}{d\eta} \frac{d^2 \theta}{d\eta^2} + \frac{d^2 \mu}{d\eta^2} \frac{d\theta}{d\eta} \right)_l (rst)_l.$$

Mas, facilmente, verifica-se que

$$\frac{d^2 \mu}{d\eta^2} \frac{d\theta}{d\eta} = \frac{1}{V} \frac{d^2 \mu}{d\eta^2} \frac{d\mu}{d\eta} = f$$

e

$$\frac{d\mu}{d\eta} \frac{d^2 \theta}{d\eta^2} = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2 \mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3 = g$$

e, então,

$$\kappa_{rst} = -\phi \sum (f + 2g)_l (rst)_l.$$

Agora,

$$\begin{aligned}\kappa_{r,st} &= E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial^2 L(\beta)}{\partial \beta_s \partial \beta_t} \right\} \\ &= \phi^2 E \left\{ \sum_{l,j} \left\{ (y - \mu)_l (y - \mu)_j w_l \frac{d\eta_l}{d\mu_l} \frac{d^2 \theta_j}{d\eta_j^2} - (y - \mu)_l w_l w_j \frac{d\eta_l}{d\mu_l} \right\} (r)_l (st)_j \right\}.\end{aligned}$$

Como y_l e y_j , $l \neq j$, são independentes, vem

$$\kappa_{r,st} = \phi^2 \sum \text{Var}(y_l) \left(w \frac{d\eta}{d\mu} \frac{d^2\theta}{d\eta^2} \right)_l (rst)_l.$$

Substituindo $\text{Var}(y_l)$ por V_l/ϕ chega-se a

$$\kappa_{r,st} = \phi \sum g_l(rst)_l.$$

De forma análoga e usando o fato de que o terceiro e o quarto momentos centrais de y são, respectivamente,

$$E((y - \mu)^3) = \frac{V}{\phi^2} \frac{dV}{d\mu}$$

e

$$E((y - \mu)^4) = \frac{V}{\phi^3} \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2V}{d\mu^2} \right\} + \frac{3V^2}{\phi^2},$$

vem

$$\begin{aligned} \kappa_{r,s,t} &= E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial L(\beta)}{\partial \beta_s} \frac{\partial L(\beta)}{\partial \beta_t} \right\} = \phi \sum (f - g)_l(rst)_l, \\ \kappa_{r,s,t,u} &= E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial L(\beta)}{\partial \beta_s} \frac{\partial L(\beta)}{\partial \beta_t} \frac{\partial L(\beta)}{\partial \beta_u} \right\} - \kappa_{r,s} \kappa_{t,u} - \kappa_{r,t} \kappa_{s,u} - \kappa_{r,u} \kappa_{s,t} \end{aligned}$$

$$= \phi \sum b_l(rstu)_l$$

e

$$\kappa_{r,s,tu} = E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial L(\beta)}{\partial \beta_s} \frac{\partial^2 L(\beta)}{\partial \beta_t \partial \beta_u} \right\} - \kappa_{r,s} \kappa_{tu}$$

$$= \phi \sum \left\{ \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} - \frac{1}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4 \right\}_l (rstu)_l,$$

onde

$$b = \frac{1}{V^3} \left(\frac{d\mu}{d\eta} \right)^4 \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2V}{d\mu^2} \right\}.$$

É fácil ver que

$$\kappa_{r,s,t,u} + \kappa_{r,s,tu} = \phi \sum h_l(rstu)_l$$

onde

$$h = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} + \frac{1}{V^2} \frac{d^2V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4.$$

A2.2. OBTENÇÃO DE A_1 , A_2 E A_3

As quantidades A_1 , A_2 e A_3 são agora obtidas substituindo os cumulantes encontrados neste apêndice nas expressões (2.6)-(2.8). Consideremos, inicialmente, a obtenção de A_1 . De (2.6) vem

$$\begin{aligned} A_1 = & 3\phi^2 \sum' \sum f_l(rst)_l \sum f_m(ijk)_m a_{ij} a_{st} m_{kr} \\ & + 6\phi^2 \sum' \sum f_l(ijk)_l \sum (f-g)_m(rst)_m a_{ij} a_{kr} m_{st} \\ & - 6\phi^2 \sum' \sum (2g-f)_l(ijk)_l \sum f_m(rst)_m a_{js} a_{kt} m_{ir} \\ & - 6\phi \sum' \sum h_l(ijkr)_l a_{kr} m_{ij}, \end{aligned}$$

onde, $\sum'$ indica o somatório sobre todos os β 's e a_{ij} e m_{ij} são os elementos (i, j) das matrizes A e M respectivamente, com estas matrizes dadas em (2.9).

O ponto principal na obtenção de fórmulas matriciais para os A 's é a troca dos somatórios $\sum'$ e $\sum$. Fazendo isto e rearrumando os termos, tem-se

$$\begin{aligned} A_1 = & 3\phi^2 \sum f_l f_m \left(\sum' a_{ij} x_{mi} x_{mj} \right) \left(\sum' a_{st} x_{ls} x_{lt} \right) \left(\sum' m_{kr} x_{mk} x_{lr} \right) \\ & + 6\phi^2 \sum f_l (f-g)_m \left(\sum' a_{ij} x_{li} x_{lj} \right) \left(\sum' a_{kr} x_{lk} x_{mr} \right) \left(\sum' m_{st} x_{ms} x_{mt} \right) \\ & - 6\phi^2 \sum (2g-f)_l f_m \left(\sum' a_{js} x_{lj} x_{ms} \right) \left(\sum' a_{kt} x_{lk} x_{mt} \right) \left(\sum' m_{ir} x_{li} x_{mr} \right) \\ & - 6\phi \sum h_l \left(\sum' m_{ij} x_{li} x_{lj} \right) \left(\sum' a_{kr} x_{lk} x_{lr} \right). \end{aligned}$$

Agora, das definições de Z_2 e $Z - Z_2$ em (3.8) vem

$$\begin{aligned} A_1 = & 3\phi^{-1} \sum f_l f_m z_{2mm} z_{2ll} (z_{lm} - z_{2lm}) \\ & + 6\phi^{-1} \sum f_l (f-g)_m z_{2ll} z_{2lm} (z_{mm} - z_{2mm}) \\ & - 6\phi^{-1} \sum (2g-f)_l f_m z_{2lm}^2 (z_{lm} - z_{2lm}) \\ & - 6\phi^{-1} \sum h_l (z_{ll} - z_{2ll}) z_{2ll} \end{aligned}$$

que, em notação matricial, se reduz a

$$\begin{aligned} A_1 = & \phi^{-1} [3 \mathbf{1}^T F Z_{2d} (Z - Z_2) Z_{2d} F \mathbf{1} + 6 \mathbf{1}^T F Z_{2d} Z_2 (Z - Z_2)_d (F - G) \mathbf{1} \\ & - 6 \mathbf{1}^T F \{ Z_2^{(2)} \otimes (Z - Z_2) \} (2G - F) \mathbf{1} - 6 \mathbf{1}^T H (Z - Z_2)_d Z_{2d} \mathbf{1}], \end{aligned}$$

onde a notação utilizada segue a da Seção 3.3.

A obtenção de A_2 e A_3 é conduzida de maneira análoga. De (2.5) e escrevendo os termos na ordem conveniente, vem

$$\begin{aligned} A_2 = & -3\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' a_{kr} x_{lk} x_{ms} \right) \left(\sum' m_{ij} x_{li} x_{lj} \right) \left(\sum' m_{st} x_{ms} x_{mt} \right) \\ & - 6\phi^2 \sum f_l (f-g)_m \left(\sum' a_{ij} x_{li} x_{lj} \right) \left(\sum' m_{kr} x_{lk} x_{mr} \right) \left(\sum' m_{st} x_{ms} x_{mt} \right) \\ & - 6\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' a_{kt} x_{lk} x_{mt} \right) \left(\sum' m_{ir} x_{li} x_{mr} \right) \left(\sum' m_{js} x_{lj} x_{ms} \right) \\ & + 3\phi \sum b_l \left(\sum' m_{ij} x_{li} x_{lj} \right) \left(\sum' m_{kr} x_{lk} x_{lr} \right). \end{aligned}$$

Das definições de Z_2 e $Z - Z_2$ chega-se a

$$\begin{aligned} A_2 = & -3\phi^{-1} \sum (f-g)_l (f-g)_m z_{2lm} (z_{ll} - z_{2ll}) (z_{mm} - z_{2mm}) \\ & - 6\phi^{-1} \sum f_l (f-g)_m z_{2ll} (z_{lm} - z_{2lm}) (z_{mm} - z_{2mm}) \\ & - 6\phi^{-1} \sum (f-g)_l (f-g)_m z_{2lm} (z_{lm} - z_{2lm})^2 \\ & + 3\phi^{-1} \sum b_l (z_{ll} - z_{2ll})^2. \end{aligned}$$

Agora, escrevendo A_2 em forma matricial tem-se

$$\begin{aligned} A_2 = & \phi^{-1} \left[-3 \mathbf{1}^T (F-G) (Z-Z_2)_d Z_2 (Z-Z_2)_d (F-G) \mathbf{1} - 6 \mathbf{1}^T F Z_{2d} (Z-Z_2) (Z-Z_2)_d (F-G) \mathbf{1} \right. \\ & \left. - 6 \mathbf{1}^T (F-G) \{ (Z-Z_2)^{(2)} \otimes Z_2 \} (F-G) \mathbf{1} + 3 \mathbf{1}^T B (Z-Z_2)_d^{(2)} \mathbf{1} \right]. \end{aligned}$$

Finalmente, de (2.6) vem

$$\begin{aligned} A_3 = & 3\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' m_{ij} x_{li} x_{lj} \right) \left(\sum' m_{kr} x_{lk} x_{mr} \right) \left(\sum' m_{st} x_{ms} x_{mt} \right) \\ & + 2\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' m_{ir} x_{li} x_{mr} \right) \left(\sum' m_{js} x_{lj} x_{ms} \right) \left(\sum' m_{kt} x_{lk} x_{mt} \right) \end{aligned}$$

que se reduz a

$$\begin{aligned} A_3 = & \phi^{-1} \{ 3 \mathbf{1}^T (F-G) (Z-Z_2)_d (Z-Z_2) (Z-Z_2)_d (F-G) \mathbf{1} \\ & + 2 \mathbf{1}^T (F-G) (Z-Z_2)^{(3)} (F-G) \mathbf{1} \}. \end{aligned}$$

APÊNDICE 3

Apresentamos neste apêndice a obtenção das quantidades A_1 , A_2 e A_3 que determinam os testes escore aperfeiçoados em problemas com parâmetros globalmente ortogonais (vide Seção 4.1.3).

Seja $L = L(\phi, \beta)$ a log-verossimilhança de um problema regular dependendo dos vetores paramétricos $\phi = (\phi_1, \dots, \phi_p)^T$ e $\beta = (\beta_1, \dots, \beta_q)^T$ globalmente ortogonais. Utilizamos aqui a notação introduzida na Seção 4.1.3.

O desenvolvimento de fórmulas especiais para A_1 , A_2 e A_3 para o teste de $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ é baseado nas fórmulas gerais para os A 's dadas em (2.6)–(2.8), nas relações em (4.5) para cumulantes de derivadas da log-verossimilhança e na forma das matrizes A e M em (4.6). De (2.6), (4.5) e (4.6) vem

$$A_1 = 3 \sum' (\kappa_{ijr'} - 2\kappa_{ijr'}) (\kappa_{sts'} - 2\kappa_{sts'}) a_{ij} a_{st} m_{r's'} \\ - 6 \sum' (\kappa_{ijk} - 2\kappa_{ijk} + 2\kappa_{ij}^{(k)}) (2\kappa_{rr's'} - \kappa_{r's'}^{(r)}) a_{ij} a_{kr} m_{r's'}$$

$$\begin{aligned}
 & + 6 \sum' (-\kappa_{jkr'} + \kappa_{jk}^{(r')} - 2\kappa_{jkr'} + \kappa_{jk}^{(r')}) (\kappa_{sts'} - 2\kappa_{sts'}) a_{js} a_{kt} m_{r's'} \\
 & - 6 \sum' (-2\kappa_{krr's'} + 2\kappa_{kr's'}^{(r)} + 2\kappa_{rr's'}^{(k)} + \kappa_{krr'}^{(s')} + \kappa_{kr's'}^{(r')} - \kappa_{kr}^{(r's')} \\
 & \quad + \kappa_{kr',rs'} + \kappa_{ks',rr'}) a_{kr} m_{r's'},
 \end{aligned}$$

ou seja,

$$\begin{aligned}
 A_1 &= 3 \sum' \kappa_{ijr'} \kappa_{sts'} a_{ij} a_{st} m_{r's'} \\
 & - 6 \sum' (-\kappa_{ijs} + 2\kappa_{ij}^{(s)}) (2\kappa_{tr's'} - \kappa_{r's'}^{(t)}) a_{ij} a_{st} m_{r's'} \\
 & - 6 \sum' (-3\kappa_{isr'} + 2\kappa_{is}^{(r')}) \kappa_{jts'} a_{ij} a_{st} m_{r's'} \\
 & - 6 \sum' (-2\kappa_{ijr's'} + 2\kappa_{ir's'}^{(j)} + 2\kappa_{jr's'}^{(i)} + \kappa_{ijr'}^{(s')} \\
 & \quad + \kappa_{ijs'}^{(r')} - \kappa_{ij}^{(r's')} + \kappa_{ir',js'} + \kappa_{is',jr'}) a_{ij} m_{r's'}.
 \end{aligned}$$

Mas, $a_{ij} = -\kappa^{ij}$ e $m_{r's'} = -\kappa^{r's'}$ e, evidentemente, valem as igualdades: $\sum' \kappa_{jr's'}^{(i)} \kappa^{ij} \kappa^{r's'} = \sum' \kappa_{ir's'}^{(j)} \kappa^{ij} \kappa^{r's'}$, $\sum' \kappa_{ijr'}^{(s')} \kappa^{ij} \kappa^{r's'} = \sum' \kappa_{ijs'}^{(r')} \kappa^{ij} \kappa^{r's'}$ e $\sum' \kappa_{ir',js'} \kappa^{ij} \kappa^{r's'} = \sum' \kappa_{is',jr'} \kappa^{ij} \kappa^{r's'}$.

Logo, A_1 reduz-se a

$$\begin{aligned}
 A_1 &= 6 \sum' (\kappa_{i'j'}^{(t)} - 2\kappa_{ti'j'}) (\kappa_{ijs} - 2\kappa_{ij}^{(s)}) \kappa^{ij} \kappa^{st} \kappa^{i'j'} \\
 & - 3 \sum' \kappa^{i'j'} \left(\sum_{i,j=1}^q \kappa_{iji'} \kappa^{ij} \right) \left(\sum_{s,t=1}^q \kappa_{stj'} \kappa^{st} \right) \\
 & - 6 \sum' \kappa_{jtj'} (3\kappa_{isi'} - 2\kappa_{is}^{(i')}) \kappa^{ij} \kappa^{st} \kappa^{i'j'} \\
 & + 6 \sum' (2\kappa_{iji'j'} + 4\kappa_{ii'j'}^{(j)} + 2\kappa_{ijj'}^{(j')} - \kappa_{ij}^{(i'j')} - 2\kappa_{ii',jj'}) \kappa^{ij} \kappa^{i'j'}.
 \end{aligned}$$

Agora, de (2.7) e (4.6) vem

$$\begin{aligned}
 A_2 &= -3 \sum' \kappa_{i',j',k} \kappa_{r,s',t'} a_{kr} m_{i'j'} m_{s't'} \\
 & + 6 \sum' (\kappa_{ijk'} + 2\kappa_{i,jk'}) \kappa_{r',s',t'} a_{ij} m_{k'r'} m_{s't'} \\
 & - 6 \sum' \kappa_{i',j',k} \kappa_{r',s',t} a_{kt} m_{i'r'} m_{j's'} \\
 & + 3 \sum' \kappa_{i',j',k',r'} m_{i'j'} m_{k'r'}.
 \end{aligned}$$

Lembrando que $a_{ij} = -\kappa^{ij}$ e $m_{i'j'} = -\kappa^{i'j'}$ e usando (4.5) chega-se a

$$\begin{aligned}
 A_2 &= \sum' \{ 3(2\kappa_{ii'j'} - \kappa_{i'j'}^{(i)}) (2\kappa_{jk'r'} - \kappa_{k'r'}^{(j)}) + 6\kappa_{ijr'} \kappa_{i',j',k'} \\
 & + 6(2\kappa_{ii'k'} - \kappa_{i'k'}^{(i)}) (2\kappa_{jj'r'} - \kappa_{j'r'}^{(j)}) \} \kappa^{ij} \kappa^{i'j'} \kappa^{k'r'} \\
 & + 3 \sum' \kappa_{i',j',k'r'} \kappa^{i'j'} \kappa^{k'r'}.
 \end{aligned}$$

Finalmente, de (2.8) e (4.6) tem-se

$$\begin{aligned} A_3 = & -3 \sum' \kappa_{i',j',k'} \kappa_{r',s',t'} \kappa^{i'j'} \kappa^{k'r'} \kappa^{s't'} \\ & - 2 \sum' \kappa_{i',j',k'} \kappa_{r',s',t'} \kappa^{i'r'} \kappa^{j's'} \kappa^{k't'} . \end{aligned}$$

APÊNDICE 4

Neste Apêndice apresentamos a obtenção das quantidades A_1 , A_2 e A_3 referentes aos testes das hipóteses $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$ e $H_0 : \beta = \beta^{(0)}$ versus $H : \beta \neq \beta^{(0)}$ em MLGs e modelos próprios de dispersão, com parâmetro de dispersão ϕ^{-1} desconhecido mas o mesmo para todas as observações (vide Capítulo 5).

A4.1. OBTENÇÃO DE ALGUNS CUMULANTES CONJUNTOS DE DERIVADAS DA LOG-VEROSSIMILHANÇA EM MLGs

A log-verossimilhança total de vetor de parâmetros (β^T, ϕ) do MLG descrito na Seção 5.1 tem a forma

$$L(\beta, \phi) = \sum \phi \{y_l \theta_l - b(\theta_l) + c(y_l)\} + nd_1(\phi) + \sum d_2(y_l), \quad (A4.1)$$

onde $\sum$ indica o somatório sobre os dados. Os cumulantes de derivadas da log-verossimilhança em relação a β foram obtidos no Apêndice 2. Para os demais cumulantes usa-

remos notação de tipo: $\kappa_{\phi\phi} = E\{\partial^2 L/\partial\phi^2\}$, $\kappa_{r\phi} = E\{\partial^2 L/\partial\beta_r\partial\phi\}$, $r = 1, \dots, p$, etc. Para derivadas de cumulantes utilizaremos a notação usual $\kappa_{r\phi}^{(t)} = \partial\kappa_{r\phi}/\partial\beta_t$, $\kappa_{\phi\phi}^{(\phi)} = \partial\kappa_{\phi\phi}/\partial\phi$, etc. É fácil ver que $\kappa_{r\phi} = 0$ de forma que os parâmetros β e ϕ são globalmente ortogonais. Além disto, como $\kappa_{rs} = -\phi \sum w_l(rs)_l$, com a notação para $(rs)_l$ dada no Capítulo 3, e $\kappa_{\phi\phi} = nd_1''(\phi)$ tem-se a matriz de informação total de Fisher $K = \text{diag}\{\phi X^T W X, -nd_1''(\phi)\}$.

Os cumulantes satisfazem certas relações dadas na Seção 2.3 que facilitam os cálculos. A log-verossimilhança (A4.1) e estas relações conduzem a: $\kappa_{i\phi\phi} = \kappa_{ij\phi\phi} = 0$, $\kappa_{\phi\phi\phi} = -\kappa_{\phi,\phi,\phi} = nd_1'''(\phi)$, $\kappa_{\phi ij} = -\sum_{rs} w_l(rs)_l$, $\kappa_{\phi i, \phi j} = \phi^{-1} \sum w_l(rs)_l$, e $\kappa_{\phi,\phi,\phi,\phi} = -nd_1^{iv}(\phi)$.

A4.2. OBTENÇÃO DE A_1 , A_2 E A_3 PARA O TESTE DAS HIPÓTESES

$H_0: \phi = \phi^{(0)}$ VERSUS $H: \phi \neq \phi^{(0)}$ EM MLGs

Devido à ortogonalidade global de β e ϕ , as quantidades A_1 , A_2 e A_3 referentes ao teste das hipóteses $H_0: \phi = \phi^{(0)}$ versus $H: \phi \neq \phi^{(0)}$ em MLGs podem ser obtidas dos resultados da Seção 4.1.4.2. De (4.14) vem

$$\begin{aligned} A_1 = & \frac{-3\phi^2}{nd_{(2)}} \left\{ -\sum w_l \sum' (i)_l \kappa^{ij}(j)_l \right\}^2 \\ & - \frac{6\phi}{nd_{(2)}} \sum w_l w_m \left(\sum' (i)_m \kappa^{ij}(j)_l \right) \left(\sum' (r)_m \kappa^{rs}(s)_l \right) \\ & - \frac{12\phi}{nd_{(2)}} \sum w_l \left(\sum' (i)_l \kappa^{ij}(j)_l \right), \end{aligned} \quad (A4.2)$$

onde $\sum'$ indica o somatório sobre os parâmetros e $d_{(2)}$ vem de (5.3). No Capítulo 3, definimos a matriz $Z = X(X^T W X)^{-1} X^T$ de dimensão $n \times n$. Da forma da matriz de informação K , é fácil ver que κ^{ij} é o elemento (i, j) da matriz $-\phi^{-1}(X^T W X)^{-1}$ e que, portanto, $-\sum' (i)_m \kappa^{ij}(j)_l$ corresponde ao elemento (m, l) da matriz $\phi^{-1}Z$. Logo, A_1 se

reduz a

$$A_1 = -\frac{3}{nd_{(2)}} \left\{ \sum w_l z_{ll} \right\}^2 - \frac{6}{nd_{(2)}} \sum w_l w_m z_{lm}^2 + \frac{12}{nd_{(2)}} \sum w_l z_{ll}.$$

Escrevendo em notação matricial vem

$$A_1 = -\frac{3}{nd_{(2)}} \{\text{tr}(WZ)\}^2 - \frac{6}{nd_{(2)}} \text{tr}(WZWZ) + \frac{12}{nd_{(2)}} \text{tr}(WZ),$$

Mas $\text{tr}(WZ) = \text{tr}(WX(X^T W X)^{-1} X^T) = \text{tr}(X^T W X (X^T W X)^{-1}) = \text{tr}(I_p) = p$, sendo p o posto da matriz X . Aqui, I_p denota uma matriz identidade de ordem p . Da mesma forma, mostra-se também que $\text{tr}(WZWZ) = p$. Assim, temos

$$A_1 = -\frac{3p(p-2)}{nd_{(2)}}. \quad (A4.3)$$

Agora, de (4.15) vem

$$A_2 = \frac{1}{n^2 d_{(2)}^2} \left\{ -6n\phi d_{(3)} \sum w_m \left(\sum' (i)_l \kappa^{ij}(j)_l \right) - 3nd_{(4)} \right\}$$

com os d 's definidos em (5.3), que se reduz a

$$A_2 = -\frac{3}{nd_{(2)}^2} \{2pd_{(3)} + d_{(4)}\}. \quad (A4.4)$$

Finalmente, de (4.16) temos

$$A_3 = -5 \frac{d_{(3)}^2}{d_{(2)}^3}. \quad (A4.5)$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

$$\text{Tr} = ae+bg+cf+dh$$

$$(a+d)(e+h) = ae+ah+de+dh$$

A4.3. OBTENÇÃO DE A_1 , A_2 E A_3 PARA O TESTE DAS HIPÓTESES

$H_0: \beta = \beta^{(0)}$ VERSUS $H: \beta \neq \beta^{(0)}$ EM MLGs

Para o teste das hipóteses $H_0: \beta = \beta^{(0)}$ versus $H: \beta \neq \beta^{(0)}$ com ϕ representando o parâmetro de perturbação, a obtenção de A_1 , A_2 e A_3 pode ser feita através dos resultados da Seção 4.1.4.3. De (4.17) vem

$$A_1 = -\frac{6d_{(3)}}{nd_{(2)}^2} \sum w_l \left(\sum' (i)_l \kappa^{ij}(j)_l \right) - \frac{12\phi}{nd_{(2)}} \sum w_l \left(\sum' (i)_l \kappa^{ij}(j)_l \right)$$

que, escrita em notação matricial, se reduz a

$$A_1 = \frac{6d_{(3)}}{nd_{(2)}^2} \text{tr}(WZ) + \frac{12}{nd_{(2)}} \text{tr}(WZ).$$

Como $\text{tr}(WZ) = p$ segue que

$$A_1 = \frac{6p}{nd_{(2)}^2} (d_{(3)} + 2d_{(2)}). \quad (A4.6)$$

Agora de (4.18) vem

$$A_2 = \frac{3\phi^2}{nd_{(2)}} \sum w_l w_m \sum' (i)_l \kappa^{ij}(j)_l \sum' (k)_m \kappa^{kr}(r)_m + \frac{6\phi^2}{nd_{(2)}} \sum w_l w_m \sum' (i)_l \kappa^{ij}(j)_m \sum' (k)_l \kappa^{kr}(r)_m + 3\phi \sum b_l \sum (i)_l \kappa^{ij}(j)_l \sum (k)_l \kappa^{kr}(r)_l.$$

Handwritten notes: $\text{Tr}(WZ) = \sum w_l z_{ll}$, $\text{Tr}(WZ)^2 = \sum w_l z_{ll} \sum w_m z_{mm}$, $\text{Tr}(WZ \otimes WZ) = \text{Tr}(WZ) \text{Tr}(WZ)$.

Após alguma álgebra chega-se a

$$A_2 = \frac{3p(p+2)}{nd_{(2)}} + \frac{3}{\phi} 1^T B Z_d^{(2)} 1. \quad (A4.7)$$

Finalmente, de (4.9) segue

$$A_3 = -3\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' (i)_l \kappa^{ij}(j)_l \right) \left(\sum' (k)_l \kappa^{kr}(r)_m \right) \left(\sum' (s)_m \kappa^{st}(t)_m \right) - 2\phi^2 \sum (f-g)_l (f-g)_m \left(\sum' (i)_l \kappa^{ir}(r)_m \right) \left(\sum' (j)_l \kappa^{js}(s)_m \right) \left(\sum' (k)_l \kappa^{kt}(t)_m \right)$$

que, escrita em notação matricial, tem a forma

$$A_3 = \phi^{-1} 1^T (F - G) (3Z_d Z Z_d + 2Z^{(3)}) (F - G) 1. \quad (A4.8)$$

A4.4. OBTENÇÃO DE A_1 , A_2 E A_3 PARA OS TESTES DE $H_0: \phi = \phi^{(0)}$ VERSUS $H: \phi \neq \phi^{(0)}$ E $H_0: \beta = \beta^{(0)}$ VERSUS $H: \beta \neq \beta^{(0)}$ EM MODELOS PRÓPRIOS DE DISPERSÃO

A log-verossimilhança total do vetor de parâmetros (β^T, ϕ) dos modelos próprios de dispersão descritos na Seção 5.4.1 é dada por

$$L(\beta, \phi) = \sum \phi t(y_l, \theta_l) + n d_1(\phi) + \sum d_2(y). \quad (A4.9)$$

Da forma de (A4.1) e (A4.9) é fácil ver que os cumulantes de derivadas da log-verossimilhança em relação ao parâmetro ϕ somente, isto é, $\kappa_{\phi\phi}$, $\kappa_{\phi\phi\phi}$, etc., são iguais aos cumulantes correspondentes nos MLGs. Por outro lado, o mesmo não ocorre com aqueles que envolvem derivadas em relação a β . Facilmente se verifica que diversos cumulantes são nulos. Entre estes estão: $\kappa_{\phi r}$, $\kappa_{r\phi\phi}$, $\kappa_{\phi\phi rs}$, $\kappa_{\phi\phi\phi r}$, $\kappa_{\phi\phi,\phi}$ e $\kappa_{\phi\phi,\phi\phi}$. Para o cálculo dos A 's são necessários ainda os cumulantes $\kappa_{r,s,t}$, $\kappa_{rs\phi}$, $\kappa_{r\phi,s\phi}$ e $\kappa_{r,s,t,u}$. No desenvolvimento que se segue usamos a notação: $(r)_l = \partial \eta_l / \partial \beta_r$, $(r, s)_l = \partial^2 \eta_l / \partial \beta_r \partial \beta_s$ e $(r, s, t)_l = \partial^3 \eta_l / \partial \beta_r \partial \beta_s \partial \beta_t$.

Derivando $L(\beta, \phi)$ em relação a β_r tem-se

$$\frac{\partial L}{\partial \beta_r} = \phi \left\{ \sum t'(y_l, \theta_l) \left(\frac{d\theta}{d\eta} \right)_l (r)_l \right\},$$

onde $t'(y, \theta) = \partial t(y, \theta) / \partial \theta$. Então,

$$\kappa_{r,s,t} = E \left\{ \frac{\partial L}{\partial \beta_r} \frac{\partial L}{\partial \beta_s} \frac{\partial L}{\partial \beta_t} \right\} = \phi^3 \sum E \left\{ (t'(y_l, \theta_l))^3 \right\} \left(\frac{d\theta}{d\eta} \right)_l^3 (r, s, t)_l$$

uma vez que $E(t'(y, \theta)) = 0$ e que $y_1, \dots, y_n$ são independentes. A notação para $(r, s, t)_l$ segue a do Capítulo 3. Mas, por condições de regularidade,

$$\phi^{-2} E \{ (t'(y, \theta))^3 \} = 2E \{ t'''(y, \theta) \} - 3 \partial E \{ t''(y, \theta) \} / \partial \theta,$$

onde $t''(\cdot, \cdot)$ e $t'''(\cdot, \cdot)$ indicam a segunda e a terceira derivadas de $t(y, \theta)$ com respeito a θ . Logo, $\kappa_{r,s,t} = \phi \sum c_l^*(r, s, t)_l$, onde $c^* = (d\theta/d\eta)^3(2D_3 - 3D_2')$ e $D_2 = D_2(\theta, \phi) = E\{\partial^2 t(y, \theta)/\partial\theta^2\}$, $D_3 = D_3(\theta, \phi) = E\{\partial^3 t(y, \theta)/\partial\theta^3\}$ e $D_2' = \partial D_2(\theta, \phi)/\partial\theta$.

Agora

$$\kappa_{r\phi,s\phi} = E\left\{\frac{\partial^2 L}{\partial\beta_r\partial\phi}\frac{\partial^2 L}{\partial\beta_s\partial\phi}\right\} - \kappa_{r\phi}\kappa_{s\phi},$$

mas como $\kappa_{r\phi} = \kappa_{s\phi} = 0$ tem-se

$$\begin{aligned}\kappa_{r\phi,s\phi} &= E\left\{\sum\left(\frac{\partial t(y, \theta)}{\partial\beta_r}\right)_l \sum\left(\frac{\partial t(y, \theta)}{\partial\beta_s}\right)_m\right\} \\ &= E\left\{\sum\left(t'(y, \theta)\frac{\partial\theta}{\partial\eta}(r)\right)_l \sum\left(t'(y, \theta)\frac{\partial\theta}{\partial\eta}(s)\right)_m\right\}.\end{aligned}$$

Da independência de $y_1, \dots, y_n$ vem

$$\kappa_{r\phi,s\phi} = \sum E\{(t'(y, \theta))_l^2\} \left\{\left(\frac{\partial\theta}{\partial\eta}\right)^2(r, s)\right\}_l$$

e, portanto,

$$\kappa_{r\phi,s\phi} = \phi^{-1} \sum w_l^*(r, s)_l.$$

De forma análoga tem-se

$$\kappa_{rs\phi} = - \sum w_l^*(r, s)_l.$$

Finalmente,

$$\kappa_{r,s,t,u} = E\left\{\frac{\partial L}{\partial\beta_r}\frac{\partial L}{\partial\beta_s}\frac{\partial L}{\partial\beta_t}\frac{\partial L}{\partial\beta_u}\right\} - \kappa_{r,s}\kappa_{t,u} - \kappa_{r,t}\kappa_{s,u} - \kappa_{r,u}\kappa_{s,t}$$

que, após alguma álgebra, se reduz a

$$\kappa_{r,s,t,u} = \phi^4 \sum \left[E\{(t'(y, \theta))^4\} - 3[E\{(t'(y, \theta))^2\}]^2 \right] \left(\frac{d\theta}{d\eta}\right)_l^4(r, s, t, u)_l.$$

Mas, de condições de regularidade, vem

$$\begin{aligned}\phi^4 E\{(t'(y, \theta))^4\} &= -3\phi E\{t^{iv}(y, \theta)\} + 8\phi \frac{\partial E\{t'''(y, \theta)\}}{\partial\theta} \\ &\quad - 6\phi \frac{\partial^2 E\{t''(y, \theta)\}}{\partial\theta^2} + 3\phi^2 E\{(t''(y, \theta))^2\}\end{aligned}$$

e

$$\phi E\{(t'(y, \theta))^2\} = -E\{t''(y, \theta)\}.$$

Portanto,

$$\kappa_{r,s,t,u} = \phi \sum b_l^*(r, s, t, u)_l,$$

onde

$$b_l^* = \left(\frac{d\theta}{d\eta}\right)^4 (-3D_4 + 8D_3' - 6D_2'' + 3\phi D_2^{(2)} - 3\phi D_2^2)$$

com $D_4 = D_4(\theta, \phi) = E\{\partial^4 t(y, \theta)/\partial \theta^4\}$, $D_2^{(2)} = D_2^{(2)}(\theta, \phi) = E\{(\partial^2 t(y, \theta)/\partial \theta^2)^2\}$, $D_2'' = D_2''(\theta, \phi) = \partial^2 D_2(\theta, \phi)/\partial \theta^2$, $D_3' = D_3'(\theta, \phi) = \partial D_3(\theta, \phi)/\partial \theta$ e com D_2 e D_3 definidos acima.

Para o teste das hipóteses $H_0 : \phi = \phi^{(0)}$ versus $H : \phi \neq \phi^{(0)}$, vem de (4.14)

$$\begin{aligned} A_1 = & -\frac{3\phi^2}{nd_{(2)}} \left\{ -\sum w_l^* \left(\sum' (i)_l \kappa^{ij}(j)_l \right) \right\}^2 \\ & - \frac{6\phi^2}{nd_{(2)}} \sum w_l^* w_m^* \left(\sum' (i)_m \kappa^{ij}(j)_l \right) \left(\sum' (r)_m \kappa^{rs}(s)_l \right) \\ & - \frac{12\phi}{nd_{(2)}} \sum w_l^* \left(\sum' (i)_l \kappa^{ij}(j)_l \right), \end{aligned}$$

onde w_l^* é o l -ésimo elemento da diagonal de W^* . Aqui, κ^{ij} é o elemento (i, j) da matriz $-\phi^{-1}(X^{*T}W^*X^*)$ e, portanto, $-\sum (i)_m \kappa^{ij}(j)_l$ corresponde ao elemento (m, l) da matriz $\phi^{-1}Z^*$. As matrizes W^* , X^* e Z^* estão definidas na Seção 5.4.1. Nota-se que a forma de A_1 coincide com a de (A4.2) e assim A_1 se reduz a (A4.3). Analogamente, mostra-se que A_2 e A_3 reduzem-se a (A4.4) e (A4.5) respectivamente.

Para o teste das hipóteses $H_0 : \beta = \beta^{(0)}$ versus $H : \beta \neq \beta^{(0)}$ mostra-se também, sem dificuldade, que A_1 é dado por (A4.6) enquanto que A_2 e A_3 vêm, respectivamente, de (A4.7) e (A4.8) com B , Z , Z_d e $F - G$ substituídos, respectivamente, por B^* , Z^* , Z_d^* e C^* . As matrizes B^* e C^* estão definidas na Seção 5.4.3.

APÊNDICE 5

Neste apêndice apresentamos a obtenção das quantidades A_1 , A_2 e A_3 que determinam os testes escore aperfeiçoados das hipóteses $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H : \beta_1 \neq \beta_1^{(0)}$ em modelos não-lineares da família exponencial (vide Seção 5.5).

A2.1. OBTENÇÃO DE ALGUNS CUMULANTES DE DERIVADAS DA LOG- VEROSSIMILHANÇA EM MODELOS NÃO-LINEARES DA FAMÍLIA EXPONENCIAL

A log-verossimilhança total do parâmetro β dos modelos não-lineares da família exponencial descritos na Seção 5.5.1 tem a forma

$$L(\beta) = \sum [\phi\{y_l\theta_l - b(\theta_l) + c(y_l)\} + a(y_l, \phi)],$$

onde $\sum$ indica o somatório sobre todas as observações. Apresentamos, a seguir, a obtenção dos cumulantes de derivadas de $L(\beta)$ envolvidos nas expressões dos A 's. Alguns cumulantes

foram obtidos na tese de doutoramento de G.A.Paula (1988). A notação utilizada aqui foi introduzida na Seção 5.5.2.

Derivando $L(\beta)$ com respeito a β_r vem

$$\frac{\partial L(\beta)}{\partial \beta_r} = \phi \sum \left\{ (y - \mu) w \frac{d\eta}{d\mu} \right\}_l (r)_l.$$

Note-se que aqui $(r) = \partial\eta/\partial\beta_r$ depende de parâmetros, ao contrário do que ocorre para os MLGs. Esta é a razão pela qual torna-se bastante trabalhosa a obtenção das outras derivadas de $L(\beta)$ necessárias ao cálculo dos cumulantes.

Ao derivar a expressão acima com relação a β_s tem-se

$$\frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} = \phi \sum \left\{ (y - \mu) \frac{\partial}{\partial \beta_s} \left(w \frac{d\eta}{d\mu} (r) \right) + w \frac{d\eta}{d\mu} (r) \frac{\partial}{\partial \beta_s} (y - \mu) \right\}_l.$$

Mas

$$\frac{\partial}{\partial \beta_s} \left(w \frac{d\eta}{d\mu} (r) \right) = \frac{\partial}{\partial \beta_s} \left(\frac{d\theta}{d\eta} \frac{d\eta}{d\beta_r} \right) = \frac{d^2\theta}{d\eta^2} (r, s) + w \frac{d\eta}{d\mu} (rs).$$

Logo,

$$\frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} = \phi \sum \left[\left\{ (y - \mu) \frac{d^2\theta}{d\eta^2} - w \right\} (r, s) + (y - \mu) w \frac{d\eta}{d\mu} (rs) \right]_l.$$

Tomando a esperança da última expressão vem

$$\kappa_{rs} = -\phi \sum w_l (r, s)_l.$$

A obtenção da terceira derivada de $L(\beta)$ em relação aos componentes de β envolve um enorme trabalho algébrico que omitimos aqui. Obtemos

$$\begin{aligned} \frac{\partial^3 L(\beta)}{\partial \beta_r \partial \beta_s \partial \beta_t} = & \phi \sum \left[\left\{ (y - \mu) \frac{d^3\theta}{d\eta^3} - 2 \frac{d^2\theta}{d\eta^2} \frac{d\mu}{d\eta} - \frac{d^2\mu}{d\eta^2} - \frac{d^2\mu}{d\eta^2} \frac{d\theta}{d\eta} \right\} (r, s, t) \right. \\ & + \left\{ (y - \mu) \frac{d^2\theta}{d\eta^2} - w \right\} \{ (r, st) + (s, rt) + (t, rs) \} \\ & \left. + (y - \mu) \frac{d\theta}{d\eta} (rst) \right]_l. \end{aligned}$$

Sejam

$$f = \frac{d^2\mu}{d\eta^2} \frac{d\theta}{d\eta} \quad \text{e} \quad g = \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2}.$$

Vimos no Apêndice 2 que f e g podem ser escritos como

$$f = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} \quad \text{e} \quad g = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3.$$

Assim, é fácil ver que

$$\kappa_{rst} = -\phi \sum [(f + 2g)(r, s, t) + w\{(r, st) + (s, rt) + (t, rs)\}]_l.$$

Agora,

$$\begin{aligned} \kappa_{r,st} &= E \left(\frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial^2 L(\beta)}{\partial \beta_s \partial \beta_t} \right) \\ &= \phi^2 E \left\{ \sum_{l,j} \left[(y - \mu)_l (y - \mu)_j w_l \frac{d\eta_l}{d\mu_l} \frac{d^2\theta_j}{d\eta_j^2} - (y - \mu)_l w_l w_j \frac{d\eta_l}{d\mu_l} \right. \right. \\ &\quad \left. \left. + (y - \mu)_l (y - \mu)_j w_l w_j \frac{d\eta_l}{d\mu_l} \frac{d\eta_j}{d\mu_j} \right] (r)_l (st)_j \right\}. \end{aligned}$$

Como y_l e y_j , $l \neq j$, são independentes, vem

$$\kappa_{r,st} = \phi^2 \sum \left\{ \text{Var}(y_l) \left(w \frac{d\eta}{d\mu} \frac{d^2\theta}{d\eta^2} \right)_l (r, s, t)_l + \text{Var}(y_l) w_l^2 \left(\frac{d\eta}{d\mu} \right)_l^2 (r, st)_l \right\}.$$

Como $\text{Var}(y) = V/\phi$ chega-se a

$$\kappa_{r,st} = \phi \sum \{g(r, s, t) + w(r, st)\}_l.$$

Os demais cumulantes são obtidos de forma análoga. Tem-se

$$\begin{aligned} \kappa_{r,s,t} &= E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial L(\beta)}{\partial \beta_s} \frac{\partial L(\beta)}{\partial \beta_t} \right\} = \phi \sum (f - g)_l (r, s, t)_l, \\ \kappa_{r,s,t,u} &= E \left\{ \frac{\partial L(\beta)}{\partial \beta_r} \frac{\partial L(\beta)}{\partial \beta_s} \frac{\partial L(\beta)}{\partial \beta_t} \frac{\partial L(\beta)}{\partial \beta_u} \right\} - \kappa_{r,s} \kappa_{t,u} - \kappa_{r,t} \kappa_{s,u} - \kappa_{r,u} \kappa_{s,t} \\ &= \phi \sum b_l (r, s, t, u)_l, \\ \kappa_{r,s,tu} &= \phi \sum \left[\left\{ \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2} - \frac{1}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4 \right\}_l (r, s, t, u)_l \right. \\ &\quad \left. + (f - g)_l (r, s, tu)_l \right] \end{aligned}$$

e

$$\kappa_{r,s,t,u} + \kappa_{r,s,tu} = \phi \sum \{h_l(r, s, t, u) + (f - g)_l(r, s, tu)_l\},$$

onde

$$b = \frac{1}{V^3} \left(\frac{d\mu}{d\eta} \right)^4 \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2 V}{d\mu^2} \right\}$$

e

$$h = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2 \mu}{d\eta^2} + \frac{1}{V^2} \frac{d^2 V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4.$$

A5.2. OBTENÇÃO DE A_1 , A_2 E A_3

Objetivamos aqui encontrar fórmulas matriciais para A_1 , A_2 e A_3 substituindo os cumulantes obtidos neste apêndice nas expressões (2.6)–(2.8). Denotemos por A_{lm} a m -ésima parcela de A_l . De (2.6) vem

$$\begin{aligned} A_{11} &= 3 \sum' (\kappa_{ijk} + 2\kappa_{i,jk})(\kappa_{rst} + 2\kappa_{rs,t}) a_{ij} a_{st} m_{kr} \\ &= 3\phi^2 \left[\sum f_l f_m \left(\sum' a_{ij}(i)_l(j)_l \right) \left(\sum' a_{st}(s)_m(t)_m \right) \left(\sum m_{kr}(k)_l(r)_m \right) \right. \\ &\quad + w_l f_m \sum' (i, j, k)_l \{ (r, st) + (s, rt) - (t, rs) \}_m a_{ij} a_{st} m_{kr} \\ &\quad + f_l w_m \sum' \{ -(i, jk) + (j, ik) + (k, ij) \}_l (r, s, t)_m a_{ij} a_{st} m_{kr} \\ &\quad \left. + w_l w_m \sum' \{ -(i, jk) + (j, ik) + (k, ij) \}_l \{ (r, st) + (s, rt) - (t, rs) \}_m a_{ij} a_{st} m_{kr} \right]. \end{aligned}$$

Mas,

$$\begin{aligned} &\sum' (i, j, k)_l \{ (r, st) + (s, rt) - (t, rs) \}_m a_{ij} a_{st} m_{kr} \\ &= \sum' (i)_l a_{ij}(j)_l \left\{ \left(\sum' (k)_l m_{kr}(r)_m \right) \left(\sum' a_{st}(ts)_m \right) \right. \\ &\quad \left. + \sum' (k)_l m_{kr}(rt)_m a_{ts}(s)_m - \sum' (k)_l m_{kr}(rs)_m a_{st}(t)_m \right\} \\ &= \sum' (i)_l a_{ij}(j)_l \left(\sum' (k)_l m_{kr}(r)_m \right) \left(\sum' a_{st}(ts)_m \right) \\ &= \phi^{-2} z_{2l}^* (z_{lm}^* - z_{2lm}^*) \text{tr}(AD^{(m)}), \end{aligned} \tag{A5.1}$$

onde z_{2lm}^* e z_{lm}^* são os elementos (l, m) das matrizes Z_2^* e Z^* respectivamente e $D^{(l)} = \{\partial^2 \eta_l / \partial \beta_i \partial \beta_j\}$ e $A = \{a_{ij}\}$ são matrizes de dimensão $p \times p$. Particionamos a

matriz $D^{(l)}$ em

$$D^{(l)} = \begin{pmatrix} D_{11}^{(l)} & D_{12}^{(l)} \\ D_{21}^{(l)} & D_{22}^{(l)} \end{pmatrix}$$

segundo a partição de $\beta = (\beta_1^T, \beta_2^T)^T$. Agora, definimos $Q_2 = \text{diag}\{q_1, \dots, q_n\}$ com $q_l = \phi \text{tr}(K_{22}^{-1} D_{22}^{(l)})$, ou seja, com $q_l = \text{tr}((X_2^{*T} W X_2^*)^{-1} D_{22}^{(l)})$. Então, a expressão em (A5.1) se reduz a

$$\phi^{-3} z_{2l}^* (z_{lm}^* - z_{2lm}^*) q_m.$$

Analogamente, chega-se a

$$\begin{aligned} & \sum' \{-(i, jk) + (j, ik) + (k, ij)\}_l (r, s, t)_m a_{ij} a_{st} m_{kr} \\ &= \phi^{-3} z_{2mm}^* q_l (z_{lm}^* - z_{2lm}^*) \end{aligned}$$

e

$$\begin{aligned} & \sum' \{-(i, jk) + (j, ik) + (k, ij)\}_l \{(r, st) + (s, rt) - (t, rs)\}_m a_{ij} a_{st} m_{kr} \\ &= \phi^{-3} q_l (z_{lm}^* - z_{2lm}^*) q_m. \end{aligned}$$

Assim, A_{11} se reduz a

$$\begin{aligned} A_{11} &= 3\phi^{-1} \sum \{f_l z_{2l}^* (z_{lm}^* - z_{2lm}^*) z_{2mm}^* f_m \\ &\quad + w_l z_{2l}^* (z_{lm}^* - z_{2lm}^*) q_m f_m \\ &\quad + f_l q_l (z_{lm}^* - z_{2lm}^*) z_{2mm}^* w_m \\ &\quad + w_l q_l (z_{lm}^* - z_{2lm}^*) q_m w_m \} \end{aligned}$$

e pode ser escrita em notação matricial como

$$\begin{aligned} A_{11} &= 3\phi^{-1} \{1^T F Z_{2d}^* (Z^* - Z_2^*) Z_{2d}^* F 1 + 2 \, 1^T W Q_2 (Z^* - Z_2^*) Z_{2d}^* F 1 \\ &\quad + 1^T W Q_2 (Z^* - Z_2^*) Q_2 W 1\}, \end{aligned}$$

onde $F = \text{diag}\{f_1, \dots, f_n\}$ e $W = \text{diag}\{w_1, \dots, w_n\}$.

A obtenção de A_{12} a partir de (2.6) é absolutamente análoga à de A_{11} . Tem-se

$$A_{12} = 6\phi^{-1} \{1^T F Z_{2d}^* Z_2^* (Z^* - Z_2^*)_d (F - G) 1 + 1^T W Q_2 Z_2^* (Z^* - Z_2^*)_d (F - G) 1\}$$

com $G = \text{diag}\{g_1, \dots, g_n\}$.

Agora, de (2.6) vem

$$\begin{aligned}
 A_{13} &= 6 \sum' (\kappa_{i,jk} - \kappa_{i,j,k})(\kappa_{rst} + 2\kappa_{rs,t}) a_{js} a_{kt} m_{ir} \\
 &= -6\phi^2 \sum (2g - f)_l f_m \left(\sum' a_{js}(j)_l (s)_m \right) \left(\sum' a_{kt}(k)_l (t)_m \right) \left(\sum' m_{ir}(i)_l (r)_m \right) \\
 &\quad + \sum w_l f_m \sum' (i,jk)_l (r,s,t)_m a_{js} a_{kt} m_{ir} \\
 &\quad + (2g - f)_l w_m \sum' (i,j,k) \{ (r,st) + (s,rt) - (t,rs) \}_m a_{js} a_{kt} m_{ir} \\
 &\quad + w_l w_m \sum' (i,jk)_l \{ (r,st) + (s,rt) - (t,rs) \}_m a_{js} a_{kt} m_{ir}.
 \end{aligned}$$

Mas,

$$\begin{aligned}
 &\sum' (i,jk)_l (r,s,t)_m a_{js} a_{kt} m_{ir} \\
 &= \sum' (i)_l m_{ir}(r)_m \sum' (s)_m a_{sj}(jk)_l a_{kt}(t)_m \\
 &= \phi^{-1} (z_{lm}^* - z_{2lm}^*)_{lm} x_m^* A D^{(l)} A x_m^{*T},
 \end{aligned} \tag{A5.2}$$

onde $x_m^* = (x_{1m}^{*T}, x_{2m}^{*T})$ é a $\mathcal{A}$ -ésima linha da matriz modelo local X^* particionada segundo a partição de β . Definimos $C_2 = \{c_{lm}\}$, matriz de dimensão $n \times n$, com $c_{lm} = \phi^2 x_m^* A D^{(l)} A x_m^{*T}$, ou seja, com $c_{lm} = \phi^2 x_{2m}^* (X_2^{*T} W X_2^*)^{-1} D_{22}^{(l)} (X_2^{*T} W X_2^*)^{-1} x_{2m}^{*T}$. Note-se que a matriz C_2 não é simétrica. Agora, a expressão em (A5.2) se reduz a

$$\phi^{-3} (z_{lm}^* - z_{2lm}^*) c_{lm}.$$

Analogamente, obtemos

$$\begin{aligned}
 &\sum' (i,j,k)_l \{ (r,st) + (s,rt) - (t,rs) \}_m a_{js} a_{kt} m_{ir} \\
 &= \phi^{-3} (z_{lm}^* - z_{2lm}^*) c_{ml}.
 \end{aligned}$$

Da mesma forma temos que

$$\begin{aligned}
 &\sum' (i,jk)_l \{ (r,st) + (s,rt) - (t,rs) \}_m a_{js} a_{kt} m_{ir} \\
 &= \sum' (i)_l m_{ir}(r)_m \sum' (jk)_l a_{kt}(ts)_m a_{sj} \\
 &\quad - \sum' (i)_l m_{ir}(rt)_m a_{tk}(kj)_l a_{js}(s)_m \\
 &\quad - \sum' (i)_l m_{ir}(rs)_m a_{sj}(jk)_l a_{kt}(t)_m \\
 &= \phi^{-1} (z_{lm}^* - z_{2lm}^*) \text{tr}(D_{22}^{(l)} A D_{22}^{(m)} A).
 \end{aligned} \tag{A5.3}$$

Definimos $J_2 = \{j_{lm}\}$ uma matriz $n \times n$ com $j_{lm} = \text{tr}(D_{22}^{(l)} A D_{22}^{(m)} A)$, ou seja, com $j_{lm} = \text{tr}(D_{22}^{(l)} (X_2^{*T} W X_2^*)^{-1} D_{22}^{(m)} (X_2^{*T} W X_2^*)^{-1})$. Assim, escreve-se a expressão em (A5.3) como

$$\phi^{-3}(z_{lm}^* - z_{2lm}^*)j_{lm}.$$

Podemos então reduzir a parcela A_{13} a

$$\begin{aligned} A_{13} = & -6\phi^{-1} \left\{ \sum (2g - f) z_{2lm}^* z_{2lm}^* (z_{lm}^* - z_{2lm}^*) f_m \right. \\ & + w_l (z_{lm}^* - z_{2lm}^*) c_{lm} f_m \\ & + (2g - f)_l (z_{lm}^* - z_{2lm}^*) c_{ml} w_m \\ & \left. + w_l (z_{lm}^* - z_{2lm}^*) j_{lm} w_m \right\} \end{aligned}$$

que, escrita em notação matricial, tem a forma

$$\begin{aligned} A_{13} = & -6\phi^{-1} [1^T (2G - F) \{Z_2^{*(2)} \otimes (Z^* - Z_2^*)\} F 1 \\ & + 1^T W \{(Z^* - Z_2^*) \otimes C_2\} F 1 \\ & + 1^T (2G - F) \{(Z^* - Z_2^*) \otimes C_2^T\} W 1 \\ & + 1^T W \{(Z^* - Z_2^*) \otimes J_2\} W 1]. \end{aligned}$$

A última parcela de A_1 é encontrada sem dificuldade e se reduz a

$$A_{14} = -6\phi^{-1} 1^T \{H(Z^* - Z_2^*)_d Z_{2d}^* + (F - G) Q_2 (Z^* - Z_2^*)_a\} 1$$

com $H = \text{diag}\{h_1, \dots, h_n\}$.

Temos, portanto,

$$\begin{aligned} A_1 = & \phi^{-1} [3 1^T F Z_{2d}^* (Z^* - Z_2^*) Z_{2d}^* F 1 + 6 1^T F Z_{2d}^* Z_2^* (Z^* - Z_2^*)_d (F - G) 1 \\ & - 6 1^T (2G - F) \{Z_2^{*(2)} \otimes (Z^* - Z_2^*)\} F 1 - 6 1^T H (Z^* - Z_2^*)_d Z_{2d}^* 1 \\ & + 6 1^T W \{Q_2 (Z^* - Z_2^*) Z_{2d}^* - [(Z^* - Z_2^*) \otimes C_2]^T\} F 1 \\ & + 3 1^T W \{Q_2 (Z^* - Z_2^*) Q_2 + 2[(Z^* - Z_2^*) \otimes J_2]\} W 1 \\ & + 6 W Q_2 Z_2^* (Z^* - Z_2^*)_d (F - G) 1 + 6 1^T (2G - F) \{(Z^* - Z_2^*) \otimes C_2^T\} W 1 \\ & - 6 1^T (F - G) Q_2 (Z^* - Z_2^*)_d 1], \end{aligned}$$

Os passos para a determinação de fórmulas matriciais para A_2 e A_3 são análogos aos da obtenção de A_1 e, portanto, serão omitidos. Obtemos

$$\begin{aligned}
 A_2 = \phi^{-1} [& -3 \, 1^T (F - G) (Z^* - Z_2^*)_d Z_2^* (Z^* - Z_2^*)_d (F - G) 1 \\
 & - 6 \, 1^T F Z_{2d}^* (Z^* - Z_2^*) (Z^* - Z_2^*)_d (F - G) 1 \\
 & - 6 \, 1^T (F - G) \{ (Z^* - Z_2^*)^{(2)} \otimes Z_2^* \} (F - G) 1 + 3 \, 1^T B (Z - Z_2^*)_d^{(2)} 1 \\
 & - 6 \, 1^T W Q_2 (Z^* - Z_2^*) (Z^* - Z_2^*)_d (F - G) 1] \\
 A_3 = \phi^{-1} \{ & 3 \, 1^T (F - G) (Z^* - Z_2^*)_d (Z^* - Z_2^*) (Z^* - Z_2^*)_d (F - G) 1 \\
 & + 2 \, 1^T (F - G) (Z^* - Z_2^*)^{(3)} (F - G) 1 \},
 \end{aligned}$$

BIBLIOGRAFIA

- ABRAMOWITZ, M. & STEGUN, I.A., eds. (1970). *Handbook of Mathematical Functions with Formulas Graphs and Mathematical Tables*. Washington, National Bureau of Standards. 1046p. (National Bureau of Standards. Applied Mathematics Series, 55)
- AITCHISON, J. & SILVEY, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, **29**(3): 813-28.
- ATKINSON, A.C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *Journal of the Royal Statistical Society B*, **44**(1): 1-36.
- ATKINSON, A.C. & LAWRENCE, A.J. (1989). A comparison of asymptotic equivalent test statistics for regression transformation. *Biometrika*, **76**(2): 223-9.
- BARNDORFF-NIELSEN, O.E. (1986). Strings tensorial combinations, and Bartlett adjustments. *Proceedings of the Royal Society. London A*, **406**: 127-37.
- BARNDORFF-NIELSEN, O.E. (1990). Likelihood theory. In: HINKLEY, D.V., REID, N. & SNELL, E.J., eds. *Statistical Theory and Modelling: In Honor to Sir David Cox, FRS*. London, Chapman and Hall. pp. 232-64.
- BARNDORFF-NIELSEN, O.E. & BLAESILD, P. (1986). A note on the calculation of Bartlett adjustments. *Journal of the Royal Statistical Society B*, **48**(3): 353-8.

BARNDORFF-NIELSEN, O.E. & COX, D.R. (1984). Bartlett adjustments on the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society B*, 46(3): 483-95.

BARNDORFF-NIELSEN, O.E. & COX, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. London, Chapman and Hall, 252p.

BARNDORFF-NIELSEN, O.E. & HALL, P.(1988). On the level-error after Bartlett adjustment of the likelihood ratio statistics. *Biometrika*, 75(2): 374-8.

BARNDORFF-NIELSEN, O.E. & JØRGENSEN, B.(1990). Proper dispersion models. Aarhus, Department of Statistics - Aarhus University. (Research Report, 200)

BARTLETT, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of Royal Society of London A*, 160: 268-82.

BARTLETT, M.S. (1953a). Approximate confidence intervals. *Biometrika*, 40(3/4): 12-9.

BARTLETT, M.S. (1953b). Approximate confidence intervals II. More than one unknown parameter. *Biometrika*, 40(3/4), 306-17.

BERA, A.K. & McKENZIE, C.R.(1986). Alternative forms and properties of the score test. *Journal of Applied Statistics*, 13(1): 13-25. ✓

BERNDT, E.R. & SAVIN, N.E. (1977). Conflict among criteria for testing hypotheses in multivariate linear regression model. *Econometrica*, 45(5): 1263-78. ✓

BHATTACHARYA, R.N. & GHOSH, J.K. (1978). On the validity of formal Edgeworth expansions. *The Annals of Statistics*, 6(2): 434-51.

BHATTACHARYA, R.N. & GHOSH, J.K. (1988). On the moment conditions for valid formal Edgeworth expansions. *Journal of Multivariate Analysis*, 27(1): 68-79.

BISHOP, Y.M.M.; FIENBERG, S.E. & HOLLAND, P.W. (1985). *Discrete Multivariate*

- Analysis: Theory and Practice*. Cambridge, Mit Press, 557p.
- BOX, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**(3/4): 317-46.
- BRESLOW, N. (1989). Score tests in overdispersed GLM's. In: DECARLI, A. et al, eds. *Statistical Modelling: Proceedings of GLIM 89 and 4th International Workshop on Statistical Modelling, Trento, 1989*. New York, Springer, 1989. p.64-74. (Lecture Notes in Statistics, **57**).
- BREUSCH, T.S. (1979). Conflict among criteria for testing hypotheses: Extensions and comments. *Econometrica*, **47**(1): 203-7. ✓
- BREUSCH, T.S. & PAGAN, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, **47**: 239-53.
- BROWN, C.C. (1982). On a goodness of fit test for the logistic model based on score statistics. *Communications in Statistics: Theory and Methods*, **11**(10): 1087-1105.
- BUSE, A. (1982). The likelihood ratio, Wald and Lagrangian multiplier tests: An expository note. *The American Statistician*, **36**(3): 153-7. ✓
- CHAMBERS, J.M. (1977). *Computational Methods for Data Analysis*. New York, John Wiley. 268p. (Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics Section)
- CHANDRA, T.K. (1985). Asymptotic expansions of perturbed chi-square variables. *Sankhya A*, **47**(1): 100-10. ✓
- CHANDRA, T.K. & GHOSH, J.K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-squared variables. *Sankhya A*, **41**(1): 22-47. ✓
- CHANDRA, T.K. & JOSHI, S.N. (1983). Comparison of the likelihood ratio, Rao's and ✓

- Wald's tests and a conjecture of C.R. Rao. *Sankhya A*, **45**(2): 226-46.
- CHANDRA, T.K. & MUKERJEE, R. (1984). On the optimality of Rao's statistic. *Communications in Statistics: Theory and Methods*, **13**(12): 1507-15. ✓
- CHANDRA, T.K. & MUKERJEE, R. (1985). Comparison of the likelihood ratio, Wald's and Rao's tests. *Sankhya A*, **47**(2): 270-84. ✓
- CHANDRA, T.K. & MUKERJEE, R. (1991). Bartlett-type modification for Rao's efficient score statistic. *Journal of Multivariate Analysis*, **36**(1):103-12. ✓
- CHANONN, E. (1981). Matrix operations in GLIM 3. *GLIM News Letter*, **5**: 45-50.
- CHAPMAN, J.-A.W. (1976). A comparison of the X^2 , $-2 \log R$, and multinomial probability criteria for significance tests when expected frequencies are small. *Journal of the American Statistical Association*, **71**(356): 854-63.
- CHEN, C.-F. (1983). Score tests for regression models. *Journal of the American Statistical Association*, **78**(381): 158-61. ✓
- CHEN, C.-F. (1985). Robustness aspects of score tests for generalized linear and partially linear regression models. *Technometrics*, **27**(3): 277-83. ✓
- COOK, R.D. & WEISBERG, S. (1983). Diagnostics for heterocedasticity in regression. *Biometrika*, **70**(1): 1-10.
- CORDEIRO, G.M. (1982). *Improved Likelihood Ratio Statistics for Generalized Linear Models*. London. 227p. Tese(Doutorado)-Imperial College of Science and Technology-University of London.
- CORDEIRO, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B*, **45**(3): 404-13.
- CORDEIRO, G.M. (1985). The null expected deviance for an extended class of generalized

- linear models. In: GILCHRIST, R.; FRANCIS, B.; WHITTAKER, J., eds. *Generalized linear models: Proceedings of the GLIM 85 Conference. Lancaster, 1985*. Berlin, Springer, 1985. p.27-34. (Lecture Notes in Statistics, **32**)
- CORDEIRO, G.M. (1986a). A razão de máxima verossimilhança em problemas com parâmetros ortogonais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7, Campinas, 1986. *Atas*. Campinas, IMECC-UNICAMP, p.138-50.
- CORDEIRO, G.M. (1986b). *Modelos Lineares Generalizados*. Campinas, IMECC. 286p. Minicurso apresentado no 7^o Simpósio Nacional de Probabilidade e Estatística, Campinas, 1986.
- CORDEIRO, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**(2): 265-74.
- CORDEIRO, G.M. (1990). A general formula for computing Bartlett corrections. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9, São Paulo, 1990. *Atas*. São Paulo, IME-USP. p.72-7.
- CORDEIRO, G.M. & McCULLAGH, P. (1991). Bias Correction in generalized linear models. *Journal of the Royal Statistical Society B*, **53**. (A ser publicado)
- CORDEIRO, G.M. & PAULA, G.A. (1989a). Fitting non-exponential family nonlinear models in GLIM by using the offset facilities. In: DECARLI, A., ed. *Statistical Modelling: Proceedings of the GLIM 89 and the 4th International Workshop on Statistical Modelling, Trento, 1989*. New York, Springer, 1989. p.105-14. (Lecture Notes in Statistics, **57**)
- CORDEIRO, G.M. & PAULA, G.A. (1989b). Improved likelihood statistics for exponential nonlinear models. *Biometrika*, **76**(1): 93-100.
- CORDEIRO, G.M. & PAULA, G.A. (1990). *Estimation, large-sample parametric tests*

- and diagnostics for non-exponential family nonlinear models.* São Paulo, IME-USP, 1990. 31p. (RT-MAE-9015)
- CORDEIRO, G.M., PAULA, G.A. & BOTTER, D.A. (1991). *Improved likelihood tests for dispersion models.* São Paulo, IME-USP, 1991. 32p. (RT-MAE-9107)
- COX, D.R. (1984). Effective degrees of freedom and the likelihood ratio test. *Biometrika*, **71**(3): 487-93.
- COX, D.R. (1988). Some aspects of conditional and asymptotic inference: a review. *Sankhya A*, **50**(3): 314-37.
- COX, D.R. & HINKLEY, D.V. (1974). *Theoretical Statistics.* New York, John Wiley. 511p.
- COX, D.R. & REID, N. (1987a). Approximations to noncentral distributions. *The Canadian Journal of Statistics*, **15**(2): 105-14.
- COX, D.R. & REID, N. (1987b). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, **49**(1): 1-39.
- COX, D.R. & SNELL, E.J. (1981). *Applied Statistic.* London, Chapman and Hall, 189p.
- CRESSIE, N. & READ, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B*, **46**(3): 440-64.
- DAVISON, A. & TSAI, C.-L. (1990). Regression model diagnostics. *Preprint.* Department of Statistics, Oxford University.
- DAY, N.E. & BYAR, D.P. (1979). Testing hypotheses in case-control studies - Equivalence of Mantel-Haenzel statistics and logit score tests. *Biometrics*, **35**(3): 623-30.
- DEL PINO, G.E. (1990). A coordinate free-approach to score tests. *Communications in Statistics: Theory and Methods*, **19**(1): 155-67. ✓

- ENGLE, R.F. (1984). Wald, likelihood ratio and Lagrange multiplier tests in econometrics. In: GRILICHES, Z. & INTRILIGATOR, M., eds. *Handbook of Econometrics*. Amsterdam, North-Holland, pp. 776-826. (Handbook of Econometrics, 2)
- EVANS, G.B.A. & SAVIN, N.E. (1982). Conflict among and criteria revisited; the W, LR and LM tests. *Econometrica*, **50**: 737-48. ✓
- FIENBERG, S.E. (1979). The use of chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society B*, **41**(1): 54-64.
- FRYDENBERG, M. & JENSEN, J.L. (1989). Is the 'improved likelihood ratio statistic' really improved in discrete case? *Biometrika*, **76**(4): 655-61.
- GALLANT, A.R. (1987). *Nonlinear Statistical Models*. New York, John Wiley. 610p. (Wiley Series in Probability and Mathematical Statistics)
- GART, J.J. (1985). Approximate tests and interval estimates of the common relative risk in the combination of 2x2 tables. *Biometrika*, **72**(3): 673-7.
- GART, J.J. & TARONE, R.E. (1983). The relation between score tests and approximate UMPU tests in exponential models common in biometry. *Biometrics*, **39**(3): 781-6.
- GODFREY, L.G. (1981). On the invariance of the Lagrange multiplier test with respect to certain changes in alternative hypothesis. *Econometrica*, **49**(6): 1443-55. ✓
- GREEN, P.J. (1984). Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternatives. *Journal of the Royal Statistical Society B*, **46**(2): 149-92.
- HALDANE, J.B.S. (1937). The exact value of the moments of the distribution of χ^2 , used as a test of goodness of fit, when the expectations are small. *Biometrika*, **29**(2): 133-43.
- HARRIS, P. (1985). An asymptotic expansion for the null distribution of the efficient score ✓

- statistic. *Biometrika*, **72**(3): 653-9.
- HARRIS, P. (1987). Discussion of the paper by Professors Cox and Reid. *Journal of the Royal Statistical Society B*, **49**(1): 30.
- HARRIS, P. & PEERS, H.W. (1980). The local power of the efficient scores test statistics. *Biometrika*, **67**(3): 525-9. ✓
- HAYAKAWA, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika*, **62**(2): 451-60.
- HAYAKAWA, T. (1976). The new test criterion for homogeneity of parameters of several populations. *Annals of the Institute of Statistical Mathematics A*, **28**(3): 309-28.
- HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics A*, **29**(3): 359-78. ✓
- HAYAKAWA, T. & PURI, M.L. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics A*, **37**(1): 95-108. ✓
- HEALY, M.J.R. (1986). *Matrices for Statistics*. Oxford: Clarendon Press.
- HILL, G.W. & DAVIS, A.W. (1968). Generalized asymptotic expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics*, **39**(4): 1246-73.
- HINKLEY, D. (1985). Transformation diagnostics for linear models. *Biometrika*, **72**(3): 487-96.
- HINKLEY, D. (1988). More on score tests for transformation in regression. *Biometrika*, **75**(2): 366-9. ✓
- HONDA, Y. (1988). A size correction to the Lagrange multiplier test for heterocedasticity. *Journal of Econometrics*, **38**(3): 375-86. ✓
- HOSKING, J.R.M. (1980). Lagrange multiplier tests of time-series models. *Journal of the* ✓

Royal Statistical Society B, 42(2): 170-81.

HOSKING, J.R.M. (1981). Lagrange multiplier tests of multivariate time-series models. ✓

Journal of the Royal Statistical Society B, 43(2): 219-30.

HUZURBAZAR, V.S. (1950). Probability distributions and orthogonal paramaters. *Proceedings of the Cambridge Philosophical Society*, 46(2): 281-4.

IOCHVIDOV, I.S. (1982). *Hankel and Toeplitz Matrices and Forms: Algebraic Theory*. Boston, Birkhauser. 244p.

JENSEN, J.L. (1991). *A historical sketch and some new results on the improved log likelihood ratio statistic*. (Não publicado)

JENSEN, J.L. & KRISTENSEN, L.B. (1990). *Improved likelihood ratio tests for the gamma distribution*. Aarhus, Department of Theoretical Statistics-Aarhus University. 18p. (Research Reports, 204)

JØRGENSEN, B. (1983). Maximum likelihood estimation and large-sample inference in generalized linear and nonlinear regression models. *Biometrika*, 70(1): 19-28.

JØRGENSEN, B. (1984). The delta algorithm and GLIM. *International Statistical Review*, 52(3): 283-300.

JØRGENSEN, B. (1987). Small dispersion asymptotics. *REBRAPE: Revista Brasileira de Estatística*, 1(1): 59-90.

KAHLE, W. (1991). Estimation of the parameters of the Weibull distribution for several sample structures - asymptotic normality, likelihood ratio and Bartlett adjustment. (Não publicado).

KOEHLER, K.J. & LARNTZ, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*,

75(370): 336-44.

LARNTZ, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*, 73(362): 253-63.

LAWAL, H.B. (1984). Comparisons of the X^2 , Y^2 , Freeman-Tukey and Williams's improved G^2 test statistics in small samples of one-way multinomials. *Biometrika*, 71(2): 415-8.

LAWLESS, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. New York, John Wiley. 580p. (Wiley Series in Probability and Mathematical Statistics).

LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43(3/4): 295-303.

LAWRANCE, A.J. (1987). The score statistics for regression transformation. *Biometrika*, 74(2): 275-9.

LEE, E. T. (1980). *Statistical Methods for Survival Data Analysis*. Belmont, Lifetime Learning Publications, 557p.

LIANG, K.Y. & SELF, S.G. (1985). Tests for homogeneity of odds ratio when the data are sparse. *Biometrika*, 72(2): 353-8.

MALLER, R.A. (1986). Some asymptotic properties of the simple loglinear model. *Australian Journal of Statistics*, 28(1): 52-71.

MARGOLIN, B.H. & LIGHT, R.J. (1974). An analysis of variance for categorical data II: Small sample comparisons with chi-square and other competitors. *Journal of the American Statistical Association*, 69(347): 755-64.

MARDIA, K.V. (1972). *Statistics of Directional Data*. London, Academic Press. 357p. (Probability and Mathematical Statistics, 13).

- McCULLAGH, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika*, **71**(3): 461-76.
- McCULLAGH, P. (1987). *Tensor Methods in Statistics*. London, Chapman and Hall. 285p.
- McCULLAGH, P. (1989). Some statistical properties of a family of continuous univariate distributions. *Journal of the American Statistical Association*, **84**(405): 125-41.
- McCULLAGH, P. & COX, D.R. (1986). Invariants and likelihood ratio statistics. *The Annals of Statistics*, **14**(4): 1419-30.
- McCULLAGH, P. & NELDER, J.A. (1989). *Generalized Linear Models*. London, Chapman and Hall, 511p. (Monographs on Statistics and Applied Probability)
- MITCHELL, A.F.S.(1962). Sufficient statistics and orthogonal parameters. *Proceedings of the Cambridge Philosophical Society*, **58**(2): 326-37.
- MØLLER, J. (1986). Bartlett adjustments for structured covariances. *Scandinavian Journal of Statistics*, **13**(1): 1-15.
- MUKERJEE, R. (1990a). Comparison of tests in the multiparameter case I: Second-order power. *Journal of Multivariate Analysis*, **33**(1): 17-30.
- MUKERJEE, R. (1990b). Comparison of tests in the multiparameter case II: A third-order optimality property of Rao's test. *Journal of Multivariate Analysis*, **33**(1): 31-48.
- NELDER, J.A. & WEDDERBURN, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**(3): 370-84.
- NEYMAN, J. (1959). Optimal asymptotic tests of composite hypotheses. In: GRENNANDER, U. ed. *Probability and Statistics*. New York, John Wiley. p.213-34.
- NEYMAN, J. (1979). $C(\alpha)$ tests and their use. *Sankhya A*, **41**(1): 1-21.
- O'QUIGLEY, J. & PESSIONE, F. (1989). Score tests for homogeneity of regression effect

- in the proportional hazards model. *Biometrics*, **45**(1): 135-144.
- OHTANI, K. & TOYODA, T. (1985). A Monte Carlo study of the Wald, LM and LR tests in a heterocedastic linear model. *Communication in Statistics: Simulation and Computation*, **14**(3): 735-46.
- PAULA, G.A. (1988). *Correções de Bartlett e Métodos de Diagnóstico para Modelos Não-Lineares da Família Exponencial*. São Paulo, 117p. Tese(Doutorado). Instituto de Matemática e Estatística da USP.
- PAYNE, C.D., ed. (1986). *The GLIM Manual: Release 3.77*. Oxford, NAG.
- PEERS, H.W. (1971). Likelihood ratio and associated test criteria. *Biometrika*, **58**(3): 577-87.
- PORTEOUS, B.T. (1985). Improved likelihood ratio statistics for covariance selection models. *Biometrika*, **72**(1): 97-101.
- POSKITT, D.S. & TREMAYNE, A.R. (1981). An approach to testing linear time series models. *The Annals of Statistics*, **9**(5): 974-86.
- POSKITT, D.S. & TREMAYNE, A.R. (1982). Diagnostic tests for multiple time series models. *The Annals of Statistics*, **10**(1): 114-20.
- PREGIBON, D. (1982). Score tests in GLIM with applications. In: GILCHRIST, R., ed. *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, 1, London, 1982. Berlin, Springer. p.87-97. (Lecture Notes in Statistics, 14).
- RAO, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**: 50-7. ✓
- RAO, C.R. (1973). *Linear Statistical Inference and its Applications*. 2.ed. New York,

John Wiley. 625p. (Wiley Series in Probability and Mathematical Statistics)

RATKOWSKY, D.A. (1983). *Nonlinear Regression Modelling: A Unified Practical Approach*. New York, Marcel Dekker. 276p. (Statistics, Textbooks and Monographs, 48)

RAYNA, G. (1987). *REDUCE: Software for Algebraic Computation*. New York, Springer. 329p. (Symbolic Computation, Artificial Intelligence)

REID, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, 3(2): 213-38.

ROSS, W.H. (1987). The expectation of the likelihood ratio criterion. *International Statistical Review*, 55(3): 315-30.

SAVIN, N.F. (1976). Conflict among test procedures in a linear regression model with autoregressive disturbances. *Econometrica*, 44: 1303-15.

SCHIFFMAN, D.A. (1986). The score statistic in constancy testing for a discrete hazard rate. *I.E.E.E. Transactions on Reliability*, 35(5): 590-4.

SILVEY, S.D. (1959). The Lagrangian multiplier test. *Annals of Mathematical Statistics*, 30(2): 389-407. ✓-

STEFANSKY, L.A. & CARROL, R.J. (1990). Score tests in generalized linear measurement error models. *Journal of the Royal Statistical Society B*, 52(2): 345-59.

TARONE, R.E. (1986). Tests for comparing a sequence of Bernoulli random variables to a sequence of known probabilities. *Communications in Statistics: Theory and Methods*, 15(3): 981-97.

TARONE, R.E. (1988). Homogeneity score tests with nuisance parameters. *Communications in Statistics: Theory and Methods*, 17(5): 1549-56.

- TARONE, R.E. & GART, J.J. (1989). Significance tests for cancer screening trials. *Biometrics*, 45(3): 883-90.
- TOSTESON, T.D. & TSIATS, A.A. (1988). The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika*, 75(3): 507-14.
- TSAI, C.-L. (1988). Power transformation and reparameterizations in nonlinear regression models. *Technometrics*, 30(4): 441-8.
- WANG, P.C. (1985). Adding a variable in generalized linear models. *Technometrics*, 27(3): 273-6.
- WILLIAMS, D.A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika*, 63(1): 33-7.
- WILMOTT, G.E. (1990). On the construction of a parameter orthogonal to the mean. *Biometrika*, 77(2):424-8.