Testes para Hipóteses Restritas
em Modelos de Regressão
Log-gama Generalizado e Estrutural

Oscar Villanueva Rojas

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU
DE
DOUTOR EM ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Gilberto Alvarenga Paula**

-São Paulo, junho de 1996-

*aos meus pais Margarita e César Julio
à minha esposa Nefer
aos meus filhos
Paola Andrea, Oscar Andrés e Liseth Karina.*

AGRADECIMENTOS

Eu gostaria de agradecer ao Prof. Dr. Gilberto Alvarenga Paula pela confiança em mim depositada e especialmente pelo oportuno, decidido e constante apoio durante o desenvolvimento deste trabalho.

Ao Prof. Dr. Heleno Bolfarine pelas valiosas sugestões e pelo incentivo durante todo o curso.

A meus amigos e colegas de turma: Antonio, Filidor e Miguel; a meu compatriota Luis Eduardo; a meus amigos Manuel e Patricia, Fabio e Olga, Manoel e Claudia e Victor e Isabel pelo companherismo e pela amizade que ofertaram a mim e a minha família.

Também gostaria de agradecer à Universidad de la Amazonia por ter-me liberado durante o tempo necessário para culminar o meu Doutorado e a meus colegas de trabalho do Departamento de Matemáticas y Física pela credibilidade que depositaram em mim.

Finalmente, gostaria de agradecer a todos aqueles que colaboraram de uma forma ou de outra para a realização deste trabalho.

À Capes (Brasil) e a Colciencias (Colombia) pelo apoio financeiro.

RESUMO

Neste trabalho estudamos a teoria de testes do tipo *one-sided* em duas classes de modelos de regressão com grande aplicabilidade em diversas áreas do conhecimento. Discutimos inicialmente a teoria para modelos gerais de regressão e sob determinadas condições de regularidade mostramos que testes do tipo *one-sided* podem ser estendidos para modelos de regressão com distribuição log-gama generalizada com parâmetro de escala não-constante. Obtemos também a distribuição nula assintótica de estatísticas assintoticamente equivalentes. Para alguns casos particulares de hipóteses restritas, tais como ordem simples, ordem simples em forma de árvore, ordem do tipo *loop* e ordem unimodal em g populações e para os interceptos em modelos de regressão de retas paralelas mostramos que a distribuição nula assintótica reduz para a distribuição qui-quadrado barra, $\bar{\chi}^2$. Numa segunda etapa e de forma similar ao caso inicial, investigamos e aplicamos essa teoria ao modelo de regressão estrutural com erros nas variáveis. Apresentamos alguns estudos de simulação e ilustramos a metodologia proposta com exemplos da literatura.

ABSTRACT

This manuscript discusses the application of one-sided tests in two general classes of regression models with large application in many areas of knowledge. We first give a general approach for regression models and discuss the extension of the methodology to generalized log-gamma regression models with non constant scale parameter. We also obtain the asymptotic null distribution of statistics asymptotically equivalent. For some particular cases, such as that of simple order, simple tree order, loop ordering and unimodal ordering in g populations as well as for the intercepts in parallel regression line models we show that the asymptotic null distribution reduces to the well-known chi-bar-squared distribution, $\tilde{\chi}^2$. In the second part of the work we investigate the application of the methodology in structural linear regression models. Some simulation studies that investigate approximations for the level probabilities are also given. Finally, two examples from the statistical literature illustrate applications of the methodology.

Índice

1	Introdução	7
1.1	Generalidades	7
1.2	Regressão normal linear	10
1.3	Modelos lineares generalizados	14
1.4	Tópicos de pesquisa	17
1.5	Organização do trabalho	19
2	Hipóteses na forma de desigualdades não-lineares	21
2.1	Introdução	21
2.2	Distância de Kodde e Palm	22

	2
2.2.1 Distribuição nula assintótica de D	25
2.3 Condições de regularidade de Gourieroux e Monfort	27
2.3.1 Estimador restrito por desigualdades	28
2.3.2 Convergência	29
2.3.3 Teste da hipótese $H_0 : h(\theta) = 0$ contra $H_1 : h(\theta) \geq 0$, $h(\theta) \neq 0$ (caso I)	30
2.3.4 Equivalência assintótica das estatísticas do teste	30
2.3.5 Distribuição nula assintótica das estatísticas	36
2.3.6 Teste da hipótese $H_0 : h(\theta) \geq 0$ contra $H_1 : \theta \in R^p$ (caso II)	43
2.4 Lema de Wolak	44
2.4.1 Exemplo	49
2.5 Probabilidades de nível	51
2.5.1 Introdução	51
2.5.2 Método da redução da dimensão	52

	3
2.5.3	Outros métodos de aproximação 55
2.5.4	Aproximação da probabilidade do quadrante por séries infinitas 56
3	Hipóteses restritas em modelos de regressão com distribuição log-gama generalizada 67
3.1	Introdução 67
3.2	Distribuição log-gama generalizada 68
3.3	Modelo de regressão log-gama generalizado 73
3.4	Hipóteses restritas 76
3.4.1	Ordem simples 78
3.4.2	Ordem simples em forma de árvore 80
3.4.3	Hipótese unimodal 82
3.4.4	Hipótese do tipo <i>loop</i> 84
3.5	Modelo de regressão do valor extremo 85
3.5.1	Modelo de regressão de retas paralelas 87

	4
3.5.2 Hipóteses de ordem para os percentis	89
3.5.3 Hipótese nula restrita	92
3.6 Estudo da robustez da distribuição $\tilde{\chi}^2$	93
3.7 Aplicações	97
3.7.1 Desempenho de turbinas	97
3.7.2 Experimento com vírus	100
4 Hipóteses restritas no modelo de regressão estrutural	105
4.1 Introdução	105
4.2 Parametrização ortogonal	107
4.3 Estimadores de máxima verossimilhança e matriz de informação	109
4.4 Hipóteses restritas	112
4.5 Coeficiente de confiabilidade conhecido	113
4.5.1 Ordem simples	114
4.5.2 Ordem simples na forma de árvore	116

	5
4.6 Razão de variâncias conhecida	117
4.6.1 Ordem simples	118
4.6.2 Ordem simples na forma de árvore	120
4.7 Simulações	121
5 Conclusões	128
5.1 Futuras pesquisas	134
Apêndice	135
A Aproximações para $P(\ell, k; w)$	135
A.0.1 Um algoritmo geral para calcular $P(\ell, k; w)$	136
A.0.2 Ordem simples com pesos arbitrários	139
A.0.3 Ordem simples com pesos iguais	141
A.0.4 Ordem simples na forma de árvore	143
A.0.5 Ordem do tipo <i>loop</i>	145

A.1 Aproximações para as probabilidades de nível	150
B Condições de regularidade de Bradley e Gart	156
B.1 Introdução	156
B.2 Consistência	158
B.3 Normalidade assintótica	160
C Resultados da Simulação	162
D Programas	171
Bibliografia	179

Capítulo 1

Introdução

1.1 Generalidades

A teoria de testes do tipo *one-sided* em regressão linear é datada da década de 60 (vide Kudô, 1963). Todavia, a extensão do método para situações mais gerais, tais como hipóteses com desigualdades não-lineares em modelos não-lineares, é bastante recente (vide Kodde e Palm, 1986). Nos últimos 30 anos, um grande número de artigos e pelo menos dois livros foram publicados na área de Inferência com Restrições de Ordem. Uma excelente referência é o livro de Robertson, Wright e Dykstra (1988). O estudo da distribuição nula do teste da razão de verossimilhança para hipóteses com restrições de ordem em populações normais é talvez o principal tópico no assunto. Mostra-se que a distribuição nula do teste da razão de verossimilhança é uma mistura de qui-quadrados (ou de distribuições do tipo F), com pesos que não dependem dos parâmetros sob a hipótese nula. Esses pesos não são em geral expressos em

forma fechada para quatro ou mais restrições. O uso de métodos de integração numérica (Bohrer e Chow, 1978; Sun, 1988), de aproximações para os pesos (Lee, Robertson e Wright, 1993; Singh e Wright, 1986, 1993; Lucas, Robertsom e Wright, 1989; Robertsom e Wright, 1983; Wright e Tran, 1985) ou o estudo de situações de importância prática em que os pesos tomam formas simplificadas (Robertson, Wright e Dykstra, 1988), têm sido alguns dos tópicos de pesquisa no assunto. A extensão dos resultados de k populações normais para a família exponencial, bem como para outras distribuições, por exemplo a multinomial, tem sido também largamente estudada. Nesses casos, sob condições gerais de regularidade, o teste da razão de verossimilhança tem distribuição nula assintótica que é uma mistura de qui-quadrados com pesos em forma similar aos do caso normal.

O estudo do teste da razão de verossimilhança para hipóteses com restrições de ordem é motivado pelo fato desse teste ser mais poderoso do que seus competidores. Pode ser mostrado que a utilização da informação de que os parâmetros estão restritos, na formulação das hipóteses, frequentemente leva a consideráveis ganhos no poder do teste. Bartholomew (1961) calculou o poder do teste de hipóteses com restrições de ordem em populações normais e mostrou que o poder do teste da razão de verossimilhança é superior ao poder dos testes usuais baseados na distribuição qui-quadrado. A confirmação desses resultados tem sido feita por vários autores, entre eles, Robertson e Wegman (1978); Hillier (1986); Robertson, Wright e Dykstra (1988) e Paula e Sen (1994, 1995).

Alguns artigos com resultados relevantes relacionados com testes *one-sided* em regressão foram publicados nos últimos anos. Silvapulle (1991) tem in-

vestigado a aplicação de um tipo particular de teste *one-sided* em modelos de regressão com log-verossimilhança côncava. Entre esses modelos estão os modelos lineares generalizados com ligação canônica e alguns modelos para a análise de dados de sobrevivência. Wolak (1991) tem mostrado para uma classe geral de modelos de regressão, que se a hipótese nula é expressa na forma de desigualdades não-lineares, não necessariamente a situação menos favorável ocorre quando todas as restrições são satisfeitas em forma de igualdades. Esse resultado se verifica, em particular, para o caso usual de regressão normal linear com desigualdades lineares (Perlman, 1969). Wolak (1991) propôs um lema em que define um subconjunto da hipótese nula que contém a situação menos favorável. Em geral, pode ser relativamente complexo obter tal subconjunto. Primeiro, pelo fato da distribuição nula assintótica, que continua sendo uma mistura de qui-quadrados, ter uma dimensão que depende da região menos favorável. Segundo, devido a dificuldade de estimar os pesos, uma vez que esses dependem dos parâmetros sob a hipótese nula, o que levaria a situação menos favorável a ser obtida quando todas as desigualdades forem satisfeitas em forma de igualdades. Wolak (1989) sugere, alternativamente, a utilização de testes para hipóteses locais onde os resultados assintóticos se verificam sob condições gerais de regularidade. Nas seções seguintes procuramos descrever brevemente os principais resultados relacionados com testes do tipo *one-sided* em regressão, o que propomos investigar no assunto, bem como a forma da apresentação do trabalho. Na Seção 1.2 fazemos uma revisão para o modelo normal linear com alguns resultados recentes. A extensão para modelos lineares generalizados é discutida na Seção 1.3. A Seção 1.4 é dedicada aos tópicos de pesquisa e na seção final descrevemos a forma como será apresentado o trabalho.

1.2 Regressão normal linear

Sejam $Y_1, \dots, Y_n$ variáveis aleatórias independentes e normalmente distribuídas com variância constante σ^2 e médias $E(Y_i) = \mu_i$, $i = 1, \dots, n$. Adicionalmente, suponha que μ_i está relacionado linearmente com $x_i = (x_{i1}, \dots, x_{ip})^t$, isto é $\mu(x_i) = \sum_1^p x_{ij}\beta_j$, onde $x_{i1}, \dots, x_{ip}$ representam valores conhecidos de p variáveis explicativas.

As hipóteses usuais do tipo *one-sided* são definidas por

$$H_0 : R\beta = r \text{ contra } H_1 : R\beta \geq r,$$

com pelo menos uma desigualdade estrita sob H_1 . A matriz R tem dimensão $q \times p$ ($q \leq p$), e é assumida de posto completo, $\beta = (\beta_1, \dots, \beta_p)^t$ são parâmetros a serem estimados e r é um vetor de constantes em R^q . Inicialmente, Bartholomew (1959 a, b) mostrou que a distribuição nula da estatística da razão de verossimilhança $\mathcal{W}$ é uma mistura de qui-quadrados e provou que os testes baseados nessa estatística são uniformemente mais poderosos do que os testes tradicionais usados para testar igualdade contra diferença. Aplicando resultados da distribuição normal multivariada obtidos por David (1953) e McFadden (1956), Bartholomew apresenta formas fechadas para os pesos no caso de $q \leq 3$. Essas probabilidades são em geral muito complexas de serem obtidas para $q \geq 4$.

Kudô (1963) e independentemente Nüesch (1966) provaram esses resultados para o quadrante não-negativo. Kudô define a estatística $\mathcal{W}$, discute sua existência e natureza geométrica, propõe um algoritmo para seu cálculo e mos-

tra que sua distribuição nula é uma mistura de qui-quadrados

$$P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^q Q(q, \ell; \Delta) P_r\{\chi_\ell^2 \geq c\}, \quad (1.1)$$

onde $c \geq 0$, χ_0^2 é a distribuição degenerada na origem, $\Delta = R(X^T X)^{-1} R^t$ e $Q(q, \ell; \Delta)$ denota a probabilidade do vetor $R\hat{\beta}^* - r$ ter exatamente ℓ valores positivos, com $\hat{\beta}^*$ sendo a estimativa de máxima verossimilhança sob a hipótese H_1 . Esses resultados são estendidos para o caso em que a matriz de covariância é conhecida.

Tem havido também interesse por problemas de teste de hipóteses restritas por desigualdades lineares ou não-lineares em modelos não-lineares. Particularmente Gourieroux, Holly e Monford (1982) consideram o problema de testar hipóteses em modelos de regressão com restrições em desigualdades lineares sobre os coeficientes. Supondo que os erros seguem uma distribuição normal com média μ e matriz de covariância Σ , conhecida e não-singular, eles definem o teste estatístico do multiplicador de Kuhn-Tucker e estabelecem as relações com o teste da razão de verossimilhança, de Wald e de escore. É enfatizado que a distribuição da estatística comum $\mathcal{W}$, sob a hipótese nula, é uma mistura de distribuições do tipo qui-quadrado. Se Σ é desconhecida e depende de um número finito de parâmetros, e se $\Sigma = \sigma^2 \Omega$ com σ^2 desconhecido, e supondo a existência e unicidade dos estimadores de máxima verossimilhança para σ , Ω , e β , Gourieroux et al. (1982) demostram que a distribuição assintótica da estatística da razão de verossimilhança sob H_0 , continua sendo uma mistura de distribuições do tipo qui-quadrado. Tabelas com alguns valores críticos para níveis de significância de 5% e 10% são fornecidas.

Outra situação de interesse prático é a seguinte:

$$H_1 : R\beta \geq r \text{ contra } H_2 - H_1,$$

onde $H_2 : \beta \in R^p$. Nesse caso, Perlman (1969) tem mostrado quando Σ é conhecida ou é desconhecida porém não depende de β que a situação menos favorável sob H_1 é dada por

$$P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^q Q(q, \ell; \Delta) P_r\{\chi_{q-\ell}^2 \geq c\}, \quad (1.2)$$

que coincide com a distribuição dada em (1.1), com χ_ℓ^2 substituído por $\chi_{q-\ell}^2$.

Kodde e Palm (1986) derivam um teste aproximado, que chamam de distância D para testar hipóteses da forma geral $H_0 : h_1(\theta) = 0, h_2(\theta) = 0$ contra $H_1 : h_1(\theta) \neq 0, h_2(\theta) \geq 0$ e $H_0 : h_1(\theta) = 0, h_2(\theta) \geq 0$ contra $H_1 : \theta \in R^k$ onde $h_1(\theta)$ e $h_2(\theta)$ são subvetores de $h(\theta)$.

Wolak (1991) deriva várias propriedades para testes de hipóteses nulas restritas por desigualdades lineares ou não-lineares as quais elucidam a utilidade prática desses testes. Essas propriedades aparecem em virtude da dependência funcional de $V(\theta)$, matriz de covariância assintótica de $\sqrt{n}(\hat{\theta} - \theta_0)$, em θ_0 , valor verdadeiro de θ , onde $\hat{\theta}$ é um estimador consistente de θ , assintoticamente normal. Três características são particularmente anotadas por Wolak (1991). Uma que limita a utilidade prática da maioria dessas técnicas de teste, é a falta de um procedimento facilmente aplicável para calcular, assintoticamente, o valor crítico exato. Para testes de hipóteses nulas restritas por desigualdades lineares ou não-lineares, Wolak mostra que a distribuição nula assintótica menos favorável pode unicamente ser limitada a um conjunto de possíveis distribuições e não a uma única distribuição como no caso de testes restritos por

desigualdades lineares em modelos lineares. Uma segunda propriedade, a qual também dificulta o conteúdo prático dessas técnicas é a ausência de limites estreitos superior e inferior para a distribuição nula assintótica. A terceira propriedade é o surpreendente resultado que a distribuição nula assintótica menos favorável pode não ocorrer no valor único do parâmetro que satisfaz todas as restrições por desigualdades com igualdade.

A distribuição nula da estatística da razão de verossimilhança quando σ^2 é substituído pela estimativa usual s^2 , é uma mistura de distribuições do tipo F, tanto em (1.1) como em (1.2) (vide Wolak, 1987). Esse resultado contrasta com aqueles obtidos por Gourieroux et al. (1982) que substituem σ^2 pela correspondente estimativa de máxima verossimilhança e verificam que os resultados valem apenas assintoticamente. Simplificações e aproximações para as probabilidades $Q(q, \ell; \Delta)$'s são em geral obtidas quando a matriz de correlação associada à matriz Δ não depende dos parâmetros sob a hipótese nula. Essa propriedade se verifica, em particular, para o caso de k populações.

Existem algumas situações de interesse prático em que a teoria é bastante desenvolvida. Por exemplo, suponhamos que temos $Y_{1j}, \dots, Y_{kj}$ variáveis aleatórias independentes, tais que $Y_{ij} \sim N(\mu_i, \sigma^2)$, σ^2 conhecido, $i = 1, \dots, k$ e $j = 1, \dots, r_{ij}$. Podemos estar interessados em testar

$$H_0 : \mu_1 = \dots = \mu_k \text{ contra } H_1 : \mu_1 \leq \dots \leq \mu_k,$$

com pelo menos uma desigualdade estrita sob H_1 . A distribuição nula da estatística da razão de verossimilhança, também conhecida como qui-quadrado barra (vide Barlow e al., 1972; Robertsom et al., 1988), é dada por

$$Pr\{\bar{\chi}_{01}^2 \geq c\} = \sum_{\ell=1}^k P(\ell, k; w) Pr\{\chi_{\ell-1}^2 \geq c\}, \quad (1.3)$$

onde $w = (w_1, \dots, w_k)^t$ denotam os pesos e nesse caso são definidos por $w_i = r_i$. Aqui, os níveis de probabilidade $P(\ell, k; w) = Q(k-1, \ell-1; \Delta)$ não dependem das médias μ_i sob H_0 . No entanto, a dificuldade de computação permanece, especialmente para $k \geq 5$.

Outra situação de interesse é a seguinte:

$$H_0 : \mu_1 = \dots = \mu_k \text{ contra } H_1 : \mu_1 \leq [\mu_2, \dots, \mu_k],$$

onde $[\mu_2, \dots, \mu_k]$ denota qualquer ordem e é assumido pelo menos uma desigualdade estrita em H_1 . A distribuição nula da estatística da razão de verossimilhança é a mesma dada em (1.3), no entanto os coeficientes de correlação associados a Δ são diferentes, porém dependem somente dos pesos.

Aproximações, algoritmos e limites para as probabilidades $P(\ell, k; w)$ bem como para as distribuições $\bar{\chi}^2$ e $\bar{E}^2$ (quando σ^2 é desconhecido) em situações particulares são dadas em Bohrer e Chow (1978), Sun (1988), Chase (1974), Sis-kind (1976), Grove (1980), Robertson e Wright (1983), Sasabuchi e Kulatunga (1985), Wright e Tran (1985), Singh e Wright (1986), Lucas, Robertson e Wright (1989), Singh e Wright (1993), Lee, Robertson e Wright (1993) e Robertson, Wright e Dykstra (1988).

1.3 Modelos lineares generalizados

Suponha $Y_1, \dots, Y_n$ variáveis aleatórias independentes tais que Y_i pertence à família exponencial de distribuições. A média e a variância de Y_i são dadas por $E(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \phi^{-1}V_i$, onde ϕ^{-1} denota o parâmetro de dispersão

e $V_i = V(\mu_i)$ é a função de variância. Adicionando a suposição de que $g(\mu_i) = \eta_i$, onde $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$, tem-se os Modelos Lineares Generalizados (vide McCullagh e Nelder, 1989). Silvapulle (1994) tem estudado a distribuição nula assintótica da estatística da razão de verossimilhança para as seguintes hipóteses:

$$H_0 : h_1(\beta) = 0, \quad h_2(\beta) = 0 \quad \text{contra} \quad H_1 : h_1(\beta) \neq 0, \quad h_2(\beta) \geq 0,$$

onde $h(\beta)$ é um vetor k -dimensional, enquanto $h_1(\beta)$ e $h_2(\beta)$ são subvetores com dimensões k_1 e $k - k_1$, respectivamente. Quando $k_1 = 0$, há pelo menos uma desigualdade estrita em H_1 . É assumido que $\Pi(\beta) = \partial h(\beta)/\partial \beta$ é contínua e de posto completo. Silvapulle (1994) propôs um lema bastante geral para a obtenção da distribuição nula assintótica da estatística da razão de verossimilhança no caso acima. Em particular, o lema proposto pode também ser estendido para os MLGs com ligação canônica sob as condições de regularidade (D) e (N) dadas em Fahrmeir e Kaufmann (1985). Com a suposição adicional de que a log-verossimilhança é côncava e usando os resultados de Kodde e Palm (1986, caso II), mostra-se assintoticamente que a distribuição nula da estatística da razão de verossimilhança é uma mistura de qui-quadrados

$$P_r\{\mathcal{W} \geq c\} \cong \sum_{\ell=0}^{k-k_1} Q(\ell, k - k_1; \Delta(\beta)) P_r\{\chi_{\ell+k_1}^2 \geq c\},$$

onde $\Delta(\beta) = \Pi_2(\beta)(X^T V X)^{-1} \Pi_2(\beta)^T$, $\Pi_2(\beta) = \partial h_2(\beta)/\partial \beta$, $V = \text{diag}\{v_1, \dots, v_n\}$ e as demais quantidades são as mesmas da regressão linear. É importante lembrar que a concavidade da log-verossimilhança é garantida para MLGs com ligação canônica. Na prática, a matriz $\Delta(\beta)$ pode ser estimada substituindo β pela estimativa irrestrita $\hat{\beta}$ embora não seja garantido que a distribuição obtida corresponda à situação menos favorável. A simplificação das probabilidades $Q(\ell, k - k_1; \Delta(\beta))$ continua sendo um tópico relevante de

pesquisa, especialmente para $k - k_1 \geq 5$. O outro caso de hipóteses do tipo *one-sided* é o seguinte:

$$H_1 : h(\beta) \geq 0 \text{ contra } H_2 = H_1,$$

onde $H_2 : \beta \in R^p$. Conforme mencionado anteriormente, Wolak (1991) tem mostrado para uma classe geral de modelos não-lineares que devido a dependência de $\Delta(\beta)$ em β , a distribuição nula assintótica menos favorável da estatística da razão de verossimilhança pode não ser alcançada quando $h(\beta) = 0$. Definindo o subconjunto de $h(\beta)$, $h_b(\beta) = (h_{b_1}(\beta), \dots, h_{b_m}(\beta))^t$, onde $1 \leq m \leq k$ e $1 \leq b_j \leq k$, Wolak (1991) mostra para grandes amostras que a distribuição nula sob a hipótese $h_b(\beta) = 0$, tal que $h_{\bar{b}}(\beta) > 0$, é dada por

$$Pr\{\mathcal{W} \geq c\} \cong \sum_{\ell=0}^m Q(\ell, m; \Delta_b(\beta)) Pr\{\chi_{m-\ell}^2 \geq c\},$$

onde $h_{\bar{b}}(\beta)$ é o complemento de $h_b(\beta)$, $\Delta_b(\beta) = \Pi_b(\beta)(X^T V X)^{-1} \Pi_b(\beta)^t$ e $\Pi_b(\beta) = \partial h_b(\beta) / \partial \beta$ tem dimensão $m \times p$. É possível mostrar apenas que a hipótese menos favorável ocorre quando $h_b(\beta) = 0$. O problema é definir esse subconjunto. Não obstante, Wolak (1991) tem encontrado para $c \geq 0$ os limites

$$\frac{1}{2} Pr\{\chi_1^2 \geq c\} \leq Pr\{\mathcal{W} \geq c\} \leq \frac{1}{2} Pr\{\chi_{k-1}^2 \geq c\} + \frac{1}{2} Pr\{\chi_k^2 \geq c\},$$

que em geral são conservadores, especialmente quando k cresce. Kodde e Palm (1986) apresentam os valores críticos correspondentes a vários níveis de significância. Um resultado importante que pode ser aplicado aqui (vide Perlman, 1969), é o fato da distribuição nula menos favorável ser alcançada quando $m = k$ nos casos em que $\Delta(\beta)$ não depende de β . Aplicações desse resultado são apresentadas em Paula e Sen (1994, 1995).

1.4 Tópicos de pesquisa

Como foi mencionado anteriormente, a concavidade de alguns modelos paramétricos tem facilitado o desenvolvimento da teoria assintótica de testes do tipo *one-sided* em regressão. Nesse sentido, propomos investigar inicialmente a teoria de testes *one-sided* para a seguinte classe de modelos de regressão:

$$Y_i = x_i^t \beta + \sigma Z_i, \quad i = 1, \dots, n, \quad (1.4)$$

onde $x_i^t = (x_{i1}, \dots, x_{ip})$, $\beta = (\beta_1, \dots, \beta_p)'$, $\sigma > 0$, com os erros Z_i sendo independentes e igualmente distribuídos com distribuição log-gama generalizada, cuja densidade é dada por (ver Lawless, 1980)

$$f(z) = \frac{\lambda^{2\lambda^2-1} \exp[\lambda z + \lambda^2 c(\lambda) - \lambda^2 \exp\{\lambda^{-1} z + c(\lambda)\}]}{\Gamma(\lambda^2)},$$

onde $-\infty < z < \infty$, $c(\lambda) = \psi(\lambda^2) - 2\log\lambda$ e $\psi(x) = \partial \log \Gamma(x) / \partial x$ é a função digama. O parâmetro β é desconhecido, porém λ é assumido conhecido. Quando $\lambda = 1$, a densidade acima reduz à distribuição do valor extremo padrão e para $\lambda \rightarrow \infty$ tem-se a distribuição normal padrão. Testes do tipo *one-sided* envolvendo β são investigados aqui. A extensão dos resultados para o caso de dados censurados poderia ser também investigada. Em particular, no caso de réplicas para os valores das variáveis explicativas, teremos uma outra versão para (1.4) dada abaixo

$$Y_{ij} = x_i^T \beta + \sigma_i Z_{ij},$$

$i = 1, \dots, g$, $j = 1, \dots, n_i$, onde $\sigma_1, \dots, \sigma_g$ são desconhecidos e Z_{ij} segue uma distribuição log-gama generalizada com parâmetro λ desconhecido. Aqui, é de interesse investigar hipóteses restritas envolvendo $\sigma_1, \dots, \sigma_g$, como por exemplo testar $H_0 : \sigma_1 = \dots = \sigma_g$ contra $H_1 : \sigma_1 \leq \dots \leq \sigma_g$, com pelo menos

uma desigualdade estrita em H_1 . Esse tipo de modelo tem sido largamente aplicado na análise de dados de sobrevivência. Nessa situação particular de importância prática procuramos formas simplificadas para os pesos. Muitos dos resultados desenvolvidos nesse assunto, particularmente para $\lambda = 1$, são descritos também em Paula e Rojas (1996).

Numa segunda etapa, o objetivo é a inferência sobre o parâmetro de inclinação no modelo estrutural de regressão normal com erros nas variáveis. Consideramos g populações, tal que para a população i , existe uma variável aleatória bivariada (y_i, x_i) , tal que

$$y_{ij} = \alpha_i + \beta_i x_{ij},$$

$j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, g$, α_i e β_i serão estimados e x_{ij} e y_{ij} não são observados diretamente. Devido aos problemas de identificabilidade implícitos no modelo, ele é considerado em suas duas versões: $\lambda_{\epsilon i} = \frac{\sigma_{\epsilon i}^2}{\sigma_{u i}^2}$ conhecida e $k_{x i} = \frac{\sigma_{x i}^2}{\sigma_{x i}^2 + \sigma_{u i}^2}$ conhecida. Aqui é de interesse investigar hipóteses restritas envolvendo $\beta_1, \dots, \beta_g$, como por exemplo

$$H_0 : \beta_1 = \dots = \beta_g \text{ contra } H_1 : \beta_1 \leq [\beta_2, \dots, \beta_g].$$

Esse modelo tem sido usado em múltiplas aplicações industriais. Parte dos resultados descritos nesse assunto podem também ser encontrados em Rojas, Paula e Bolfarine (1995).

Vários estudos de simulação foram realizados com o objetivo de estudar a conveniência do uso de aproximações mais simples para os coeficientes $Q(k, \ell; \Delta)$'s em situações de interesse prático.

1.5 Organização do trabalho

Este trabalho é organizado em quatro capítulos e quatro apêndices. O Capítulo 2 é dedicado ao estudo de teste de hipóteses restritas por desigualdades não-lineares. Nesse capítulo definimos as condições gerais de regularidade e as diversas estatísticas do teste para as quais se estabelecem as equivalências assintóticas. Também é demonstrado que a distribuição nula assintótica da estatística do teste é uma mistura de distribuições do tipo qui-quadrado. Definimos as probabilidades de nível, suas simplificações e aproximações.

No Capítulo 3, estendemos os resultados do Capítulo 2 para o modelo de regressão com distribuição log-gama generalizada. Definimos as estatísticas do teste e mostramos que a distribuição nula assintótica continua sendo uma mistura de distribuições do tipo qui-quadrado. É mostrado que as probabilidades de nível dependem unicamente dos pesos em várias situações de importância prática. Estudos de simulação são feitos para aproximar esses pesos pelos coeficientes de pesos iguais.

O Capítulo 4 é dedicado ao modelo de regressão estrutural normal com erros nas variáveis. Estabelecemos as estatísticas do teste e demonstramos que a distribuição nula assintótica continua sendo uma mistura de distribuições do tipo qui-quadrado. Nesse caso, as probabilidades de nível dependem unicamente dos pesos para alguns casos particulares. Similarmente ao Capítulo 3, estudos de simulação são realizados.

Finalmente, resumimos nos apêndices, alguns resultados da literatura utilizados na tese. Também apresentamos as tabelas geradas por simulação e os

programas usados nos exemplos.

Capítulo 2

Hipóteses na forma de desigualdades não-lineares

2.1 Introdução

Existem inúmeros problemas, particularmente em Econometria, em que o espaço paramétrico é expresso na forma de restrições do tipo $h(\theta) \geq 0$, onde h é uma função contínua de R^p em R^k de posto k , sendo que o sinal $\geq$ significa que cada componente h_j de h é positiva ou nula. É de nosso interesse desenvolver procedimentos de estimação e testes que levem em consideração essas restrições. O estudo desses procedimentos são mais complexos do que no caso usual em que θ varia livremente no R^p . Propomos neste capítulo fazer uma apresentação dos resultados mais importantes relacionados com teste de hipóteses na forma de desigualdades não-lineares. Inicialmente, discutimos os resultados de Kodde e Palm (1986) que propõem um teste do tipo Wald ou distância D para testar hipóteses restritas por desigualdades não

-lineares. Devido às dificuldades para obter a distribuição nula assintótica da estatística do teste quando o número de desigualdades aumenta, Kodde e Palm apresentam limites para os valores críticos do teste sob a hipótese nula. Eles também provam que a distribuição nula assintótica de D é uma mistura de distribuições do tipo qui-quadrado. Em seguida, apresentamos os resultados de Gourieroux e Monfort (1995) que mostram formalmente a equivalência assintótica das estatísticas da razão de verossimilhança, de escore, de Wald e de Kuhn-Tucker para o caso do teste de hipóteses na forma de desigualdades não-lineares. Em particular, é mostrado que a distribuição nula assintótica da estatística de Kuhn-Tucker é uma mistura de distribuições do tipo qui-quadrado e que a mesma é assintoticamente equivalente à distância D de Kodde e Palm (1986). O caso de hipóteses nulas restritas esbarra no problema da determinação da distribuição nula menos favorável, que embora seja uma mistura de qui-quadrados, é de difícil determinação. Um lema proposto por Wolak (1991), em que esse problema é discutido, é apresentado na Seção 2.4. Finalmente, discutimos a computação dos coeficientes das distribuições qui-quadrado na Seção 2.5.

2.2 Distância de Kodde e Palm

Kodde e Palm (1986) propõem um teste do tipo Wald para testar hipóteses restritas na forma de desigualdades não-lineares. A fim de apresentar esses resultados, vamos supor que as restrições sob o vetor de parâmetros $\theta \in R^p$ são formuladas em termos de k funções contínuas independentes $h(\theta)$, as quais são diferenciáveis em alguma vizinhança aberta do verdadeiro parâmetro θ_0 .

A hipótese a ser testada é da forma

$$H_0 : h_1(\theta) = 0, h_2(\theta) = 0 \text{ contra } H_1 : h_1(\theta) \neq 0, h_2(\theta) \geq 0; \quad (2.1)$$

ou

$$H_0 : h_1(\theta) = 0, h_2(\theta) \geq 0 \text{ contra } H_1 : h_1(\theta) \neq 0, h_2(\theta) \not\geq 0. \quad (2.2)$$

As dimensões da partição de $h(\theta)$ em $h_1(\theta)$ e $h_2(\theta)$ são k_1 e $k - k_1$, respectivamente. Quando $k_1 = 0$, assumimos que sob H_1 existe pelo menos uma desigualdade estrita em (2.1). Sob H_1 em (2.2), os parâmetros são completamente não restritos. Para apresentar o teste de Wald para restrições por igualdades ou desigualdades, assumimos que θ pode ser consistentemente estimado por $\hat{\theta}$ tal que a distribuição assintótica é dada por

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_0), \quad (2.3)$$

onde I_0 pode ser consistentemente estimada por $\hat{I}_0$.

Transformamos as funções dos parâmetros $h(\theta)$ em um novo vetor de parâmetros $\gamma = (\gamma_1^t, \gamma_2^t)^t$ e $\hat{\gamma} = (\hat{\gamma}_1^t, \hat{\gamma}_2^t)^t$, onde

$$\gamma_i = \sqrt{n}h_i(\theta) \text{ e } \hat{\gamma}_i = \sqrt{n}h_i(\hat{\theta}). \quad (2.4)$$

A matriz de covariância assintótica de $\hat{\gamma}$ é portanto dada por

$$\Sigma = \Pi I_0^{-1} \Pi^t, \quad (2.5)$$

onde o argumento θ tem sido omitido para efeito de simplicidade e $\Pi = \partial h / \partial \theta^t$ é a matriz $k \times p$ das derivadas parciais de $h(\theta)$ com relação aos elementos de θ . A matriz de covariância Σ pode ser estimada por $\hat{\Sigma}$, com a expressão (2.5) sendo avaliada em $\hat{\theta}$.

Sejam R_0 e R_1 os espaços possíveis para γ sob as hipóteses nula e alternativa, respectivamente. Sob as hipóteses (2.1) e (2.2) os espaços prováveis são convexos em γ . Definimos a função distância na métrica Σ , com relação ao vetor x , por

$$\|x\|^2 = x^t \Sigma^{-1} x. \quad (2.6)$$

Usualmente a matriz Σ não é conhecida e deve ser substituída por uma estimativa consistente $\hat{\Sigma}$. Sejam $\hat{\gamma}^0$ e $\hat{\gamma}^c$ os estimadores de mínima distância os quais satisfazem as restrições sob H_0 e H_1 , respectivamente. Definimos

$$D_0 = \|\hat{\gamma} - \hat{\gamma}^0\|^2 = \min_{\gamma \in R_0} \|\hat{\gamma} - \gamma\|^2 \quad (2.7)$$

e

$$D_1 = \|\hat{\gamma} - \hat{\gamma}^c\|^2 = \min_{\gamma \in R_1} \|\hat{\gamma} - \gamma\|^2, \quad (2.8)$$

onde D_i , $i = 0, 1$, é a distância mínima do estimador irrestrito (isto é $\hat{\gamma}$) ao ponto mais próximo sob H_i , $i = 0, 1$. De fato, $\hat{\gamma}^0$ e $\hat{\gamma}^c$ são as projeções ortogonais de $\hat{\gamma}$ sobre R_0 e R_1 , respectivamente. Como R_0 e R_1 são convexos, $\hat{\gamma}^0$ e $\hat{\gamma}^c$ são determinados de forma única. Definimos o teste da distância ou teste do tipo Wald como

$$D = D_0 - D_1. \quad (2.9)$$

Usando propriedades de projeções ortogonais, obtemos uma formulação alternativa para esse teste,

$$D = \|\hat{\gamma}^c\|^2 - \|\hat{\gamma}^0\|^2. \quad (2.10)$$

Como $R_0 \subset R_1$, D será sempre não-negativo. Se D exceder o valor crítico para um nível de significância pré-estabelecido, rejeitamos a hipótese nula.

2.2.1 Distribuição nula assintótica de D

A seguir discutimos as hipóteses compostas (2.1) e (2.2) e apresentamos a distribuição assintótica de D sob H_0 .

Consideramos inicialmente (2.1). Quando $H_0 : \gamma = (\gamma_1^t, \gamma_2^t)^t = 0$ e $H_1 : \gamma_1 \neq 0, \gamma_2 \geq 0$, temos $\hat{\gamma}^0 = 0$ e $\tilde{\gamma}^c$ é tal que

$$\tilde{\gamma}_1^c = \hat{\gamma}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\tilde{\gamma}_2^c - \hat{\gamma}_2), \quad (2.11)$$

onde $\tilde{\gamma}_2^c$ é solução do problema

$$\min_{\gamma_2 \geq 0} (\hat{\gamma}_2 - \gamma_2)^t \Sigma_{22}^{-1} (\hat{\gamma}_2 - \gamma_2). \quad (2.12)$$

A partição de Σ está em concordância com γ . O teste do tipo Wald é dado por

$$\begin{aligned} D &= \|\tilde{\gamma}^c\|^2 \\ &= (\hat{\gamma}_1 - \Sigma_{12}\Sigma_{22}^{-1}\hat{\gamma}_2)^t \Sigma_{1.2}^{-1} (\hat{\gamma}_1 - \Sigma_{12}\Sigma_{22}^{-1}\hat{\gamma}_2) + (\hat{\gamma}_2^t \Sigma_{22}^{-1} \tilde{\gamma}_2^c), \end{aligned} \quad (2.13)$$

onde $\Sigma_{1.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Os dois termos do lado direito de (2.13) são assintoticamente independentes. O primeiro tem distribuição $\chi_{k_1}^2$. A distribuição do segundo é uma mistura de $q = k - k_1$ distribuições do tipo χ^2 , tal que a distribuição para grandes amostras de (2.13) pode ser escrita como

$$P_r(D \geq c | \Sigma) = \sum_{\ell=0}^q Q(q, \ell; \Sigma_{22}) P_r(\chi_{\ell+k_1}^2 \geq c), \quad (2.14)$$

onde $c \geq 0$ e $Q(q, \ell; \Sigma_{22})$ denota a probabilidade que exatamente ℓ componentes da variável aleatória $\tilde{\gamma}_2^c$ sejam positivos. O resultado (2.14) tem sido obtido por Gourieroux, Holly e Monfort (1982), Kudô (1963) e Perlman (1969) no

caso de estimação por máxima verossimilhança e $k_1 = 0$. Para $k_1 = 0$, $\chi_{k_1}^2$ é a distribuição degenerada na origem.

Para o caso (2.2), com $H_0 : \gamma_1 = 0, \gamma_2 \geq 0$ e $H_1 : \gamma_1 \neq 0, \gamma_2 \not\geq 0$, temos $\hat{\gamma}^c = \hat{\gamma}$, $\hat{\gamma}_1^0 = 0$ e $\hat{\gamma}_2^0$ é a solução de

$$\min_{\gamma_2 \geq 0} (\hat{\gamma}_2 - \gamma_2 - \Sigma_{21} \Sigma_{11}^{-1} \hat{\gamma}_1)^t \Sigma_{2.1}^{-1} (\hat{\gamma}_2 - \gamma_2 - \Sigma_{21} \Sigma_{11}^{-1} \hat{\gamma}_1), \quad (2.15)$$

onde $\Sigma_{2.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{21}$. O teste do tipo Wald fica dado por

$$\begin{aligned} D &= \| \hat{\gamma} - \hat{\gamma}^0 \|^2 \\ &= \hat{\gamma}_1^t \Sigma_{11}^{-1} \hat{\gamma}_1 + (\hat{\gamma}_2 - \hat{\gamma}_2^0 - \Sigma_{21} \Sigma_{11}^{-1} \hat{\gamma}_1)^t \Sigma_{2.1}^{-1} (\hat{\gamma}_2 - \hat{\gamma}_2^0 - \Sigma_{21} \Sigma_{11}^{-1} \hat{\gamma}_1). \end{aligned} \quad (2.16)$$

Quando Σ não depende de θ , o máximo sob a hipótese nula da distribuição para grandes amostras (vide Perlman 1969, teorema 8.3; Nuesch 1966), é igual a

$$\sup_{\gamma_2 \geq 0} P_r(D \geq c | \Sigma) = \sum_{\ell=0}^q Q(q, \ell; \Sigma_{2.1}) P_r\{\chi_{k-\ell}^2 \geq c\}, \quad (2.17)$$

onde o peso $Q(q, \ell; \Sigma_{2.1})$ denota a probabilidade que ℓ componentes da variável aleatória $\hat{\gamma}_2^0$ sejam positivos. A matriz $\Sigma_{2.1}$ é a matriz de covariância condicional de $\hat{\gamma}_2$ dado $\hat{\gamma}_1$. Quando a matriz Σ depende, sob H_0 , dos parâmetros, não necessariamente a situação menos favorável é alcançada com todas as desigualdades na forma de igualdades, isto é quando $h(\theta) = 0$. Esse problema será discutido na Seção 2.4. Os coeficientes $Q(q, \ell; \Sigma)$'s, também chamados de probabilidades de nível, serão apresentados analiticamente na Seção 2.5.

A principal dificuldade quando testamos restrições por desigualdades consiste no cálculo dos coeficientes em (2.17) a medida que o número de restrições por desigualdades aumenta. Porém, do teorema (6.2) de Perlman (1969) podemos obter limites superior e inferior para os valores críticos correspondentes a uma

determinada escolha do nível de significância α . Esses limites são obtidos resolvendo-se as equações:

$$\alpha = \inf_{\theta_0 \in H_0} P_r(D \geq c | \Sigma) = \frac{1}{2} P_r(\chi_{k_1}^2 \geq c) + \frac{1}{2} P_r(\chi_{k_1+1}^2 \geq c) \quad (2.18)$$

$$\alpha = \sup_{\theta_0 \in H_0} P_r(D \geq c | \Sigma) = \frac{1}{2} P_r(\chi_{k-1}^2 \geq c) + \frac{1}{2} P_r(\chi_k^2 \geq c),$$

respectivamente, para c .

2.3 Condições de regularidade de Gourieroux e Monfort

Gourieroux e Monfort (1995, Capítulo 21) tratam de maneira formal o problema de testar hipóteses na forma de desigualdades não-lineares. Apresentamos a seguir as condições de regularidade.

- Seja $L(\theta)$ uma função de verossimilhança e denotaremos por θ_0 o verdadeiro valor do parâmetro θ . Suponhamos que

$$\frac{1}{\sqrt{n}} U(\theta_0) \xrightarrow{d} N(0, I_0) \text{ e } -\frac{1}{n} K(\theta_0) \xrightarrow{q.c.} J_0, \quad (2.19)$$

onde $U(\theta) = \partial L / \partial \theta$ e $K(\theta) = \partial^2 L / \partial \theta \partial \theta^t$. Para simplificar, vamos supor que as matrizes I_0 e J_0 são iguais.

- θ_0 satisfaz k restrições por desigualdades escritas em forma explícita

$$h(\theta) \geq 0 \Leftrightarrow h_j(\theta) \geq 0, \quad j = 1, 2, \dots, k, \quad (2.20)$$

onde h é uma função de R^p em R^k , $k \leq p$, com o Jacobiano de posto k .

- O estimador não restrito de θ , denotado por $\hat{\theta}$, existe e é solução do problema de maximizar $L(\theta)$, $\theta \in \Theta$. Quando θ_0 pertence ao interior de Θ , o estimador $\hat{\theta}$ é tal que

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, J_0^{-1}).$$

Finalmente, podemos definir outros estimadores restritos. Se A é um subconjunto de índices de $\{1, \dots, k\}$, podemos considerar a solução do problema $\max_{\{\theta: h_j(\theta)=0\}} L(\theta)$, para todo $j \in A$. O estimador restrito pela igualdade é denotado $\hat{\theta}^0$.

2.3.1 Estimador restrito por desigualdades

Definição 2.1: o estimador $\tilde{\theta}^c$, restrito pela desigualdade é a solução do problema de maximizar $L(\theta)$, $\theta \in \Theta$, sujeito às restrições $h_1(\theta) \geq 0, \dots, h_k(\theta) \geq 0$. Esse estimador admite formas diferentes conforme as restrições sejam ou não satisfeitas.

Pode-se também resolver, quando $L(\theta)$ é uma função côncava, o problema de otimização introduzindo multiplicadores, nesse caso, chamados de Kuhn-Tucker (KT). Precisamos de k multiplicadores $\tilde{\lambda}_j$ associados às diversas restrições $h_j(\theta) \geq 0$, $j = 1, \dots, k$. O Lagrangiano fica dado por

$$l_n = L(\theta) + \sum_{j=1}^k \tilde{\lambda}_j h_j(\theta),$$

e as condições de primeira ordem são :

- $\frac{\partial l_n}{\partial \theta}(\tilde{\theta}^c, \tilde{\lambda}) = 0 \Leftrightarrow U'(\tilde{\theta}^c) + \Pi'(\tilde{\theta}^c)\tilde{\lambda} = 0$, onde $\lambda = (\lambda_1, \dots, \lambda_k)'$.

- Condição do sinal: para todo $j = 1, 2, \dots, k$, $h_j(\tilde{\theta}^c) \geq 0$, $\tilde{\lambda}_j \geq 0$.
- Condição de exclusão: para todo $j = 1, 2, \dots, k$, $\tilde{\lambda}_j h_j(\tilde{\theta}^c) = 0$.

Definição 2.2 : dado que o Jacobiano de h é de posto k , podemos mostrar facilmente a relação entre os multiplicadores de Kuhn-Tucker e a função escore,

$$\tilde{\lambda} = -[\Pi(\tilde{\theta}^c)\Sigma(\tilde{\theta}^c)\Pi'(\tilde{\theta}^c)]^{-1}\Pi(\tilde{\theta}^c)\Sigma(\tilde{\theta}^c)U'(\tilde{\theta}^c), \quad (2.21)$$

onde Σ é positiva definida.

2.3.2 Convergência

As propriedades de convergência de um estimador obtido pela maximização da função L não dependem da ausência ou presença de restrições; daí segue a convergência de $\tilde{\theta}^c$ para θ_0 . Também podemos concluir que

$$\frac{1}{n}\tilde{\lambda} = -[\Pi(\theta_0)\Sigma(\theta_0)\Pi'(\theta_0)]^{-1}\Pi(\theta_0)\Sigma(\theta_0)\frac{1}{n}U'(\theta_0) \xrightarrow{p} 0,$$

pois a condição usual de identificação assintótica é definida para todo $\epsilon \geq 0$, como

$$\lim_{n \rightarrow \infty} P(|\frac{1}{n}U'(\theta_0)| \geq \epsilon) = 0,$$

portanto $\frac{1}{n}\tilde{\lambda} \xrightarrow{p} 0$.

Por outro lado, o estudo das distribuições assintóticas de $\sqrt{n}(\tilde{\theta}^c - \theta_0)$ e de $\frac{1}{\sqrt{n}}\tilde{\lambda}$ é muito mais difícil. Se θ_0 pertence ao interior do domínio, isto é, se $h_j(\theta_0) > 0$, $j = 1, \dots, k$, $\tilde{\theta}^c$ é assintoticamente equivalente ao estimador não restrito $\hat{\theta}$ e é desse fato, assintoticamente normalmente distribuído. Ao

contrário, quando θ_0 se aproxima da fronteira do domínio, a distribuição limite toma uma forma muito complexa. A seguir determinamos as distribuições das funções de $\tilde{\theta}^c$, quando θ_0 satisfaz $h_j(\theta_0) = 0$, $j = 1, \dots, k$.

2.3.3 Teste da hipótese $H_0 : h(\theta) = 0$ contra $H_1 : h(\theta) \geq 0$, $h(\theta) \neq 0$ (caso I)

Suponhamos que a função $L(\theta)$ satisfaz as condições de regularidade definidas em (2.19) e que as duas matrizes I_0 e J_0 são iguais. Conforme foi definido acima, denotaremos por $\tilde{\theta}^c$ o estimador restrito pela desigualdade e seja $\hat{\theta}^0$ o estimador restrito pela igualdade. Assim, $\tilde{\theta}^c$ é solução do problema

$$\max_{\{\theta: h(\theta) \geq 0\}} L(\theta). \quad (2.22)$$

$\tilde{\lambda}$ designará o vetor de multiplicadores de KT associado a essas restrições e $U(\tilde{\theta}^c)$ denotará o vetor de escores avaliado em $\tilde{\theta}^c$. Similarmente, $\hat{\theta}^0$ é solução do problema

$$\max_{\{\theta: h(\theta) = 0\}} L(\theta). \quad (2.23)$$

Da mesma forma, $\hat{\lambda}^0$ denotará o vetor de multiplicadores de Lagrange associado às restrições e $U(\hat{\theta}^0)$ será o vetor de escores avaliado em $\hat{\theta}^0$.

2.3.4 Equivalência assintótica das estatísticas do teste

Definição 2.3: uma estatística do tipo da razão de verossimilhança pode ser definida como

$$LR = 2[L(\tilde{\theta}^c) - L(\hat{\theta}^0)]$$

$$= 2\left[\max_{\{\theta: h(\theta) \geq 0\}} L(\theta) - \max_{\{\theta: h(\theta) = 0\}} L(\theta)\right].$$

Sob a hipótese nula $H_0 : h(\theta) = 0$, os dois estimadores $\tilde{\theta}^c$ e $\hat{\theta}^0$ convergem para θ_0 . Podemos então aplicar uma expansão (por exemplo em série de Taylor numa vizinhança de θ_0) na estatística LR para obter uma estatística equivalente

$$\begin{aligned} LR &= 2\left[\frac{1}{\sqrt{n}}U^t(\theta_0)\sqrt{n}(\tilde{\theta}^c - \theta_0) - \frac{1}{2}n(\tilde{\theta}^c - \theta_0)^t I_0(\tilde{\theta}^c - \theta_0) + o_p(\|\hat{\theta}_n^c - \theta_0\|^2)\right. \\ &\quad \left.- \left\{\frac{1}{\sqrt{n}}U^t(\theta_0)\sqrt{n}(\hat{\theta}^0 - \theta_0) - \frac{1}{2}n(\hat{\theta}^0 - \theta_0)^t I_0(\hat{\theta}^0 - \theta_0) + o_p(\|\hat{\theta}_n^c - \theta_0\|^2)\right\}\right] \\ &= 2\left[\frac{1}{\sqrt{n}}U^t(\theta_0)\sqrt{n}(\tilde{\theta}^c - \theta_0) - \frac{1}{\sqrt{n}}U^t(\theta_0)\sqrt{n}(\hat{\theta}^0 - \theta_0)\right. \\ &\quad \left.- \frac{1}{2}n(\tilde{\theta}^c - \theta_0)^t I_0(\tilde{\theta}^c - \theta_0) + \frac{1}{2}n(\hat{\theta}^0 - \theta_0)^t I_0(\hat{\theta}^0 - \theta_0)\right. \\ &\quad \left.+ o_p(\max(\|\hat{\theta}_n^c - \theta_0\|^2, \|\hat{\theta}_n^0 - \theta_0\|^2))\right]. \end{aligned}$$

Nessa expansão admitimos implicitamente que L é definida numa vizinhança de θ_0 . Essa suposição é mantida ao longo deste capítulo.

Expandindo em série de Taylor a condição de primeira ordem sobre o lagrangiano

$$U(\tilde{\theta}^c) + \Pi^t(\tilde{\theta}^c)\tilde{\lambda} = 0,$$

obtemos

$$\frac{1}{\sqrt{n}}U(\theta_0) - I_0\sqrt{n}(\tilde{\theta}^c - \theta_0) + \Pi^t(\theta_0)\frac{\tilde{\lambda}}{\sqrt{n}} + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) = 0. \quad (2.24)$$

Da mesma forma obtemos

$$\frac{1}{\sqrt{n}}U(\theta_0) - I_0\sqrt{n}(\hat{\theta}^0 - \theta_0) + \Pi^t(\theta_0)\frac{\hat{\lambda}^0}{\sqrt{n}} + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) = 0. \quad (2.25)$$

Em (2.24) e (2.25) obtemos os valores de $\frac{1}{\sqrt{n}}U(\theta_0)$ para serem substituídos na expansão de LR ,

$$\begin{aligned}
LR &= 2\{I_0\sqrt{n}(\tilde{\theta}^c - \theta_0) - \Pi'(\theta_0)\frac{\tilde{\lambda}}{\sqrt{n}} - o_p(\|\hat{\theta}_n^c - \theta_0\|^2)\}\sqrt{n}(\tilde{\theta}^c - \theta_0) \\
&\quad - \{I_0\sqrt{n}(\hat{\theta}^0 - \theta_0) - \Pi'(\theta_0)\frac{\hat{\lambda}^0}{\sqrt{n}} - o_p(\|\tilde{\theta}_n^c - \theta_0\|^2)\}\sqrt{n}(\hat{\theta}^0 - \theta_0) \\
&\quad - \frac{1}{2}n(\tilde{\theta}^c - \theta_0)'I_0(\tilde{\theta}^c - \theta_0) + \frac{1}{2}n(\hat{\theta}^0 - \theta_0)'I_0(\hat{\theta}^0 - \theta_0) \\
&\quad + o_p(\max(\|\hat{\theta}_n^c - \theta_0\|^2, \|\tilde{\theta}_n^c - \theta_0\|^2)) \\
&= n(\tilde{\theta}^c - \theta_0)'I_0(\tilde{\theta}^c - \theta_0) - n(\hat{\theta}^0 - \theta_0)'I_0(\hat{\theta}^0 - \theta_0) \\
&\quad - 2\sqrt{n}(\tilde{\theta}^c - \theta_0)'\Pi'(\theta_0)\frac{\tilde{\lambda}}{\sqrt{n}} + 2\sqrt{n}(\hat{\theta}^0 - \theta_0)'\Pi'(\theta_0)\frac{\hat{\lambda}^0}{\sqrt{n}} \\
&\quad + \sqrt{n}(\hat{\theta}^0 - \theta_0)'o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) - \sqrt{n}(\tilde{\theta}^c - \theta_0)'o_p(\|\hat{\theta}_n^c - \theta_0\|^2) \\
&\quad + o_p(\max(\|\hat{\theta}_n^c - \theta_0\|^2, \|\tilde{\theta}_n^c - \theta_0\|^2)).
\end{aligned}$$

Agora vamos expandir a condição de exclusão $h_j(\tilde{\theta}^c)\tilde{\lambda}_j = 0$, para todo $j = 1, \dots, k$. Como $h(\theta_0) = 0$, obtemos

$$\begin{aligned}
h_j(\tilde{\theta}^c)\tilde{\lambda}_j &= h_j(\theta_0)\tilde{\lambda}_j + \Pi'(\theta_0)\sqrt{n}(\tilde{\theta}^c - \theta_0)\frac{\tilde{\lambda}_j}{\sqrt{n}} + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) \\
&= \Pi'(\theta_0)\sqrt{n}(\tilde{\theta}^c - \theta_0)\frac{\tilde{\lambda}_j}{\sqrt{n}} + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) = 0.
\end{aligned} \tag{2.26}$$

Similarmente, obtemos

$$h_j(\hat{\theta}^0)\hat{\lambda}_j^0 = \Pi'(\theta_0)\sqrt{n}(\hat{\theta}^0 - \theta_0)\frac{\hat{\lambda}_j^0}{\sqrt{n}} + o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) = 0, \tag{2.27}$$

pois $h(\hat{\theta}_0) = h(\theta_0) = 0$. Substituindo na última expansão de LR , obtemos

$$\begin{aligned}
LR &= n(\tilde{\theta}^c - \theta_0)'I_0(\tilde{\theta}^c - \theta_0) - n(\hat{\theta}^0 - \theta_0)'I_0(\hat{\theta}^0 - \theta_0) \\
&\quad + \sqrt{n}(\hat{\theta}^0 - \theta_0)'o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) - \sqrt{n}(\tilde{\theta}^c - \theta_0)'o_p(\|\hat{\theta}_n^c - \theta_0\|^2).
\end{aligned}$$

Aplicando o teorema de Slutsky aos dois últimos termos da expansão anterior podemos concluir que

$$LR \sim n(\tilde{\theta}^c - \theta_0)' I_0(\tilde{\theta}^c - \theta_0) - n(\hat{\theta}^0 - \theta_0)' I_0(\hat{\theta}^0 - \theta_0), \quad (2.28)$$

onde $\sim$ significa assintoticamente equivalente.

Definição 2.4: uma estatística do tipo Wald é definida como

$$W_H = n(\tilde{\theta}^c - \hat{\theta}^0)' \hat{I}(\tilde{\theta}^c - \hat{\theta}^0), \quad (2.29)$$

onde $\hat{I}$ é um estimador consistente de I_0 . A idéia dessa estatística é comparar os dois estimadores $\tilde{\theta}^c$ e $\hat{\theta}^0$. Notamos que W_H está diretamente ligada com a estatística LR anterior.

Teorema 2.1: sob H_0 , as estatísticas de W_H e LR são assintoticamente equivalentes.

Demonstração : seja

$$o_p[g(\theta_0, \tilde{\theta}^c, \hat{\theta}^0)] = \sqrt{n}(\hat{\theta}^0 - \theta_0)' o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) - \sqrt{n}(\tilde{\theta}^c - \theta_0)' o_p(\|\hat{\theta}_n^c - \theta_0\|^2),$$

então

$$\begin{aligned} LR &= n(\tilde{\theta}^c - \theta_0)' I_0(\tilde{\theta}^c - \theta_0) - n(\hat{\theta}^0 - \theta_0)' I_0(\hat{\theta}^0 - \theta_0) + o_p[g(\theta_0, \tilde{\theta}^c, \hat{\theta}^0)] \\ &= n(\tilde{\theta}^c - \hat{\theta}^0 + \hat{\theta}^0 - \theta_0)' I_0(\tilde{\theta}^c - \hat{\theta}^0 + \hat{\theta}^0 - \theta_0) - n(\hat{\theta}^0 - \theta_0)' I_0(\hat{\theta}^0 - \theta_0) \\ &\quad + o_p[g(\theta_0, \tilde{\theta}^c, \hat{\theta}^0)] \\ &= n(\tilde{\theta}^c - \hat{\theta}^0)' I_0(\tilde{\theta}^c - \hat{\theta}^0) + 2n(\tilde{\theta}^c - \hat{\theta}^0)' I_0(\hat{\theta}_n^c - \theta_0) \\ &\quad + o_p[g(\theta_0, \tilde{\theta}^c, \hat{\theta}^0)]. \end{aligned} \quad (2.30)$$

Agora, fazendo a diferença entre (2.25) e (2.24) obtemos

$$I_0\sqrt{n}(\tilde{\theta}^c - \hat{\theta}^0) = \Pi'(\theta_0) \frac{\tilde{\lambda} - \hat{\lambda}^0}{\sqrt{n}} + o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) - o_p(\|\hat{\theta}_n^c - \theta_0\|^2). \quad (2.31)$$

Substituindo (2.31) em (2.30) temos que

$$\begin{aligned} LR = & n(\tilde{\theta}^c - \hat{\theta}^0)' I_0(\tilde{\theta}^c - \hat{\theta}^0) + 2\Pi'(\theta_0) \frac{(\tilde{\lambda} - \hat{\lambda}^0)}{\sqrt{n}} \sqrt{n}(\hat{\theta}^0 - \theta_0) \\ & + \sqrt{n}(\hat{\theta}^0 - \theta_0) o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) - \sqrt{n}(\hat{\theta}^0 - \theta_0) \\ & + o_p[g(\theta_0, \tilde{\theta}^c, \hat{\theta}^0)]. \end{aligned}$$

De (2.26) e (2.27) segue que $\Pi'(\theta_0) \frac{(\tilde{\lambda} - \hat{\lambda}^0)}{\sqrt{n}} \sqrt{n}(\hat{\theta}^0 - \theta_0) \xrightarrow{p} 0$ e aplicando o teorema de Slutsky vemos que os três últimos termos na expansão de LR vão em probabilidade para zero. Portanto

$$LR \sim n(\tilde{\theta}^c - \hat{\theta}^0)' I_0(\tilde{\theta}^c - \hat{\theta}^0).$$

Definição 2.5: a estatística de KT compara os multiplicadores $\tilde{\lambda}$ e $\hat{\lambda}^0$. Essa estatística é definida por

$$KT = \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)' \hat{\Pi} \hat{I}^{-1} \hat{\Pi}' (\tilde{\lambda} - \hat{\lambda}^0), \quad (2.32)$$

onde, sob a hipótese nula, $\hat{I}$ e $\hat{\Pi}$ são estimadores consistentes de I_0 e $\Pi(\theta_0)$, respectivamente.

Teorema 2.2: sob H_0 , a estatística KT é assintoticamente equivalente a LR e W_H .

Demonstração : a prova desse teorema segue diretamente de (2.31).

Definição 2.6: A estatística de Wald é construída diretamente a partir das restrições avaliadas sob a hipótese geral. Essa estatística é definida como

$$W = nh'(\tilde{\theta}^c)[\Pi(\tilde{\theta}^c)\hat{I}^{-1}\Pi'(\tilde{\theta}^c)]^{-1}h(\tilde{\theta}^c). \quad (2.33)$$

Teorema 2.3: sob a hipótese nula a estatística de Wald é assintoticamente equivalente a LR , W_H e KT . **Demonstração :**

$$\begin{aligned}
 h(\tilde{\theta}^c) &= (h(\tilde{\theta}^c) - h(\hat{\theta}^0)) \\
 &= h(\theta_0) + \Pi(\theta_0)(\tilde{\theta}^c - \theta_0) + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) \\
 &\quad - h(\theta_0) - \Pi(\theta_0)(\hat{\theta}^0 - \theta_0) - o_p(\|\tilde{\theta}_n^c - \theta_0\|^2) \\
 &= \Pi(\theta_0)(\tilde{\theta}^c - \hat{\theta}^0) + o_p(\|\hat{\theta}_n^c - \theta_0\|^2) - o_p(\|\tilde{\theta}_n^c - \theta_0\|^2). \quad (2.34)
 \end{aligned}$$

Agora substituindo (2.31) em (2.34), obtemos

$$\begin{aligned}
 h(\tilde{\theta}^c) &= \Pi(\theta_0) \frac{\hat{J}^{-1}}{\sqrt{n}} \Pi'(\theta_0) \frac{\tilde{\lambda} - \hat{\lambda}^0}{\sqrt{n}} \\
 &\quad + [\Pi(\theta_0) \frac{\hat{J}^{-1}}{\sqrt{n}} - 1] o_p(\|\hat{\theta}_n^c - \theta_0\|^2) - [\Pi(\theta_0) \frac{\hat{J}^{-1}}{\sqrt{n}} - 1] o_p(\|\hat{\theta}_n^c - \theta_0\|^2).
 \end{aligned}$$

Daí segue que

$$\sqrt{n}h(\tilde{\theta}^c) \sim \Pi(\theta_0) \hat{J}^{-1} \Pi'(\theta_0) \frac{\tilde{\lambda} - \hat{\lambda}^0}{\sqrt{n}}. \quad (2.35)$$

O resultado desejado é obtido ao substituir (2.35) em (2.33) e comparar com (2.32). Em particular se I_0 é identidade e se H_0 é da forma $H_0 : \theta = 0$, a estatística W reduz a

$$W = n \|\tilde{\theta}^c\|^2.$$

Definição 2.7: a estatística de escore é definida por

$$S = \frac{1}{n} [U(\hat{\theta}^0) - U(\tilde{\theta}^c)]' \hat{J}^{-1} \hat{\Pi}' [\Pi \hat{J}^{-1} \hat{\Pi}']^{-1} \Pi \hat{J}^{-1} [U(\hat{\theta}^0) - U(\tilde{\theta}^c)], \quad (2.36)$$

onde $\hat{J}$ e $\hat{\Pi}$ são avaliados em $\hat{\theta}^0$.

Teorema 2.4: sob H_0 , a estatística de escore é assintoticamente equivalente às estatísticas LR , W_H , KT e W .

Demonstração : da definição (2.2), com $\Sigma = I_0^{-1}$ temos :

$$\frac{1}{\sqrt{n}}\tilde{\lambda} = -[\Pi(\tilde{\theta}^c)I_0^{-1}\Pi'(\tilde{\theta}^c)]^{-1}\Pi(\tilde{\theta}^c)I_0^{-1}\frac{1}{\sqrt{n}}U(\tilde{\theta}^c),$$

então

$$\begin{aligned}\frac{1}{\sqrt{n}}\tilde{\lambda} &\sim -[\Pi(\theta_0)I_0^{-1}\Pi'(\theta_0)]^{-1}\Pi(\theta_0)I_0^{-1}\frac{1}{\sqrt{n}}U(\theta_0) \text{ e} \\ \frac{1}{\sqrt{n}}\hat{\lambda}^0 &\sim -[\Pi(\hat{\theta}^0)I_0^{-1}\Pi'(\hat{\theta}^0)]^{-1}\Pi(\hat{\theta}^0)I_0^{-1}\frac{1}{\sqrt{n}}U(\hat{\theta}^0),\end{aligned}$$

portanto

$$\begin{aligned}KT &= \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' \Pi(\theta_0) \hat{I}^{-1} \Pi'(\theta_0) \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0) \\ &\sim \frac{1}{\sqrt{n}}[U(\hat{\theta}^0) - U(\tilde{\theta}^c)]' I_0^{-1} \Pi'(\theta_0) [\Pi(\theta_0) I_0^{-1} \Pi'(\theta_0)]^{-1} \times \\ &\quad \Pi(\theta_0) I_0^{-1} \frac{1}{\sqrt{n}}[U(\hat{\theta}^0) - U(\tilde{\theta}^c)] \sim S.\end{aligned}$$

2.3.5 Distribuição nula assintótica das estatísticas

Vamos examinar a distribuição assintótica das estatísticas do teste sob a hipótese nula $H_0 : h(\theta) = 0$. Como elas são todas equivalentes é suficiente encontrar a distribuição de uma delas. Partindo da estatística dual e conceitos afins, com $I_0 = J_0$, obtemos

$$\Sigma_0^{-1} = V_{as}(\frac{1}{\sqrt{n}}\hat{\lambda}^0) = [\Pi(\theta_0)I^{-1}(\theta_0)\Pi'(\theta_0)]^{-1},$$

onde V_{as} denota variância assintótica. Dessa maneira procuramos a distribuição nula de

$$\Xi = \min_{\{\lambda: \lambda \geq 0\}} \left[\frac{1}{n} (\lambda - \hat{\lambda}^0)' \Sigma_0 (\lambda - \hat{\lambda}^0) \right].$$

Note que a estatística Ξ é assintoticamente equivalente à estatística KT dada em (2.32). Seja $\hat{\mu}^0 = \Sigma_0^{-\frac{1}{2}} \hat{\lambda}^0 / \sqrt{n}$, $\mu = \Sigma_0^{-\frac{1}{2}} \lambda / \sqrt{n}$ e determinemos a distribuição nula de

$$\Xi = \min_{\{\mu: \Sigma_0^{-\frac{1}{2}} \mu \geq 0\}} \|\mu - \hat{\mu}^0\|^2,$$

onde $\hat{\mu}^0$ segue assintoticamente distribuição normal padrão.

Lema 2.1: seja X um vetor aleatório de dimensão k , com distribuição normal de média zero e matriz de covariância identidade de ordem k , (I_k) e seja R uma matriz quadrada simétrica, inversível de posto k . A distribuição de

$$\Xi = \min_{\{x: Rx \geq 0\}} \|X - x\|^2$$

é uma mistura de distribuições do tipo qui-quadrado, isto é, $\Xi \sim \sum_0^k Q(\ell) \chi_\ell^2$, onde $Q(\ell)$, $\ell = 0, \dots, k$, é um conjunto de pesos tais que $Q(\ell) \geq 0$, $\sum_0^k Q(\ell) = 1$ e χ_0^2 denota a massa pontual em zero.

De modo a simplificar a demonstração não discutiremos aqui a forma dos pesos $Q(\ell)$ s, (esses pesos serão objeto de tratamento especial na Seção 2.5 e no Apêndice A). A demonstração será dada apenas para o caso $k = 2$.

Demonstração : as restrições $Rx \geq 0$ valem para todas as componentes de Rx . Como R é uma matriz simétrica, elas se escrevem: $R_1^t x \geq 0$, $R_2^t x \geq 0$, onde R_1 e R_2 são as colunas de R . O conjunto de pontos $\{x : R_1^t x \geq 0, R_2^t x \geq 0\}$ forma o cone convexo $\mathcal{C}_{12}$ em R^2 . Para obtê-lo é suficiente procurar dois vetores $\tilde{R}_1$ e $\tilde{R}_2$ tais que $\tilde{R}_1^t R_1 > 0$, $\tilde{R}_2^t R_2 > 0$, $\tilde{R}_1^t R_2 = 0$ e $\tilde{R}_2^t R_1 = 0$ (vide Figura 2.1).

O cone $\mathcal{C}_{12}$ é então dado por

$$\mathcal{C}_{12} = \{x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2, \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0\}.$$

A escolha dos vetores $\tilde{R}_1$ e $\tilde{R}_2$ não é única de tal modo que cada um deles pode ser definido de forma diferente bastando para isso ser multiplicado por um coeficiente estritamente positivo. Entre essas escolhas, pode-se optar por dois vetores coluna da inversa da matriz R : $\tilde{R} = (\tilde{R}_1, \tilde{R}_2) = R^{-1}$. Com efeito, $\tilde{R}_1^t R_1 = 1$, $\tilde{R}_2^t R_2 = 1$, $\tilde{R}_2^t R_1 = 0$ e $\tilde{R}_1^t R_2 = 0$.

Figura 2.1: Cone $\mathcal{C}_{12}$

Ficam também definidos os cones convexos

$$\mathcal{C}_1 = \{x : \tilde{R}_1^t x \geq 0, \tilde{R}_2^t x \leq 0\} = \{x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2, \lambda_1 \geq 0, \lambda_2 \leq 0\};$$

$$\mathcal{C}_2 = \{x : R_1^t x \leq 0, \tilde{R}_2^t x \geq 0\} = \{x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2, \lambda_1 \leq 0, \lambda_2 \geq 0\} \text{ e}$$

$$\mathcal{C}_\phi = \{x : \tilde{R}_1^t x \leq 0, \tilde{R}_2^t x \leq 0\} = \{x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2, \lambda_1 \leq 0, \lambda_2 \leq 0\}.$$

O valor ótimo de Ξ é dado por

$$\Xi = \|X - Proj_{\mathcal{C}_{12}} X\|^2 = \|Proj_{\mathcal{C}_\phi} X\|^2,$$

onde $Proj_{\mathcal{C}} X$ denota a projeção de X sobre o cone $\mathcal{C}$. A projeção toma diversas formas conforme a posição de X

$$\text{Se } X \in \mathcal{C}_{12} : \|Proj_{\mathcal{C}_\phi} X\|^2 = 0;$$

$$\text{Se } X \in \mathcal{C}_1 : \|Proj_{\mathcal{C}_\phi} X\|^2 = \|Proj_{R_2} X\|^2;$$

$$\text{Se } X \in \mathcal{C}_2 : \|Proj_{\mathcal{C}_\phi} X\|^2 = \|Proj_{R_1} X\|^2 \text{ e}$$

$$\text{Se } X \in \mathcal{C}_\phi : \|Proj_{\mathcal{C}_\phi} X\|^2 = \|X\|^2.$$

A projeção de X sobre o cone coincide com as projeções sobre um subespaço de dimensão zero se $X \in \mathcal{C}_{12}$; de dimensão 1 se $x \in \mathcal{C}_1 \cup \mathcal{C}_2$; de dimensão 2 se $X \in \mathcal{C}_\phi$ (vide Figura 2.2).

Se denotarmos os eventos $B = (\Xi \in A)$, $B_0 = (X \in \mathcal{C}_\phi)$, $B_1 = (X \in \mathcal{C}_1)$, $B_2 = (X \in \mathcal{C}_2)$, $B_3 = (X \in \mathcal{C}_{12})$, onde A é um evento qualquer, temos que

$$P_r(B) = \sum_{j=0}^3 P_r(B|B_j) P_r(B_j).$$

Como $X \sim N(0, I_2)$, temos independência entre $X_1^2 + X_2^2$ e X_2/X_1 . O evento B_0 é função unicamente da variável X_2/X_1 . Então segue que

$$P_r(B|B_0) = P_r(X_1^2 + X_2^2 \in A|B_0) = P_r(X_1^2 + X_2^2 \in A) = P_2(A),$$

onde $P_2(A)$ denota a distribuição de probabilidades de uma distribuição qui-quadrado com dois graus de liberdade.

Também segue que

$$P_r(B|B_3) = P_r(0 \in A|B_3) = P_0(A),$$

onde P_0 denota a massa pontual em zero. Como a distribuição normal padrão é invariante por transformações ortogonais, podemos supor que os cones $\mathcal{C}_1$ e $\mathcal{C}_2$ coincidem com o quadrante positivo. Devemos então calcular

$$P_r(X_1^2 \in A | X_1 \geq 0, X_2 \geq 0) = P_r(X_1^2 \in A | X_1 \geq 0) = P_r(X_1^2 \in A) = P_1(A),$$

uma vez que X_1 e X_2 são independentes e a distribuição de X_1 é simétrica. P_1 denota a distribuição de uma χ_1^2 . Finalmente, notamos que

$$Q(0) = P_r(B_3), \quad Q(1) = P_r(B_1) + P_r(B_2) \quad \text{e} \quad Q(2) = P_r(B_0),$$

daí segue que

$$\Xi \sim Q(0)\chi_0^2 + Q(1)\chi_1^2 + Q(2)\chi_2^2.$$

A demonstração geral seguirá de forma similar. Quando $k \geq 3$ precisaremos introduzir os vetores coluna R_j , $j = 1, \dots, k$, da matriz R e os vetores coluna $\tilde{R}_j$, $j = 1, \dots, k$, da matriz R^{-1} . Para todo subconjunto de índices $\alpha \in \{1, \dots, k\}$ definimos os cones

$$\mathcal{C}_A = \{x = \sum_{j \in A} \lambda_j \tilde{R}_j + \sum_{j \in \bar{A}} \lambda_j R_j, \quad \lambda_j \geq 0 \text{ se } j \in A \text{ e } \lambda_j \leq 0 \text{ se } j \in \bar{A}\},$$

e segue então que

$$\Xi \sim \sum_{\ell=0}^k Q(\ell) \chi_\ell^2,$$

onde

$$Q(\ell) = \sum_{\{A: \#(\bar{A})=\ell\}} P_r(X \in \mathcal{C}_A). \quad (2.37)$$

Figura 2.2: Projeções sobre o cone $\mathcal{C}_\phi$

Teorema 2.5: as estatísticas LR , W_H , W , KT e S seguem assintoticamente, sob a hipótese nula, uma mistura de distribuições do tipo qui-quadrado

$$\sum_{\ell=0}^k Q(\ell) \chi^2_\ell,$$

com os pesos calculados segundo (2.37) e com

$$R = [\Pi(\theta_0)I^{-1}(\theta_0)\Pi^t(\theta_0)]^{-\frac{1}{2}}.$$

Sob o ponto de vista prático, os pesos que dependem de θ_0 devem ser estimados. Os pesos podem ser completamente explicitados para o caso $k = 2$, uma vez que eles são calculados a partir das probabilidades $P_r(X \in \mathcal{C}_A)$. Como a distribuição de X é invariante por rotação, essa probabilidade é igual a $\alpha_A/2\pi$, onde α_A é o ângulo no vértice do cone $\mathcal{C}_A$. Segue imediatamente que $Q(1) = \frac{1}{2}$.

Por outro lado

$$\cos \alpha_{12} = \frac{\tilde{R}_1^t \tilde{R}_2}{\sqrt{\tilde{R}_1^t R_1} \sqrt{\tilde{R}_2^t R_2}}.$$

Como R^{-1} é matriz simétrica, essa razão é igual á $r_{12}/\sqrt{r_{11}}\sqrt{r_{22}}$ onde r_{ij} é o termo geral de $R^{-2} = \Pi(\theta_0)I^{-1}(\theta_0)\Pi^t(\theta_0)$. Portanto

$$\cos \alpha_{12} = \frac{\Pi_1(\theta_0)I^{-1}(\theta_0)\Pi_2^t(\theta_0)}{\sqrt{\Pi_1(\theta_0)I^{-1}(\theta_0)\Pi_1^t(\theta_0)}\sqrt{\Pi_2(\theta_0)I^{-1}(\theta_0)\Pi_2^t(\theta_0)}},$$

onde Π_j , $j = 1, 2$, é a matriz de primeiras derivadas de h_j , $j = 1, 2$, com relação a θ e $Q(0) = \alpha_{12}/2\pi$, $Q(2) = \frac{1}{2} - \alpha_{12}/2\pi$. A mistura de qui-quadrados tem a forma

$$Q(0)\chi_0^2 + \frac{1}{2}\chi_1^2 + \left\{\frac{1}{2} - Q(0)\right\}\chi_2^2,$$

onde $Q(0)$ se calcula como: $Q(0) = \frac{1}{2\pi} \cos^{-1}[\rho_{as}(h_1(\hat{\theta}), h_2(\hat{\theta}))]$, assim os pesos estão diretamente relacionados com o coeficiente de correlação linear assintótico entre as restrições estimadas. Podemos então escrever, seguindo a notação da Seção 2.2, que $Q(\ell) = Q(\ell, k; \Sigma_0)$, onde $\Sigma_0 = \Pi_0(\theta_0)I^{-1}(\theta_0)\Pi_0^t(\theta_0)$.

2.3.6 Teste da hipótese $H_0 : h(\theta) \geq 0$ contra $H_1 : \theta \in R^p$ (caso II)

Neste caso, a hipótese nula H_0 corresponde a restrições por desigualdades sobre os parâmetros, enquanto que sob a hipótese geral H_1 os parâmetros são irrestritos. Podemos construir diversas medidas da distância entre as duas hipóteses H_0 e H_1 . Denotamos por $\hat{\theta}$ e $\tilde{\theta}^c$ os estimadores não restrito e restrito pelas desigualdades dos parâmetros, respectivamente. Podemos, em particular, introduzir uma estatística baseada nos valores ótimos da função objetivo

$$LR = 2[L(\hat{\theta}) - L(\tilde{\theta}^c)],$$

ou uma estatística do tipo Wald

$$W_H = \min_{\{\theta: h(\theta) \geq 0\}} \left[n(\hat{\theta} - \theta)' I(\hat{\theta})(\hat{\theta} - \theta) \right].$$

De fato, destacamos que apesar disso, o problema do teste é muito diferente do problema tratado na Seção 2.3.3. A hipótese nula é agora composta de um interior não -vazio. Daí seguem as seguintes observações :

1. Fica difícil propor os testes baseados em idéias locais, em particular introduzir os testes do tipo multiplicadores de Lagrange ou do tipo escore, pois determinadas vizinhanças de pontos da hipótese nula estão inteiramente incluídos nessa hipótese.
2. Sob a hipótese nula, a distribuição assintótica de W_H (se ela existe) admite diversas formas dependendo do verdadeiro valor do parâmetro θ_0 , $h(\theta_0) \geq 0$. Se temos uma região crítica do tipo $[W_H \geq c]$, as

condições sobre o erro de tipo um α são

$$\sup_{\{\theta_0: h(\theta_0) \geq 0\}} P_{\theta_0}[W_H \geq c] \geq \alpha.$$

A seguir apresentamos um lema que discute o problema de definir a situação menos favorável no caso II.

2.4 Lema de Wolak

Para estabelecer as bases necessárias para os resultados de Wolak (1991), primeiro são revisados os fatos relevantes para o teste linear simples restrito por desigualdades. Esse teste toma a forma $H_0 : \mu \geq 0$ contra $H_1 : \mu \in R^p$ baseado em $\hat{\mu} \sim N(\mu, J_0)$, onde J_0 é uma matriz positiva definida e conhecida. Perlman (1969) apresenta um procedimento geral para analisar esse tipo de problema de teste. Seguindo suas argumentações, a estatística $\mathcal{W}$ é definida como a distância mínima do tipo qui-quadrado entre $\hat{\mu}$ e o conjunto ao qual μ pertence sob a hipótese nula. Nesse caso temos

$$\mathcal{W} = \min_{\{m: m \geq 0\}} [(\hat{\mu} - m)^t J_0^{-1} (\hat{\mu} - m)]. \quad (2.38)$$

Essa hipótese nula não especifica um único μ . Para obter um valor crítico exato de nível α para um teste de hipótese restrito por desigualdades lineares, o valor verdadeiro de μ menos favorável pode ser escolhido no conjunto determinado sob a hipótese nula. Do lema 8.2 de Perlman (1969) temos que para $c \geq 0$,

$$\sup_{\mu \geq 0} P_r[\mathcal{W} \geq c | \mu, J_0] = P_r[\mathcal{W} \geq c | 0, J_0], \quad (2.39)$$

onde $P_r[\mathcal{W} \geq c|\mu, J_0]$ é a probabilidade do evento $[\mathcal{W} \geq c]$ dado que $\hat{\mu}$ em (2.39) é $N(\mu, J_0)$. Portanto, o valor menos favorável de $\mu \geq 0$ o qual determina a distribuição nula de nível exato é $\mu = 0$. É crucial para a validade do lema 8.2 de Perlman (1969) a independência funcional entre J_0 e o parâmetro μ , de tal forma que a variação em μ não pode afetar os elementos de J_0 . Infelizmente com modelos não-lineares, a dependência funcional da matriz de covariância assintótica de funções do estimador restrito do verdadeiro vetor de parâmetros, é intrínseca ao problema.

Apresentamos agora o lema de Wolak (1991), que pode ser obtido do citado lema 8.2 e que segue rigorosamente o problema geral de testes com restrições não-lineares por desigualdades. Para facilidade de exposição, o lema é apresentado para o método de máxima verossimilhança, embora sob adequadas condições de regularidade o resultado vale para situações mais gerais. Seja $\theta \in \Theta$ o vetor paramétrico, $\Theta \subset R^p$ um espaço paramétrico compacto e $L(\theta)$ a função de verossimilhança baseada em n observações. Sob condições gerais de regularidade $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$ para todo $\theta_0 \in (\text{interior de } \Theta)$, onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ_0 e

$$I(\theta) = \lim_{n \rightarrow \infty} E_{\theta_0} \left[-\frac{1}{n} K(\theta) \right],$$

onde $I(\theta)$ é a matriz de informação de Fisher. Essa matriz pode ser consistentemente estimada por $\hat{J}(\theta) = -\frac{1}{n} K(\theta)$. Assumindo que θ_0 é o verdadeiro valor de θ , podemos mostrar que $\hat{J}(\hat{\theta})$ converge em probabilidade para $I(\theta_0)$, tal que $\hat{J}(\hat{\theta})^{-1}$ é um estimador consistente da matriz de covariância assintótica de $\sqrt{n}(\hat{\theta} - \theta_0)$.

Consideremos o teste restrito por desigualdades não-lineares

$$H_0 : h(\theta) \geq 0 \text{ contra } H_1 : \theta \in R^p, \quad (2.40)$$

onde $h(\theta) = (h_1(\theta), \dots, h_k(\theta))^t$ é uma função vetorial não linear p -dimensional de R^p em R^k , ($k \leq p$). Seja $\Pi(\theta)$ uma matriz $k \times p$ das derivadas parciais $\partial h_i(\theta)/\partial \theta_j$, $i = 1, \dots, k$ e $j = 1, \dots, p$, de $h(\theta)$ com relação aos elementos de θ . De (2.12) segue que uma das três formas para a estatística do teste, assintoticamente equivalentes, é dada por

$$W_n = \min_{\{t: t \geq 0\}} [n(h(\hat{\theta}) - t)^t \Sigma(\hat{\theta})^{-1} (h(\hat{\theta}) - t)],$$

onde $\Sigma(\hat{\theta}) = [\Pi(\hat{\theta})\hat{J}(\hat{\theta})^{-1}\Pi(\hat{\theta})^t]$ e $\Sigma(\theta_0) = [\Pi(\theta_0)I(\theta_0)^{-1}\Pi(\theta_0)^t]$.

A fim de enunciar o lema, cuja demonstração pode ser vista em Wolak (1991), definimos a seguinte notação : seja $C = \{\theta | h(\theta) \geq 0\}$, $C^i = \{\theta | h(\theta) > 0\}$ e $C^b = C - C^i$. Particionando C^b em dois conjuntos $A = \{\theta | \theta \in C^b \text{ e } h_j(\theta) = 0 \text{ para um único } j \in P\}$, onde $P = \{1, 2, \dots, k\}$ e $B = C^b - A$. Para todo $\theta_0 \in C^b$, seja $h_b(\theta)$ o vetor de elementos de $h(\theta)$ tal que $h_j(\theta_0) = 0$, $j \in P$, e $m \leq k$ denota a dimensão de $h_b(\theta)$. Seja $\Pi_b(\theta_0)$ a matriz de primeiras derivadas parciais de $h_b(\theta)$ com relação a θ .

Lema 2.2: para testar (2.40) temos os três resultados seguintes:

(i) Para todo $\theta_0 \in C^i$,

$$\lim_{n \rightarrow \infty} P_r[W_n = 0 | \theta_0] = 1.$$

(ii) Para qualquer $c \geq 0$ e $\theta_0 \in C^b$ tal que $\Pi_b(\theta_0)$ seja de posto completo

$$\lim_{n \rightarrow \infty} P_r[W_n \geq c | \theta_0] = \sum_{\ell=0}^m Q(m, m - \ell; \Sigma_b(\theta_0)) P_r\{\chi_\ell^2 \geq c\}, \quad (2.41)$$

onde $\Sigma_b(\theta_0) = [\Pi_b(\theta_0)I(\theta_0)^{-1}\Pi_b(\theta_0)']$ e os $Q(m, m - \ell; \Sigma_b(\theta_0))$'s são as funções peso que serão definidas na Seção 2.5.

(iii) Para todo $\theta_0 \in B$, $\theta_0^* \in A$ e fixado $c \geq 0$,

$$\lim_{n \rightarrow \infty} P_r[W_n \geq c|\theta_0] \geq \lim_{n \rightarrow \infty} P_r[W_n \geq c|\theta_0^*].$$

A parte (i) deste lema estabelece que todos os elementos em C^i , o interior do conjunto de restrições, pode não ser levado em consideração como o valor menos favorável, uma vez que $W_n \xrightarrow{p} 0$ para esses valores de θ_0 . A parte (ii) dá uma caracterização geral da distribuição assintótica de W_n para todos os valores de $\theta_0 \in C^b$. Em particular, mostra o impacto na distribuição de W_n de qualquer uma das desigualdades que não satisfazem a igualdade em θ_0 . Finalmente, a parte (iii) elimina de consideração determinados elementos de C^b como potenciais valores menos favoráveis para θ_0 . Juntos, esses três resultados mostram que o valor menos favorável de θ_0 é um elemento de B . Embora esse lema especifique que $B \subset C$ deva conter o valor menos favorável de θ_0 sob a hipótese nula, em geral, ele não produz a solução em θ_0 tal que

$$\sup_{\theta_0 \in C} \lim_{n \rightarrow \infty} P_r[W_n \geq c|\theta_0], \quad (2.42)$$

para um dado $c \geq 0$, o qual é o problema de otimização que deve ser resolvido para computar o $\theta_0 \in C$ menos favorável para o valor crítico c . Devemos notar o seguinte:

- (a) O valor de $\theta_0 \in B$ o qual maximiza (2.42) é o valor menos favorável de θ_0 que determina o tamanho assintoticamente exato do teste para o valor crítico c . Em vista da dependência funcional desse valor menos

favorável em c , o mesmo é denotado por $\theta_0^*(c)$. O valor crítico para um teste assintoticamente exato, é o c que resolve

$$\lim_{n \rightarrow \infty} P_r[W_n \geq c | \theta_0^*(c)] = \alpha.$$

- (b) O cálculo das funções pesos $Q(q, \ell; \Sigma)$, $\ell = 1, \dots, k$ tem sido de alguma forma resolvido e assim a distribuição de (2.41) pode ser calculada. Porém, o problema de determinar $\Sigma_b(\theta_0)$ para cada θ_0 ainda permanece. Como definido na parte (ii) do lema (2.2), $\Sigma_b(\theta_0)$ é uma função da matriz de informação, a qual não tem uma forma analítica para vários modelos não-lineares. Em consequência para esses modelos, a não existência de uma expressão analítica para a matriz $I(\theta)$ impede o cálculo do valor crítico assintoticamente exato.
- (c) Segundo Wolak (1991), há duas situações em que o valor único menos favorável de θ_0 existe para o teste de (2.40). Se a matriz $I(\theta_0)$ é diagonal para todo $\theta_0 \in B$ e as restrições por desigualdades tomam as formas de limites simples, o valor menos favorável é o único θ_0 o qual satisfaz todas as desigualdades com igualdade. Para o caso de um teste unidimensional restrito por desigualdade não-linear, todos os valores de θ_0 os quais satisfazem as restrições por desigualdade com igualdade produzem a mesma distribuição assintótica. Esse resultado é um teste não-linear em grandes amostras, analogo ao teste t usual unicaudal. No entanto, há situações em modelos não-lineares em que a distribuição nula assintótica menos favorável é a mesma para todo θ_0 tal que $h(\theta_0) = 0$. Paula e Sen (1995) mostram para modelos lineares generalizados de retas paralelas, sob certas condições experimentais comumente adotadas na prática, que a situação menos favorável ocorre quando $h(\theta_0) = 0$ e é a mesma para todo θ_0 nessas condições. Mostramos no Capítulo 3 que também sob

certas condições experimentais esse resultado vale para modelos não-gaussianos com distribuição do valor extremo (vide também Paula e Rojas, 1996).

Com a determinação do conjunto B , Wolak (1991) obtém os seguintes limites para o nível de significância α :

$$\alpha = \inf_{\{\theta_0 \in B\}} P_r(D \geq c | \Sigma(\theta_0)) = \frac{1}{2} P_r(\chi_1^2 \geq c)$$

e

$$\alpha = \sup_{\{\theta_0 \in B\}} P_r(D \geq c | \Sigma(\theta_0)) = \frac{1}{2} P_r(\chi_{k-1}^2 \geq c) + \frac{1}{2} P_r(\chi_k^2 \geq c),$$

onde D denota aqui a estatística $\mathcal{W}_n$ para grandes amostras. Note que esses limites coincidem, quando $k_1 = 0$, com os limites (2.18) obtidos por Kodde e Palm (1986).

2.4.1 Exemplo

Apresentamos um exemplo bidimensional para ilustrar o caso no qual podem ocorrer múltiplas distribuições nulas assintóticas para teste de hipóteses não-lineares restritas por desigualdades. Consideremos o problema de testar

$$H_0 : h(\theta) \geq 0 \quad \text{contra} \quad H_1 : \theta \in R^2 \quad (2.43)$$

baseado em $\gamma = \theta + \nu$ com $\nu \sim N(0, I_2)$, onde I_2 é a matriz identidade de ordem 2. Definimos os elementos de $h(\theta) \in R^2$ da seguinte forma: $h_1(\theta) = \theta_2 - \theta_1^2$ e $h_2(\theta) = \theta_1 - \theta_2$ para $\theta = (\theta_1, \theta_2)^t$. A matriz de primeiras derivadas de $h(\theta)$ toma a forma

$$\Pi(\theta) = \begin{bmatrix} -2\theta_1 & 1 \\ 1 & -1 \end{bmatrix}, \quad (2.44)$$

para todo $\theta \in R^2$. Dadas n observações de γ , o estimador de máxima verossimilhança de θ_0 é $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \gamma_i$, o qual tem distribuição $N(\theta_0, \frac{1}{n} I_2)$. Segue então que

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_2)$$

e pelo método delta temos que

$$\sqrt{n}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} N(0, J(\theta_0)),$$

onde $J(\theta_0) = \Pi(\theta_0)\Pi(\theta_0)^t$. A estatística para (2.43) é dada por

$$W_n = \min_{\{t: t \geq 0\}} [n(h(\hat{\theta}) - t)^t J(\hat{\theta})^{-1} (h(\hat{\theta}) - t)], \quad (2.45)$$

onde $J(\hat{\theta}) = \Pi(\hat{\theta})\Pi(\hat{\theta})^t$. Aplicando o lema 2.2 e como $h(\theta) \in R^2$, o valor menos favorável para θ_0 sob a hipótese nula deve pertencer ao conjunto $T = \{\theta | h(\theta) = 0\}$. Para esse problema existem dois $\theta_0 \in T$, isto é, $\theta_0^{00} = (0, 0)^t$ e $\theta_0^{11} = (1, 1)^t$. Como $\hat{\theta}$ converge em probabilidade para θ_0 , sabemos que $J(\hat{\theta})$ converge em probabilidade para $J(\theta_0)$ para todo θ_0 no espaço paramétrico, tal que temos:

$$J(\theta_0^{00}) = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \text{ e } J(\theta_0^{11}) = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}.$$

Como a distribuição assintótica de (2.45) depende de θ_0 através de $J(\theta_0)$, cada valor de θ_0 implica numa distribuição nula assintótica. A distribuição assintótica de W_n dado $\theta_0 \in T$ é dada por

$$\lim_{n \rightarrow \infty} P_r[W_n \geq c] = \sum_{\ell=0}^2 Q(2, 2 - \ell; J(\theta_0)) P_r\{\chi_\ell^2 \geq c\}. \quad (2.46)$$

A distribuição assintótica é funcionalmente dependente de θ_0 através dos pesos $Q(2, 2 - \ell; J(\theta_0))$ os quais são funções dos coeficientes de correlação obtidos a partir da matriz $J(\theta_0)$. Os pesos para $J(\theta_0^{00})$ e $J(\theta_0^{11})$ podem ser obtidos das

formas fechadas apresentadas em Wolak (1987) (vide Seção 2.5). Esses pesos nos possibilitam obter duas distribuições nulas assintóticas as quais levam a dois valores críticos diferentes para o teste. Wolak (1991) utilizando transformações afins, mostra que o valor θ_0^{00} leva à distribuição nula assintótica menos favorável. Isso pode ser verificado, alternativamente, calculando-se a probabilidade do lado direito de (2.46) para θ_0^{00} e θ_0^{11} e variando-se o valor crítico c .

2.5 Probabilidades de nível

2.5.1 Introdução

Seja X uma variável aleatória k -dimensional com vetor de médias μ , matriz de covariância Σ e função de densidade

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right\}, \quad x \in R^k. \quad (2.47)$$

Seja $A \subset R^k$ um conjunto Borel-mensurável. O problema a ser considerado é encontrar métodos para calcular ou aproximar a probabilidade

$$P_r(X \in A) = \int_A f(x; \mu, \Sigma) \partial x. \quad (2.48)$$

Sob o ponto de vista aplicado este é um problema importante, uma vez que essas probabilidades são frequentemente necessárias em estimação paramétrica, teste de hipóteses e outros problemas relativos à análise multivariada. Em particular, o problema de aproximar a probabilidade do quadrante para uma distribuição normal multivariada tem longa história e consiste em parte importante da estatística computacional em análise multivariada. O problema

geral está relacionado com métodos numéricos para aproximar

$$P_{\mu, \Sigma} \left[\bigcap_{i=1}^k \{X_i \leq a_i\} \right] \quad \text{ou} \quad P_{\mu, \Sigma} \left[\bigcap_{i=1}^k \{X_i \geq b_i\} \right], \quad (2.49)$$

onde $a_1, \dots, a_k, b_1, \dots, b_k$ são números reais arbitrários fixados e $X = (X_1, \dots, X_k)^t$ tem distribuição normal multivariada com média μ e matriz de covariância Σ . Como podemos escrever

$$P_{\mu, \Sigma} \left[\bigcap_{i=1}^k \{X_i \leq a_i\} \right] = P_{0, R} \left[\bigcap_{i=1}^k \left\{ Z_i \leq \frac{a_i - \mu_i}{\sigma_i} \right\} \right], \quad (2.50)$$

onde $Z_i = (X_i - \mu_i)/\sigma_i$ $i = 1, \dots, k$ e $R = \{\rho_{ij}\}$ é a correspondente matriz de correlações, sem perda de generalidade podemos assumir que o vetor de médias de X é zero e sua matriz de covariância R . Assim, o problema de interesse é calcular a probabilidade

$$g(\rho_{12}, \dots, \rho_{k-1, k}; a_1, \dots, a_k) = P_{0, R} \left[\bigcap_{i=1}^k \{X_i \leq a_i\} \right] \quad (2.51)$$

como uma função dos a_i 's e dos ρ_{ij} 's. Os principais resultados nesta área estão relacionados com expansões assintóticas para a função de densidade normal multivariada e portanto para a própria g . A seguir faremos uma apresentação dos principais métodos usados para o cálculo e aproximações de integrais da distribuição normal multivariada [vide Tong, 1990], assim como dos resultados e das aplicações dessas aproximações em teste de hipóteses.

2.5.2 Método da redução da dimensão

Um resultado significativo na redução da dimensão para calcular a integral da distribuição normal multivariada é devido a Plackett (1954). Ele provou a

identidade

$$\frac{\partial}{\partial \rho_{ij}} f(x; 0, \Sigma) = \frac{\partial^2}{\partial x_i \partial x_j} f(x; 0, \Sigma), \quad i \neq j, \quad (2.52)$$

onde $f(x; 0, \Sigma)$ é a função normal multivariada com vetor de médias zero e matriz de covariância Σ positiva definida. Como a função de distribuição é da forma

$$F(a; 0, \Sigma) = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_k} f(x; 0, \Sigma) \prod_{i=1}^n \partial x_i, \quad (2.53)$$

trocando diferenciação e integração, essa identidade leva a

$$\begin{aligned} \frac{\partial}{\partial \rho_{12}} F(a; 0, \Sigma) &= \int_{-\infty}^{a_3} \dots \int_{-\infty}^{a_k} f(a_1, a_2, x_3, \dots, x_k; 0, \Sigma) \prod_{i=1}^k \partial x_i \\ &= f_1(a_1, a_2; 0, \Sigma) \int_{-\infty}^{a_3} \dots \int_{-\infty}^{a_k} f_{2|1}(x_2; \mu_{2|1}, \Sigma_{22.1}) \prod_{i=3}^k \partial x_i, \end{aligned} \quad (2.54)$$

onde f_1 é a função de densidade marginal de $(X_1, X_2)^t$ avaliada em $(a_1, a_2)^t$ e $f_{2|1}$ é a função de densidade condicional de $(X_3, \dots, X_k)^t$ dado $(X_1, X_2)^t = (a_1, a_2)^t$. Usando (2.54), $F(a; 0, \Sigma)$ pode ser expressa como função de ρ_{12} na forma de uma integral simples cujo integrando envolve f_1 e uma integral das restantes $k-2$ variáveis $X_3, \dots, X_k$. Esse método tem sido utilizado por vários autores. Em particular, Bacon (1963) usou essa identidade para desenvolver métodos de redução da dimensão para calcular a probabilidade do quadrante positivo de variáveis normais igualmente correlacionadas.

Particularmente, Bacon considerou k variáveis aleatórias correlacionadas com distribuição normal de média 0 e variância 1. Logo, se $f_k(x_1, x_2, \dots, x_k)$ é a função de densidade, então a probabilidade de que todas as k variáveis aleatórias sejam positivas é dada por

$$P_k(c_{11}, c_{12}, \dots, c_{kk}) = \int_0^\infty \dots \int_0^\infty f_k(x_1, \dots, x_k) \partial x_1 \dots \partial x_k,$$

onde $c_{11}, \dots, c_{kk}$ denotam os elementos da inversa da matriz de covariância. Sejam ρ_{ij} os coeficientes de correlação. Aplicando a identidade (2.52), temos que

$$\begin{aligned} \frac{\partial P_k}{\partial \rho_{12}} &= \int_0^\infty \dots \int_0^\infty \left[\int_0^\infty \int_0^\infty \frac{\partial^2 f_k}{\partial x_1 \partial x_2} \partial x_1 \partial x_2 \right] \partial x_3 \dots \partial x_k \\ &= \frac{1}{2\pi(1 - \rho_{12}^2)^{\frac{1}{2}}} P_{k-2}(c_{33}, \dots, c_{kk}). \end{aligned} \quad (2.55)$$

Assim teremos $k(k-1)/2$ equações, uma para cada coeficiente de correlação. Quando os ρ 's são iguais, somando as $k(k-1)/2$ equações do tipo (2.55) obtemos

$$\frac{\partial P_k(\rho)}{\partial \rho} = \frac{k(k-1)}{4\pi(1 - \rho^2)^{\frac{1}{2}}} P_{k-2}\left(\frac{\rho}{1 + 2\rho}\right).$$

Como $P_k(0) = (\frac{1}{2})^k$, temos

$$P_k(\rho) = \left(\frac{1}{2}\right)^k + \frac{k(k-1)}{4\pi} \int_0^\rho P_{k-2}\left(\frac{\lambda}{1 + 2\lambda}\right) \frac{\partial \lambda}{(1 - \lambda^2)^{\frac{1}{2}}}. \quad (2.56)$$

Daí segue que

$$P_1 = \frac{1}{2}, \quad (2.57)$$

$$P_2 = \frac{1}{4} \left[1 + \frac{2}{\pi} \sin^{-1} \rho \right] \text{ e} \quad (2.58)$$

$$P_3 = \frac{1}{8} \left[1 + \frac{2}{\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{23} + \sin^{-1} \rho_{13}) \right]. \quad (2.59)$$

Combinando (2.56) e (2.58) obtemos

$$P_4(\rho) = \frac{1}{16} \left[1 + \frac{12}{\pi} \sin^{-1} \rho + \frac{24}{\pi^2} \int_0^\rho \sin^{-1} \left(\frac{\lambda}{1 + 2\lambda} \right) \frac{\partial \lambda}{(1 - \lambda^2)^{\frac{1}{2}}} \right]. \quad (2.60)$$

Combinando (2.56) e (2.59) obtemos

$$P_5(\rho) = \frac{1}{32} \left[1 + \frac{20}{\pi} \sin^{-1} \rho + \frac{120}{\pi^2} \int_0^\rho \sin^{-1} \left(\frac{\lambda}{1 + 2\lambda} \right) \frac{\partial \lambda}{(1 - \lambda^2)^{\frac{1}{2}}} \right]. \quad (2.61)$$

Continuando o processo de combinar (2.56) com as equações seguintes obtemos a equação geral

$$\begin{aligned}
 P_k(\rho) = & \left(\frac{1}{2} \right)^k \left[1 + \frac{k(k-1)}{\pi} \sin^{-1} \rho + \frac{k(k-1)(k-2)(k-3)}{\pi^2} \int_0^\rho \sin^{-1} \left(\frac{\lambda}{1+2\lambda} \right) \right. \\
 & \frac{\partial \lambda}{(1-\lambda^2)^{\frac{1}{2}}} + \frac{k(k-1) \dots (k-5)}{\pi^3} \int_0^\rho \int_0^{\frac{u}{1+2u}} \sin^{-1} \left(\frac{\lambda}{1+2\lambda} \right) \\
 & \frac{\partial \lambda}{(1-\lambda^2)^{\frac{1}{2}}} + \frac{k(k-1) \dots (k-7)}{\pi^4} \int_0^\rho \int_0^{\frac{v}{1+2v}} \int_0^{\frac{u}{1+2u}} \sin^{-1} \left(\frac{\lambda}{1+2\lambda} \right) \\
 & \left. \frac{\partial \lambda}{(1-\lambda^2)^{\frac{1}{2}}} \frac{\partial u}{(1-u^2)^{\frac{1}{2}}} \frac{\partial v}{(1-v^2)^{\frac{1}{2}}} + \dots \right]. \quad (2.62)
 \end{aligned}$$

Portanto, (2.62) é uma expressão exata para $P_k(\rho)$; as integrais não podem ser expressas em termos de funções elementares, mais elas podem ser calculadas por integração numérica. Comparações com outras aproximações tem sido feitas (McFadden, 1956; Plackett, 1954; Sondhi, 1961).

2.5.3 Outros métodos de aproximação

Outros métodos para aproximar a integral normal multivariada incluem métodos de simulação de Monte Carlo, métodos de mudanças de variáveis, métodos de quadratura em análise numérica e condicionamento. Alguns desses métodos não se limitam somente a aproximar a probabilidade do quadrante, mas também se aplicam ao cálculo de integrais sobre outras regiões geométricas.

Métodos de Monte Carlo podem ser usados para aproximar o valor de uma integral quando resultados analíticos não são disponíveis. Porém, esses métodos não são recomendados para uso prático uma vez que existem procedimentos mais eficientes que levam em conta princípios de redução do erro (vide Moran,

1984). A partir da aplicação de métodos de Monte Carlo, Dunn, Kronmal e Yee (1968) fornecem extensivas tabelas para a distribuição t multivariada.

Cadwell (1951) aplicou transformações lineares para calcular a integral de uma função de densidade normal bivariada sobre qualquer polígono. Posteriormente, John (1966) estudou o cálculo de integrais de probabilidade sobre poliedros convexos. Milton (1972), por intermédio de uma modificação na quadratura adaptativa multidimensional de Simpson com erro controlado, propõe um algoritmo computacional para calcular a integral de probabilidade de uma distribuição normal multivariada sobre qualquer região geométrica. Genz e Kahaner (1986) descrevem um método para o caso especial na qual $\Sigma^{-1} = (\tau_{ij})$ satisfaz $\tau_{ij} = 0$ para todo $|i - j| > 1$. Uma outra aproximação é usar condicionamento. John (1959) considerou o caso da probabilidade do quadrante de uma variável normal k -dimensional através do condicionamento em $X_1 = x_1$, reduzindo assim a dimensão de k para $k - 1$. Ihm (1959) e Marsaglia (1963) estudaram a integral para uma distribuição normal multivariada quando a matriz de covariância é da forma $\Sigma = D + B$, onde D é uma matriz diagonal.

2.5.4 Aproximação da probabilidade do quadrante por séries infinitas

Os principais resultados a serem considerados, estão relacionados com expansões assintóticas para a função de densidade normal multivariada, e portanto para a própria g , na forma de séries infinitas. Consideramos a função g na forma definida em (2.51). Depois disso feito, aproximações para g podem

então ser obtidas computando-se um número finito de termos. Por exemplo, Pearson (1901) apresenta um método para calcular integrais da distribuição normal bivariada usando séries do tipo *tetrachoric*. Kendall (1941) considerou uma generalização dessas séries para diversas variáveis e investigou a expansão assintótica da probabilidade do quadrante para variáveis normais em R^3 . Moran (1948) adotou uma aproximação similar. David (1953) discute a utilidade do método de Moran e fornece expressões para as probabilidades do quadrante usando métodos geométricos. Ela comparou os seus resultados com os de Kendall e Moran e declarou que os mesmos são ineficientes por causa da demora na convergência das séries infinitas usadas nas expansões assintóticas. Recentemente, Harris e Sons (1980) estudaram a convergência das séries *tetrachoric*, e descobriram que a afirmação relativa à convergência das séries para todo $k \geq 2$ e toda matriz de covariância $k \times k$ é falsa. Ruben (1954) discutiu o significado geométrico dos momentos de estatísticas de ordem de uma população normal multivariada correlacionada, e apresentou resultados na forma de uma expansão assintótica para a probabilidade do quadrante. MacFadden (1956) forneceu resultados para o caso especial de $k = 4$ e $\rho_{ij} = \rho \in (-1/3, 1)$ e, posteriormente, MacFadden (1960), impondo condições mais fracas sobre os ρ'_{ij} s, obteve duas expansões em séries infinitas para $k = 4$. A aproximação proposta por Henery (1981) é encontrar uma expressão assintótica condicionada nas outras variáveis que são truncadas. A distribuição condicional das variáveis restantes pode ser escrita por uma expansão, a qual é similar a uma série *tetrachoric* no caso bivariado, mas tem convergência rápida sob determinadas condições.

A maioria dos artigos citados consideram em (2.51) o caso especial $a_1 = \dots =$

$a_k = 0$, isto é,

$$P_k = P_{0,R} \left[\bigcap_{i=1}^k \{X_i \leq 0\} \right], \quad (2.63)$$

e o problema de interesse é encontrar expressões para a probabilidade do quadrante P_k (em forma fechada se possível), em termos dos coeficientes de correlação. David (1953) usa a fórmula de inclusão -exclusão para estudar este problema; ela obteve o seguinte resultado:

$$P_2 = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho, \quad (2.64)$$

onde ρ é o coeficiente de correlação da distribuição normal bivariada, e então mostrou que

$$P_3 = \frac{1}{8} + \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{13} + \sin^{-1} \rho_{23}). \quad (2.65)$$

Se $k \geq 4$, então uma representação em forma fechada para P_k não é possível. Abrahamson (1964) estudou esse problema para $k = 4$ e obteve uma representação integral para P_4 . Impondo a condição adicional de que $\rho_{ij} = \rho$ para todo $i \neq j$, alguns casos especiais podem ser encontrados em Bacon (1963). Alguns desses casos são dados por (2.60) e (2.61). Em particular, se $\rho = 1/2$, pode ser verificado que

$$P_k = \frac{1}{n+1} \text{ para todo } k. \quad (2.66)$$

Childs (1967) estendeu os resultados de David (1953) e McFadden (1955). A integral de ordem k sobre um quadrante generalizado da distribuição normal multivariada foi transformada para uma série de integrais sobre um hipervolume. As séries foram derivadas para todo k e exibidas explicitamente até a ordem 6. Seus principais resultados são apresentados a seguir. Desejamos calcular

$$P_k = \int_0^\infty f_k(x) \partial^k(x), \quad (2.67)$$

onde

$$f_k(x) = \frac{1}{(2\pi)^{\frac{k}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x^t R^{-1} x\right)$$

com R matriz de correlação, R^{-1} sua inversa, $|R|$ seu determinante, ρ'_{ij} s os coeficientes de correlação e x o vetor coluna das k variáveis independentes. Sem perda de generalidade estamos supondo os x 's padronizados. Como $f_k(x) = f_k(-x)$, podemos escrever (2.67) como

$$P_k = \int_{-\infty}^{\infty} \prod_{i=1}^k u_i(x_i) f_k(-x) \partial^k(x), \quad (2.68)$$

onde

$$u_i(x_i) = \begin{cases} 1 & \text{se } x_i \geq 0 \\ 0 & \text{se } x_i < 0. \end{cases}$$

Agora, denotemos por $H(t)$ e $G(t)$ as transformadas de Fourier de $h(x)$ e $g(x)$, respectivamente. Segue do teorema de Parseval que

$$\int_{-\infty}^{\infty} h(x) g(x) dx = \int_{-\infty}^{\infty} H(t) G(-t) dt.$$

As transformadas de Fourier para $f_k(x)$ e $u_i(x_i)$ são

$$F_k(w) = \frac{1}{(2\pi)^k} \exp\left(-\frac{1}{2} w^t R w\right) \text{ e} \quad (2.69)$$

$$U_i(w_i) = \lim_{\alpha_i \rightarrow 0} \left\{ \pi \delta(w_i) - \frac{i w_i}{w_i^2 + \alpha_i^2} \right\}.$$

É necessário tomar $\lim_{\alpha_i \rightarrow 0} w_i / (w_i^2 + \alpha_i^2)$ depois de ter aplicado os outros limites, pois

$$\lim_{\alpha \rightarrow 0} \lim_{w \rightarrow 0} \frac{w}{w^2 + \alpha^2} = 0 \text{ e } \lim_{w \rightarrow 0} \lim_{\alpha \rightarrow 0} \frac{w}{w^2 + \alpha^2} = \infty. \quad (2.70)$$

Portanto, a integral normal multivariada pode ser escrita na forma

$$P_k = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \prod_{j=1}^k \left\{ \lim_{\alpha_j \rightarrow 0} \pi \delta(w_j) - \frac{i w_j}{w_j^2 + \alpha_j^2} \right\} \exp\left(-\frac{1}{2} w^t R w\right) \partial^k(w), \quad (2.71)$$

onde i denota a unidade imaginária. Para $k = 1$ e $k = 2$, segue de (2.71) que

$$\begin{aligned} P_1 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{\alpha \rightarrow 0} \left\{ \pi \delta(w) - \frac{iw}{w^2 + \alpha^2} \right\} \exp\left(-\frac{1}{2}w^2\right) \partial w \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \pi \delta(w) \exp\left(-\frac{1}{2}w^2\right) \partial w = \frac{1}{2} \text{ e} \end{aligned} \quad (2.72)$$

$$\begin{aligned} P_2 &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lim_{\alpha_1, \alpha_2 \rightarrow 0} \left\{ \pi \delta(w_1) - \frac{iw_1}{w_1^2 + \alpha_1^2} \right\} \left\{ \pi \delta(w_2) - \frac{iw_2}{w_2^2 + \alpha_2^2} \right\} \times \\ &\quad \exp\left\{-\frac{1}{2}(w_1^2 + 2\rho_{12}w_1w_2 + w_2^2)\right\} \partial w_1 \partial w_2. \end{aligned} \quad (2.73)$$

Expandindo o integrando em (2.73) temos que $P_2 = \mathcal{Z}_1 + 2\mathcal{Z}_2 + \mathcal{Z}_3$, onde

$$\begin{aligned} \mathcal{Z}_1 &= \frac{1}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lim_{\alpha_1, \alpha_2 \rightarrow 0} \{\delta(w_1)\delta(w_2)\} f(\rho, w) \partial w_1 \partial w_2 \\ &= 1/4, \\ \mathcal{Z}_2 &= -\frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lim_{\alpha_1, \alpha_2 \rightarrow 0} \{\pi \delta(w_1)\} \left\{ \frac{iw_2}{w_2^2 + \alpha_2^2} \right\} f(\rho, w) \partial w_1 \partial w_2 \\ &= 0 \text{ e} \\ f(\rho, w) &= \exp\left\{-\frac{1}{2}(w_1^2 + 2\rho_{12}w_1w_2 + w_2^2)\right\}. \end{aligned}$$

Para o terceiro termo temos para todo $a_1 > 0$ e $a_2 > 0$ que

$$\begin{aligned} \mathcal{Z}_3 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{w_1 w_2} \exp\left\{-\frac{1}{2}(a_1 w_1^2 + 2\rho_{12}w_1w_2 + a_2 w_2^2)\right\} \partial w_1 \partial w_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi_1 \xi_2} f(\rho, \xi) \partial \xi_1 \partial \xi_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi_1 \xi_2} [f(\rho, \xi) - f(\xi)] \partial \xi_1 \partial \xi_2 \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi_1 \xi_2} f(\xi) \partial \xi_1 \partial \xi_2, \end{aligned} \quad (2.74)$$

onde

$$\begin{aligned} f(\rho, \xi) &= \exp\left\{-(a_1 \xi_1^2 + 2\rho_{12}\xi_1\xi_2 + a_2 \xi_2^2)\right\} \text{ e} \\ f(\xi) &= \exp\{-a_1 \xi_1^2 - a_2 \xi_2^2\}. \end{aligned}$$

Esses resultados são obtidos depois de ter aplicado a transformação $w_1 = \sqrt{2}\xi_1$ e $w_2 = \sqrt{2}\xi_2$ na primeira igualdade.

Agora o último termo em (2.74) é igual a

$$\lim_{\alpha_1, \alpha_2 \rightarrow 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\xi_1}{(\xi_1^2 + \alpha_1^2)} \frac{\xi_2}{(\xi_2^2 + \alpha_2^2)} f(\xi) \partial \xi_1 \partial \xi_2 = 0,$$

pois o integrando é uma função par em ξ_1 e em ξ_2 . Portanto, (2.74) pode ser expresso como

$$\begin{aligned} \mathcal{Z}_3 &= -2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\rho_{12}} \exp(-(a_1 \xi_1^2 + 2u \xi_1 \xi_2 + a_2 \xi_2^2)) \partial u \partial \xi_1 \partial \xi_2 \\ &= -2\pi \int_0^{\rho_{12}} \frac{\partial u}{(a_1 a_2 - u^2)^{\frac{1}{2}}} \\ &= -2\pi \sin^{-1} \left\{ \frac{\rho_{12}}{\sqrt{a_1 a_2}} \right\}. \end{aligned} \quad (2.75)$$

Depois de trocar a ordem de integração segue que

$$P_2 = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho_{12}). \quad (2.76)$$

Para $k = 3$ usamos de novo (2.71) para escrever P_3 como a soma de 8 termos.

O primeiro é dado por

$$\frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \pi^3 \delta(w_1) \delta(w_2) \delta(w_3) \exp\left(-\frac{1}{2} w^t R w\right) \partial^3 w = \frac{1}{8}.$$

Os seguintes três termos e o último termo envolvem potências pares dos w 's e portanto suas integrais são nulas. Os três termos restantes têm forma similar ao caso $k = 2$; o primeiro deles tem a forma

$$\frac{\pi i^2}{(2\pi)^3} \int_{-\infty}^{\infty} \delta(w_1) \lim_{\alpha_2, \alpha_3 \rightarrow 0} \frac{w_2 w_3}{(\alpha_2^2 + w_2^2)(\alpha_3^2 + w_3^2)} \exp\left(-\frac{1}{2} w^t R w\right) \partial^3 w$$

e é igual a $\sin^{-1}(\rho_{23})/4\pi$. Portanto

$$P_3 = \frac{1}{8} \left[1 + \frac{2}{\pi} \{ \sin^{-1}(\rho_{12}) + \sin^{-1}(\rho_{13}) + \sin^{-1}(\rho_{23}) \} \right], \quad (2.77)$$

conforme mostrado por David (1953) e McFadden (1955). Dado que as integrais com número par de w 's são 0, o caso $k = 4$ pode ser escrito a partir de (2.71) como

$$P_4 = \frac{1}{16} \left\{ 1 + \frac{2}{\pi} \sum_{i < j} \sin^{-1}(\rho_{ij}) \right\} + \frac{1}{16\pi^4} w_{1234}, \quad (2.78)$$

onde

$$w_{1234} = \int_{-\infty}^{\infty} \frac{1}{w_1 w_2 w_3 w_4} \exp\left(-\frac{1}{2} w^t R w\right) \partial^4 w. \quad (2.79)$$

De forma similar temos

$$P_5 = \frac{1}{32} \left\{ 1 + \frac{2}{\pi} \sum_{i < j} \sin^{-1}(\rho_{ij}) + \frac{1}{\pi^4} \sum w_{ijkl} \right\}.$$

Quando $k > 6$, temos as seguintes aproximações:

$$P_{2k} = \frac{1}{2^{2k}} \left\{ 1 + \sum_{i < j=1}^{2k} \frac{2}{\pi} \sin^{-1}(\rho_{ij}) + \sum_{j=2}^k \left(\frac{i}{\pi}\right)^{2j} \sum_{k_1 < \dots < k_{2j}} w_{k_1, \dots, k_{2j}} \right\}. \quad (2.80)$$

Temos $(2k)!/2!(2k-2)!$ termos da forma $\sin^{-1}(\rho_{ij})$ e $(2k)!/(2j)!(2k-2j)!$ termos de forma $w_{k_1, k_2, \dots, k_{2j}}$, onde os k_l 's são permutados e $k_l \neq k_m$. Segue também que

$$P_{2k+1} = \frac{1}{2^{2k+1}} \left\{ 1 + \sum_{i < j=1}^{2k+1} \frac{2}{\pi} \sin^{-1}(\rho_{ij}) + \sum_{j=2}^k \left(\frac{i}{\pi}\right)^{2j} \sum_{k_1 < \dots < k_{2j}} w_{k_1, \dots, k_{2j}} \right\} \quad (2.81)$$

e de novo temos $(2k+1)!/2!(2k-1)!$ termos da forma $\sin^{-1}(\rho_{ij})$ e $(2k+1)!/(2j)!(2k+1-2j)!$ termos da forma $w_{k_1, k_2, \dots, k_{2j}}$, onde os k_l 's são permutados, $k_l \neq k_m$ e

$$w_{k_1, \dots, k_{2j}} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\exp(-\frac{1}{2} w^t R w)}{\prod_{l=1}^{2j} w_{kl}} \partial^{2j} w.$$

Temos reduzido a integral normal multivariada para uma série de integrais definidas sobre um hipervolume, as quais facilitam o cálculo computacional.

O cálculo de w_{1234} pode ser feito da seguinte maneira:

$$I = \frac{1}{\pi^4} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\exp\{-\frac{1}{2}(\sum_{i=1}^4 w_i^2 + 2 \sum_{i < j=1}^4 \rho_{ij} w_i w_j)\}}{w_1 w_2 w_3 w_4} \partial w_1 \dots \partial w_4.$$

Se $w_i = \sqrt{2}y_i$ e

$$g(\rho, y, t)\Delta_{y,t} = \exp\left\{-(ty_1^2 + \sum_{i=2}^4 y_i^2 + 2 \sum_{i<j=1}^4 \rho_{ij}y_iy_j)\right\}\partial y_1 \dots \partial y_4 \partial t,$$

temos que

$$I = \frac{1}{\pi^4} \int_1^\infty \dots \int_{-\infty}^\infty \frac{y_1}{ty_2y_3y_4} g(\rho, y, t)\Delta_{y,t}.$$

Integrando por partes com relação a y_1 obtemos

$$\begin{aligned} I &= -\frac{1}{\pi^4} \int_1^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \frac{\rho_{12}}{ty_3y_4} g(\rho, y, t)\Delta_{y,t} \\ &\quad - \frac{1}{\pi^4} \int_1^\infty \dots \int_{-\infty}^\infty \frac{\rho_{13}}{ty_2y_4} g(\rho, y, t)\Delta_{y,t} \\ &\quad - \frac{1}{\pi^4} \int_1^\infty \dots \int_{-\infty}^\infty \frac{\rho_{14}}{ty_2y_3} g(\rho, y, t)\Delta_{y,t} \\ &= I_1 + I_2 + I_3, \end{aligned}$$

onde I_j , $j = 1, 2, 3$, representa cada componente de I . Ao integrar o primeiro termo I_1 com relação a y_1 e y_2 , obtemos

$$I_1 = -\frac{1}{\pi^4} \int_1^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{\rho_{12}}{ty_3y_4} \frac{\pi}{(t - \rho_{12}^2)^{\frac{1}{2}}} h(\rho, y, t) \partial y_3 \partial y_4 \partial t,$$

onde

$$\begin{aligned} h(\rho, y, t) &= \exp \left[-\frac{\{h_3(t)y_3^2 + h_4(t)y_4^2 + 2h_{34}(t)y_3y_4\}}{1 - \frac{\rho_{12}^2}{t}} \right], \\ h_3(t) &= 1 - \rho_{23}^2 - \frac{1}{t}\rho_{12}^2 - \frac{1}{t}\rho_{13}^2 + \frac{2}{t}\rho_{12}\rho_{13}\rho_{23}, \\ h_4(t) &= 1 - \rho_{24}^2 - \frac{1}{t}\rho_{12}^2 - \frac{1}{t}\rho_{14}^2 + \frac{2}{t}\rho_{12}\rho_{14}\rho_{24} \text{ e} \\ h_{34}(t) &= \rho_{34} - \rho_{23}\rho_{24} - \frac{1}{t}\rho_{13}\rho_{14} - \frac{1}{t}\rho_{12}^2\rho_{34} + \frac{1}{t}\rho_{12}\rho_{14}\rho_{23} + \frac{1}{t}\rho_{12}\rho_{13}\rho_{24}. \end{aligned}$$

Segue de (2.75) que

$$I_1 = \frac{2\rho_{12}}{\pi^2} \int_1^\infty \frac{1}{t(t - \rho_{12}^2)^{\frac{1}{2}}} \sin^{-1} \left[\frac{h_{34}(t)}{\{h_3(t)h_4(t)\}^{\frac{1}{2}}} \right] \partial t.$$

Se $t = u^{-2}$, I_1 toma a forma

$$\frac{4\rho_{12}}{\pi^2} \int_0^1 \frac{1}{(1 - \rho_{12}^2 u^2)^{\frac{1}{2}}} \sin^{-1} \left[\frac{h_{34}(u)}{\{h_3(u)h_4(u)\}^{\frac{1}{2}}} \right] \partial u.$$

As integrais I_2, I_3 podem ser reduzidas de forma similar

$$I_j = \frac{4}{\pi^2} \rho_{1j} \int_0^1 \frac{1}{(1 - \rho_{1j}^2 u^2)^{\frac{1}{2}}} \sin^{-1} \left[\frac{\alpha - \beta u^2}{(c - du^2)^{\frac{1}{2}}(e - hu^2)^{\frac{1}{2}}} \right] \partial u.$$

Se $\rho_{12} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, obtemos, $I_2 = 0$, $I_3 = 0$ e

$$I_1 = \frac{4}{\pi^2} \sin^{-1}(\rho_{12}) \sin^{-1}(\rho_{34}).$$

Sun (1987) generaliza o método de Childs (1967) e oferece expressões alternativas para a aproximação numérica das integrais.

Os P'_k s definidos em (2.62), (2.63) ou (2.67) correspondem à probabilidade de nível $Q(k, k; \Sigma)$, isto é

$$Q(k, k; \Sigma) = P_r(X_1 \geq 0, X_2 \geq 0, \dots, X_k \geq 0),$$

onde $X = (X_1, \dots, X_k)^t$ é um vetor aleatório com distribuição normal multivariada. As probabilidades de nível $Q(k, \ell; \Sigma)$, $\ell = 1, 2, \dots, k-1$, são obtidas de forma recursiva a partir de $Q(k, k; \Sigma)$. As expressões para $Q(k, \ell; \Sigma)$ quando $k = 2$ e $k = 3$ são obtidas por Kudô (1963) e para $k = 4$ são dadas por Shapiro (1985). Elas são apresentadas a seguir.

Quando $k = 2$ temos

$$\begin{aligned} Q(2, 0; \Sigma) &= \frac{1}{2\pi} \cos^{-1}(\rho_{12}), \\ Q(2, 1; \Sigma) &= \frac{1}{2} \text{ e} \\ Q(2, 2; \Sigma) &= \frac{1}{2} - \frac{1}{2\pi} \cos^{-1}(\rho_{12}), \end{aligned}$$

onde ρ_{12} é o coeficiente de correlação associado com a matriz de covariância Σ de ordem 2×2 . Para $k = 3$, esses coeficientes assumem os valores

$$\begin{aligned} Q(3, 0; \Sigma) &= \frac{1}{2} - Q(3, 2; \Sigma), \\ Q(3, 1; \Sigma) &= \frac{1}{2} - Q(3, 3; \Sigma), \\ Q(3, 2; \Sigma) &= \frac{1}{4\pi} \{3\pi - \cos^{-1}(\rho_{12,3}) - \cos^{-1}(\rho_{13,2}) - \cos^{-1}(\rho_{23,1})\} \text{ e} \\ Q(3, 3; \Sigma) &= \frac{1}{4\pi} \{2\pi - \cos^{-1}(\rho_{12}) - \cos^{-1}(\rho_{13}) - \cos^{-1}(\rho_{23})\}, \end{aligned}$$

onde ρ_{ij} é o (i, j) -ésimo elemento da matriz de correlações associada com a matriz Σ de ordem 3×3 e $\rho_{ij,t}$'s são as correspondentes correlações parciais as quais são definidas por

$$\rho_{ij,t} = \frac{\rho_{ij} - \rho_{it}\rho_{jt}}{\sqrt{(1 - \rho_{it}^2)(1 - \rho_{jt}^2)}}.$$

Finalmente, para $k = 4$ temos

$$\begin{aligned} Q(4, 0; \Sigma) &= \frac{1}{2} - Q(4, 4; \Sigma) - Q(4, 2; \Sigma), \\ Q(4, 1; \Sigma) &= \frac{1}{8\pi} \left\{ -4\pi + \sum_{i>j; i, j \neq k} \cos^{-1}(\rho_{ij,k}) \right\}, \\ Q(4, 2; \Sigma) &= \frac{1}{4\pi^2} \sum_{i>j, k>l; k, l \neq i, j} \cos^{-1}(\rho_{ij}) \left\{ \pi - \cos^{-1}(\rho_{kl,ij}) \right\} \text{ e} \\ Q(4, 3; \Sigma) &= \frac{1}{8\pi} \left\{ 8\pi - \sum_{i>j; i, j \neq k} \cos^{-1}(\rho_{ij,k}) \right\}. \end{aligned}$$

A notação $\rho_{kl,ij}$ é o coeficiente de correlação parcial entre X_k e X_l fixadas X_i e X_j . O peso $Q(4, 4; \Sigma)$ é a probabilidade que $X \in R^4$ tenha todas as suas componentes positivas.

Para algumas hipóteses ordenadas chamadas de quase-ordem, a matriz Σ toma formas específicas de modo que a matriz de correlações associada não depende

dos parâmetros. Nesses casos de quase-ordem os coeficientes $Q(k, \ell; \Sigma)$'s são denotados por $P(k, \ell; w)$, onde $w = (w_1, \dots, w_k)^t$ são os pesos. Cálculos e aproximações para $P(k, \ell; w)$ podem ser vistos no Apêndice A.

Capítulo 3

Hipóteses restritas em modelos de regressão com distribuição log-gama generalizada

3.1 Introdução

Este capítulo é dedicado a discutir testes do tipo *one-sided* em modelos de regressão com distribuição log-gama generalizada com parâmetro de escala não-constante e dados não censurados. Obtemos a distribuição nula assintótica de três estatísticas assintoticamente equivalentes para duas situações de hipóteses gerais do tipo *one-sided*. Para alguns casos particulares, tais como aqueles de testes de hipóteses em ordem simples ou ordem simples em forma de árvore para o parâmetro de locação, parâmetro de escala e percentis em g populações e para os interceptos de modelos de regressão em retas paralelas, a distribuição nula assintótica reduz à conhecida distribuição qui-quadrado barra, $\bar{\chi}^2$. Con-

sequentemente, todas as aproximações desenvolvidas para essa distribuição, especialmente para quatro ou mais restrições, podem ser aplicadas aos casos enunciados acima. A robustez da distribuição $\bar{\chi}^2$ é estudada para alguns casos particulares através de simulações. Finalmente, apresentamos dois exemplos ilustrativos.

3.2 Distribuição log-gama generalizada

Seja T uma variável aleatória com função de densidade

$$f(t; a, d, p) = \frac{p}{a^d \Gamma(d/p)} t^{d-1} \exp[-(t/d)^p], \quad t > 0, \quad (3.1)$$

com parâmetros a , d e p positivos [vide Stacy, 1962]. A função (3.1) pode ser vista como uma generalização da distribuição gama e propriedades elementares da variável aleatória T podem ser verificadas diretamente. Denotando sua função de distribuição acumulada por $F(t; a, d, p)$ e sua função geratriz de momentos por $M(\theta)$, temos

$$F(t; a, d, p) = \Gamma_z(d/p) / \Gamma(d/p) \text{ e}$$

$$M(\theta) = \sum_{r=0}^{\infty} \frac{(\theta a)^r}{r!} \left[\frac{\Gamma((d+r)/p)}{\Gamma(d/p)} \right],$$

onde $z = (t/a)^p$ e $\Gamma_z(d/p) = \int_0^z v^{\frac{d}{p}-1} \exp(v) dv$. Fazendo $\alpha = a$, $\beta = p$ e $k = d/p$, o modelo (3.1) pode ser escrito na forma

$$f(t; \alpha, \beta, k) = \frac{\beta}{\alpha^{\beta k} \Gamma(k)} t^{\beta k-1} \exp[-(t/\alpha)^\beta], \quad t > 0, \quad (3.2)$$

onde $\alpha > 0$, $\beta > 0$ e $k > 0$. Essa é uma família flexível de distribuições que inclui como casos especiais as distribuições exponencial ($\beta = k = 1$), Weibull

($k = 1$), e gama ($\beta = 1$). A distribuição log-normal aparece como uma forma limite de (3.2). Assim, o modelo gama generalizado abrange como casos especiais todas as principais distribuições comumente usadas para a análise de dados de sobrevivência.

Os primeiros trabalhos com a distribuição gama generalizada encontraram dificuldades no desenvolvimento de procedimentos de inferência. Por exemplo, com o modelo escrito na forma (3.2), com amostras de tamanho 200 a 300, os algoritmos para determinar as estimativas de máxima verossimilhança apresentam dificuldades para convergir (Hager e Bain, 1970). Com amostras simuladas de (3.2), Hager e Bain notaram que os problemas de convergência são maiores com k grande do que com k pequeno. Outros autores como Parr e Webster (1965), Harter (1967), Stacy e Minhram (1965) encontraram dificuldades similares com a estimação por máxima verossimilhança. Adicionalmente, foi difícil desenvolver procedimentos adequados para testes e estimação por intervalo. Métodos para grandes amostras baseados em aproximações pela normal, para a distribuição assintótica dos estimadores de máxima verossimilhança $\hat{\alpha}$, $\hat{\beta}$ e $\hat{k}$ mostraram-se insatisfatórios uma vez que a distribuição normal limite é aproximada muito lentamente. Hager e Bain notaram que a aproximação normal pode ser inadequada mesmo para tamanhos amostrais da ordem de 400.

Esses problemas limitaram o uso da distribuição gama generalizada, no entanto Prentice (1974) aclarou grande parte do assunto. Muitas das dificuldades com o modelo (3.2) são devidas ao fato das distribuições serem similares mesmo para valores muito diferentes para os parâmetros α , β e k . Prentice (1974) propôs um procedimento alternativo para amenizar esse problema, o qual descrevemos a seguir.

Sejam $Y = \log T$, $u = \log \alpha$ e $b = \beta^{-1}$. Aplicando a transformação

$$W_1 = \beta(\log T - \log \alpha),$$

temos que $w_1 \in R$, $\partial t / \partial w_1 = \beta^{-1} \exp(\frac{w_1}{\beta} + \log \alpha)$ e a família de distribuições (3.2) é transformada para

$$f(w_1; k) = \frac{1}{\Gamma(k)} \exp(kw_1 - \exp(w_1)), k > 0. \quad (3.3)$$

Para a variável aleatória W_1 temos os seguintes resultados:

$$E(W_1) = \frac{\partial}{\partial k} \log \Gamma(k) = \Psi(k) \text{ e} \quad (3.4)$$

$$E(W_1^2) = \Psi'(k) + \Psi^2(k). \quad (3.5)$$

Para $k > 0$, temos que $\Psi(k)$ é a função digama, $\Psi(k+1) = \Psi(k) + 1/k$, $\Psi(1) = -\gamma$ é a constante de Euler e

$$\Psi(k) = \log k - 1/2k - 1/12k^2 + 1/120k^4 - \dots$$

Também temos que $\Psi'(k+1) = \Psi'(k) - 1/k^2$, $\Psi'(1) = \pi^2/6$ e

$$\Psi'(k) = 1/k + 1/2k^2 + 1/6k^3 - 1/30k^5 + 1/42k^7 + \dots$$

De (3.4) e (3.5) obtemos a variância de W_1 , dada por $\text{Var}(W_1) = \frac{\partial^2}{\partial k^2} \log \Gamma(k) = \Psi'(k)$. Também obtemos

$$E[W_1 \exp(W_1)] = k\Psi(k+1) \text{ e} \quad (3.6)$$

$$E[W_1^2 \exp(W_1)] = k[\Psi'(k+1) + \Psi^2(k+1)]. \quad (3.7)$$

Quando $k \rightarrow \infty$, a média de W_1 dada por (3.4) e a variância de W_1 são infinitas. Aplicamos uma nova transformação

$$W = \sqrt{k}(W_1 - \log k).$$

Daí obtemos $\partial w_1 / \partial w = k^{-1/2}$ e portanto

$$f(w; k) = \frac{k^{k-\frac{1}{2}}}{\Gamma(k)} \exp \left[\sqrt{k}w - k \exp(w/\sqrt{k}) \right], \quad w \in R, \quad k > 0. \quad (3.8)$$

Note que $W = \sqrt{k}(W_1 - \log k) = (Y - \mu)/\sigma$, onde

$$\sigma = \frac{b}{\sqrt{k}} \quad \text{e} \quad \mu = u + b \log k.$$

Segue então que $Y = \log T$ tem função de densidade dada por

$$f(y; \mu, \sigma, k) = \frac{k^{k-\frac{1}{2}}}{\sigma \Gamma(k)} \exp \left[\sqrt{k}(y - \mu)/\sigma - k \exp((y - \mu)/\sigma \sqrt{k}) \right], \quad (3.9)$$

para $y \in R$, $\mu \in R$, $\sigma > 0$ e $k > 0$. Quando $k \rightarrow \infty$, (3.8) aproxima-se da densidade normal padrão. Com o modelo escrito na forma (3.9), $k = 1$ e $k = \infty$ levam às distribuições do valor extremo e normal padrão para $Y = \log T$, ou equivalentemente às distribuições Weibull e log-normal para T , respectivamente (vide Figura 3.1).

Como é evidente pela Figura 3.1, distribuições com valores muito diferentes para k podem ser similares e isso cria problemas de estimação, principalmente no caso da parametrização original. Duas distribuições que têm valores muito diferentes para o parâmetro k , mas valores similares para os parâmetros μ e σ terão valores completamente diferentes para os parâmetros b e u . Isso faz com que o problema de estimação seja mais sério nas parametrizações (u, b, k) e (α, β, k) do que na parametrização (μ, σ, k) .

Uma reparametrização mais usual define μ e σ como acima, mas substitui k por $\lambda = k^{-1/2}$. As distribuições normal e do valor extremo correspondem a $\lambda = 0$ e $\lambda = 1$, respectivamente.

Prentice (1974) e Farewell e Prentice (1977) generalizaram a família (3.9) e incluíram distribuições com $\lambda < 0$, considerando modelos nos quais $W = (Y - \mu)/\sigma$ tenha função de densidade

$$f(w; \lambda) = \frac{|\lambda|}{\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left[\lambda^{-2} (\lambda w - \exp(\lambda w)) \right], \quad -\infty < w < \infty, \quad (3.10)$$

onde $-\infty < \lambda < \infty$, $\lambda \neq 0$. Com essa reformulação do modelo, a estimação de máxima verossimilhança e métodos para grandes amostras são mais facilmente tratáveis.

Figura 3.1: Distribuição log-gama para diferentes valores de k

3.3 Modelo de regressão log-gama generalizado

Sejam Y_{ij} , $i = 1, \dots, g$ e $j = 1, \dots, r_i$, variáveis aleatórias mutuamente independentes tais que

$$Y_{ij} = x_i^t \beta + \sigma_i Z_{ij}, \quad (3.11)$$

onde Z_{ij} tem distribuição log-gama generalizada com função de densidade de probabilidade dada em (3.9), $x_i = (x_{i1}, \dots, x_{ip})^t$ é um vetor $p \times 1$ com os valores de p covariáveis, $\beta = (\beta_1, \dots, \beta_p)^t$ é um vetor $p \times 1$ de coeficientes de regressão e $\sigma_i > 0$ são os parâmetros de escala, assumidos desconhecidos enquanto k será assumido conhecido. A função conjunta de verossimilhança fica dada por

$$\ell(\theta) = \prod_{i=1}^g \left\{ r_i! \prod_{j=1}^{r_i} \frac{k^{k-\frac{1}{2}}}{\sigma_i \Gamma(k)} \exp \left[\sqrt{k} \frac{y_{ij} - \mu_i}{\sigma_i} - k \exp \left(\frac{y_{ij} - \mu_i}{\sqrt{k} \sigma_i} \right) \right] \right\}.$$

Fazendo $z_{ij} = (y_{ij} - x_i^t \beta) / \sigma_i$, a função log-verossimilhança de $\theta = (\beta^t, \sigma^t)^t$, onde $\sigma = (\sigma_1, \dots, \sigma_g)^t$, pode ser reescrita na forma

$$L(\theta) \propto \sum_{i=1}^g [-r_i \log(\sigma_i \Gamma(k) k^{k-\frac{1}{2}}) + \sum_{j=1}^{r_i} \{ \sqrt{k} z_{ij} - k \exp(\frac{z_{ij}}{\sqrt{k}}) \}]. \quad (3.12)$$

Segue de Pratt (1981) que $L(\theta)$ é uma função côncava em θ (vide Figura 3.2). Isso garante a unicidade das estimativas de máxima verossimilhança $(\hat{\beta}^t, \hat{\sigma}^t)^t$, se elas existem. Note que $Z_{ij} = (Y_{ij} - x_i^t \beta) / \sigma_i$ segue uma distribuição log-gama generalizada padrão. Temos então que a matriz de informação de Fisher para $(\beta^t, \sigma^t)^t$ pode ser expressa como

$$F = \begin{bmatrix} X^t W X & D^t W \\ W^t D & \Phi(k) W \end{bmatrix},$$

onde X é uma matriz $g \times p$ com linhas $x_i^t = (x_{i1}, \dots, x_{ip})$, D é uma matriz $g \times p$, G é uma matriz $p \times g$ com um em todas suas posições, $W = \text{diag}\{w_1, \dots, w_g\}$,

$$w_i = r_i \sigma_i^{-2}, \quad i = 1, \dots, g,$$

$$\begin{aligned} D^t &= [1 + \sqrt{k}\Psi(k)]G + k^{3/2}\Psi(k+1)(X^t - \sqrt{k}G) \text{ e} \\ \Phi(k) &= k^2[\Psi^2(k+1) + \Psi(k+1) + \Psi'(k+1)] - \sqrt{k}\Psi(k). \end{aligned}$$

Agora, vamos supor as seguintes condições suficientes para que $\frac{F}{n} \rightarrow V_\theta$ seja positiva definida:

1. O número de grupos é finito (g é finito), onde $n = r_1 + r_2 + \dots + r_g$.
2. $\frac{r_i}{n} \rightarrow a_i > 0$ quando $n \rightarrow \infty$;
3. $\sum_{i=1}^g x_i x_i^T$ é de posto completo, e
4. os regressores são absolutamente limitados.

Assim temos que $\lim_{n \rightarrow \infty} n^{-1}F = V_\theta$, onde V_θ é uma matriz positiva definida $(p+g) \times (p+g)$ a qual toma a mesma forma funcional de F acima com W sendo substituído por $W_a = \text{diag}\{w_{a1}, \dots, w_{ag}\}$, onde $w_{ai} = a_i \sigma_i^{-2}$, $i = 1, \dots, g$. Consequentemente, sob condições gerais de regularidade (vide Bradley e Gart, 1962; vide também Apêndice B), temos que quando $n \rightarrow \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta^{-1}),$$

e, em particular,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta^{-1}) \text{ e } \sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N(0, V_\sigma^{-1}).$$

Figura 3.2: Distribuição log-gama generalizada

onde

$$\begin{aligned} V_{\beta}^{-1} &= (X^t W_a X)^{-1} \{I_p + D^t W_a V_{\sigma}^{-1} (X^t W_a X)^{-1} D W_a^t\}, \\ V_{\sigma}^{-1} &= \{\Phi(k) - W_a^t D (X^t W_a X)^{-1} D^t\} W_a\}^{-1} \text{ e} \\ V_{\beta\sigma}^{-1} &= -(X^t W_a X)^{-1} D^t W_a V_{\sigma}^{-1}, \end{aligned}$$

onde I_p denota a matriz identidade de ordem p .

No caso de g populações, temos $D^t = \gamma(k)G$ onde $\gamma(k) = 1 + \sqrt{k}\Psi(k) + k^{3/2}\Psi(k+1)(1 + \sqrt{k})$, assim obtemos

$$V_{\theta}^{-1} = \begin{bmatrix} \theta(k)W_a^{-1} & \Pi(k)W_a^{-1} \\ \Pi(k)W_a^{-1} & \beta(k)W_a^{-1} \end{bmatrix},$$

onde $\theta(k) = 1 + \gamma^2(k)\beta(k)$, $\Pi(k) = -\gamma(k)\beta(k)$ e $\beta(k) = \{\Phi(k) - (1 + \sqrt{k}\Psi(k) + k^{3/2}\Psi(k+1)(1 - \sqrt{k}))\}^{-1}$.

3.4 Hipóteses restritas

Seja $h(\theta) = (h_1(\theta), \dots, h_k(\theta))^t$, onde $k \leq (p + g)$, e assumamos que $\Pi(\theta) = \{(\partial/\partial\theta)h(\theta)\}$ são funções contínuas de θ tal que $\Pi(\theta)$ é de posto completo. Então, quando $n \rightarrow \infty$, temos que

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \xrightarrow{d} N(0, \Pi(\theta)V_{\theta}^{-1}\Pi(\theta)^t).$$

Suponhamos agora, que h se divide em dois subvetores, h_1 e h_2 , os quais são formados pelos primeiros k_1 e os restantes $(k - k_1)$ elementos de h , respectivamente. Desejamos testar a hipótese $H_0 : h_1(\theta) = 0, h_2(\theta) = 0$ contra $H_1 : h_1(\theta) \neq 0, h_2(\theta) \geq 0$, com pelo menos uma desigualdade estrita sob H_1

quando $k_1 = 0$. Supondo que as condições de regularidade de Gouriéroux e Monfort (1995, Capítulo 21) seguem para o modelo log-gama generalizado, temos para $c \geq 0$ que

$$\lim_{n \rightarrow \infty} Pr\{\mathcal{W} \geq c\} = \sum_{\ell=0}^q Q(k, \ell; \Sigma) Pr\{\chi_{q+k_1}^2 \geq c\}, \quad (3.13)$$

onde $q = k - k_1$, $\mathcal{W}$ é a estatística da razão de verossimilhança ou alguma outra estatística assintoticamente equivalente tal como a estatística de Wald ou a estatística de escore, $Q(q, \ell; \Sigma)$'s são as probabilidades de nível definidas no Capítulo 2, $\Sigma = \Pi_2(\theta)V_\theta^{-1}\Pi_2(\theta)'$ e $\Pi_2(\theta) = (\partial/\partial\theta)h_2(\theta)$. Os coeficientes $Q(q, \ell; \Sigma)$'s sob H_0 , dependem dos coeficientes de correlação associados à matriz Σ , cuja ordem é $q \times q$. É importante observar que os coeficientes da distribuição nula assintótica dependem aqui de $\theta = (\beta', \sigma')'$, o que pode dificultar bastante a obtenção da distribuição nula assintótica menos favorável. A substituição de θ por uma estimativa consistente não garante que se esteja na distribuição nula assintótica menos favorável. Assim, o desenvolvimento de algoritmos de busca torna-se relevante apesar do custo computacional que poderá estar envolvido. Fahrmeir e Klinger (1994) discutem o problema de encontrar a distribuição nula menos favorável em modelos lineares generalizados e chegam a conclusões interessantes. Primeiro, verificam após vários estudos de simulação que a distribuição nula menos favorável ocorre, em geral, quando todas as desigualdades são satisfeitas em forma de igualdades, isto é, no conjunto $E = \{\theta : h(\theta) = 0\}$. Verificam também que mesmo quando há mais de um ponto menos favorável em E , as diferenças entre as probabilidades de rejeição do teste estatístico avaliadas em cada ponto, são desprezíveis. Sugerem então um algoritmo de busca que consiste em calcular a probabilidade de rejeição inicialmente em $\hat{\theta}^0$ e depois para um conjunto de valores no conjunto E . Se o coeficiente $Q(q, q; \Sigma)$ avaliado nesses pontos for menor do

que os demais coeficientes, deve-se continuar o cálculo das probabilidades de rejeição para valores de θ nos conjuntos em que uma ou mais igualdades não são satisfeitas, isto é, em conjuntos do tipo

$$E_m = \{\theta : h_{b_1}(\theta) = 0, \dots, h_{b_m}(\theta) = 0\}, m < q \text{ e } b_j \in P = \{1, \dots, p\}.$$

Caso contrário, há grandes chances de o ponto menos favorável estar no conjunto E .

Uma outra possibilidade, que investigamos neste trabalho, é a procura de estruturas para a matriz modelo X em que os pesos $Q(q, \ell; \Sigma)$'s deixem de depender sob a hipótese nula (através de Σ) dos parâmetros, particularmente para o caso de testarmos $H_0 : h(\theta) \geq 0$ contra $H_1 : H_0$, onde $H_1 : \theta \in R^{p+q}$. Segundo Wolak (1991) isso somente seria possível em situações triviais sem interesse prático, todavia Paula e Sen (1995) mostram que é possível $Q(q, \ell; \Sigma)$ reduzir a $P(\ell + 1, q + 1; w)$ em alguns modelos lineares generalizados com estrutura balanceada. Mostramos aqui que esses resultados se estendem para algumas situações de importância prática em modelos log-gama generalizados. Estudamos ainda a robustez da aproximação quando a estrutura balanceada proposta não se verifica mais. Apresentamos também alguns casos particulares de hipóteses do tipo *one-sided*.

3.4.1 Ordem simples

Inicialmente vamos supor $g = 5$ e que as dispersões são constantes, isto é, $\sigma_i = \sigma$. Testemos a hipótese $H_0 : \mu_1 = \dots = \mu_5$ contra $H_1 : \mu_1 \leq \dots \leq \mu_5$ com pelo menos uma desigualdade estrita. Essas hipóteses são equivalentes a

$H_0 : \Pi_2 \mu = 0$ contra $H_1 : \Pi_2 \mu \geq 0$, onde $\mu = (\mu_1, \dots, \mu_5)^t$, e

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

Lembrando que $\theta(k) = 1 + \gamma^2(k)\beta(k)$ com $\beta(k) = \{\Phi(k) - (1 + \sqrt{k}\Psi(k) + k^{\frac{3}{2}}\Psi(k+1)(1 - \sqrt{k})\}^{-1}$, temos então que

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \theta(k)(S + S_0)$$

com $S_0 = \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e

$$S = \begin{pmatrix} w_1^{-1} & -w_2^{-1} & 0 & 0 \\ -w_2^{-1} & w_2^{-1} & -w_3^{-1} & 0 \\ 0 & -w_3^{-1} & w_3^{-1} & -w_4^{-1} \\ 0 & 0 & -w_4^{-1} & w_4^{-1} \end{pmatrix},$$

onde w_i , para simplicidade de notação, é dado por $w_i = a_i \sigma_i^{-2}$.

Sob H_0 e a partir da suposição $\sigma_i = \sigma$ é facilmente visto, por exemplo, que

$$\begin{aligned} \rho_{12} &= - \left\{ \frac{a_1 a_3}{(a_1 + a_2)(a_2 + a_3)} \right\}^{1/2}, \\ \rho_{23} &= - \left\{ \frac{a_2 a_4}{(a_3 + a_2)(a_3 + a_4)} \right\}^{1/2} \text{ e} \\ \rho_{34} &= - \left\{ \frac{a_3 a_5}{(a_4 + a_3)(a_4 + a_5)} \right\}^{1/2}, \end{aligned}$$

$\rho_{ii} = 1$ e $\rho_{ij} = 0$ para $|i - j| > 1$. Na sua forma geral, supomos que $\sigma_i = \sigma$ e testamos $H_0 : \mu_1 = \dots = \mu_g$ contra $H_1 : \mu_1 \leq \dots \leq \mu_g$ com pelo menos uma desigualdade estrita. Assim, Π_2 ou equivalentemente $\Pi_2(\theta)$ será uma matriz $(g-1) \times (p+g)$ cuja j -ésima linha tem -1 na j -ésima posição, 1 na $j+1$ -ésima

posição e zeros nas posições restantes. Os coeficientes de correlação, sob H_0 , associados à matriz Σ são dados por ($\rho_{ii} = 1$)

$$\rho_{i(i+1)} = - \left[\frac{a_i a_{i+2}}{(a_{i+1} + a_i)(a_{i+1} + a_{i+2})} \right]^{1/2} \quad (3.14)$$

para $i = 1, \dots, g-2$ e $\rho_{ij} = 0$, para $|i - j| \geq 2$. Isso significa que a distribuição nula assintótica da estatística $\mathcal{W}$ para efeito de testar a homogeneidade de médias contra uma alternativa de ordem simples, reduz à conhecida distribuição qui-quadrado barra

$$\lim_{n \rightarrow \infty} Pr\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a) Pr\{\chi_{\ell-1}^2 \geq c\}, \quad (3.15)$$

onde $P(\ell, g; a) = Q(g-1, \ell-1; \Sigma)$ e $a = (a_1, \dots, a_g)^t$ são os pesos. Em particular, se temos pesos iguais, $a_1 = \dots = a_g$, os coeficientes $P(\ell, g; a)$'s tomam a fórmula recursiva $P_S(1, g) = 1/g$, $P_S(g, g) = 1/g!$ e $P_S(i, g) = \frac{1}{g} P_S(i-1, g-1) + \frac{g-1}{g} P_S(i, g-1)$, $i = 2, \dots, g-1$ (vide Apêndice A). Robertson e Wright (1983) (ver também Siskind, 1976) têm estudado a distribuição nula de estatísticas restritas para pesos iguais quando os tamanhos amostrais não são muito diferentes. Eles sugerem usar a aproximação quando o quociente entre o maior e o menor peso não for superior a 3.5. Para $\mu_i = \mu$ e hipóteses $H_0 : \sigma_1 = \dots = \sigma_g$ contra $H_1 : \sigma_1 \leq \dots \leq \sigma_g$, a distribuição nula assintótica de $\mathcal{W}$ é também dada por (3.15) com coeficientes de correlação, sob H_0 , dados em (3.14).

3.4.2 Ordem simples em forma de árvore

Outra situação de importância prática é aquela denominada de ordem simples em forma de árvore, isto é, $H_0 : \mu_1 = \dots = \mu_g$ contra $H_1 : \mu_1 \leq [\mu_2, \dots, \mu_g]$

com pelo menos uma desigualdade estrita. Vamos assumir aqui que $\sigma_i = \sigma$ e inicialmente consideremos que $g = 5$. Então H_0 e H_1 podem ser escritas como $H_0 : \Pi_2 \mu = 0$ e $H_1 : \Pi_2 \mu \geq 0$, onde $\mu = (\mu_1, \dots, \mu_5)^t$ e

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Temos então,

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \theta(k) w_1^{-1} (A + A_0)$$

com $A_0 = w_1 \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e A é uma matriz com 1 em todas suas posições. Novamente, sob H_0 e a partir da suposição $\sigma_i = \sigma$, podemos mostrar que

$$\begin{aligned} \rho_{13} &= \left\{ \frac{a_2 a_4}{(a_1 + a_2)(a_1 + a_4)} \right\}^{1/2}, \\ \rho_{34} &= \left\{ \frac{a_4 a_5}{(a_1 + a_4)(a_1 + a_5)} \right\}^{1/2}, \end{aligned}$$

os demais ρ'_{ij} s são obtidos de forma similar. Na sua forma geral, se assumimos que $\sigma_i = \sigma$, a distribuição nula assintótica de $\mathcal{W}$ toma, nesse caso, a mesma forma funcional de (3.15) com os coeficientes de correlação, sob H_0 , associados com a matriz Σ expressados como segue ($\rho_{ii} = 1$)

$$\rho_{ij} = \left[\frac{a_{i+1} a_{j+1}}{(a_1 + a_{i+1})(a_1 + a_{j+1})} \right]^{1/2}, \quad (3.16)$$

para $1 \leq i \neq j \leq g-1$. A fórmula recursiva para as probabilidades com pesos iguais, denotadas por $P_T(i, g)$, é dada em Robertson, Wright e Dykstra (1988). Wright e Tran (1985) tem investigado o uso da aproximação de pesos iguais para tamanhos amostrais desiguais. Similarmente ao caso de ordem simples,

a aproximação parece ser adequada para variações moderadas entre os pesos. De igual forma, a distribuição nula assintótica de $\mathcal{W}$ para o caso de testar $H_0 : \sigma_1 = \cdots = \sigma_g$ contra $H_1 : \sigma_1 \leq [\sigma_2, \dots, \sigma_g]$ é também dada por (3.15) com coeficientes de correlação, sob H_0 , dados em (3.16).

Limites superiores e inferiores para (3.15) também têm sido investigados para os casos de ordem simples e ordem simples em forma de árvore. Lee, Robertson e Wright (1993) apresentam uma revisão desses limites e obtêm aproximações mais precisas supondo que os pesos estejam em ordem, isto é, $a_1 \leq \cdots \leq a_g$ e $a_1 \leq a_i, i = 2, \dots, g$.

3.4.3 Hipótese unimodal

Suponhamos que estamos interessados em testar a hipótese $H_0 : \mu_1 = \cdots = \mu_g$ contra uma alternativa unimodal. Como exemplo, vamos assumir homogeneidade das variâncias e inicialmente consideremos $g = 5$. Então H_0 e H_1 podem ser escritas como $H_0 : \mu_1 = \cdots = \mu_5$ e $H_1 : \mu_1 \leq \mu_2 \leq \mu_3 \geq \mu_4 \geq \mu_5$ ou equivalentemente, $H_0 : \Pi_2 \mu = 0$ e $H_1 : \Pi_2 \mu \geq 0$, onde $\mu = (\mu_1, \dots, \mu_5)^t$ e

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Usando $\theta(k)$ e $\beta(k)$ como definidas no caso de ordem simples, temos então que

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \theta(k)(U + U_0)$$

com $U_0 = \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e

$$U = \begin{pmatrix} w_1^{-1} & -w_2^{-1} & 0 & 0 \\ -w_2^{-1} & w_2^{-1} & w_3^{-1} & 0 \\ 0 & w_3^{-1} & w_3^{-1} & -w_4^{-1} \\ 0 & 0 & -w_4^{-1} & w_4^{-1} \end{pmatrix}.$$

Sob H_0 e a partir da suposição $\sigma_i = \sigma$ é facilmente visto, por exemplo, que

$$\begin{aligned} \rho_{12} &= - \left\{ \frac{a_1 a_3}{(a_1 + a_2)(a_2 + a_3)} \right\}^{1/2}, \\ \rho_{23} &= \left\{ \frac{a_2 a_4}{(a_3 + a_2)(a_3 + a_4)} \right\}^{1/2} \text{ e} \\ \rho_{34} &= - \left\{ \frac{a_3 a_5}{(a_4 + a_3)(a_4 + a_5)} \right\}^{1/2}, \end{aligned}$$

$\rho_{ii} = 1$ e $\rho_{ij} = 0$ para $|i - j| > 1$. Na sua forma geral, supomos que $\sigma_i = \sigma$ e testamos $H_0 : \mu_1 = \dots = \mu_g$ contra uma alternativa unimodal com pelo menos uma desigualdade estrita. Assim, a distribuição nula assintótica de $\mathcal{W}$ toma, nesse caso, a mesma forma funcional de (3.15) com os coeficientes de correlação, sob H_0 , associados à matriz Σ dados por ($\rho_{ii} = 1$)

$$\rho_{i(i+1)} = (-1)^i \left[\frac{a_i a_{i+2}}{(a_{i+1} + a_i)(a_{i+1} + a_{i+2})} \right]^{1/2}, \quad (3.17)$$

para $i = 1, \dots, g-2$ e $\rho_{ij} = 0$, para $|i - j| \geq 2$.

Para $\mu_i = \mu$ e hipóteses $H_0 : \sigma_1 = \dots = \sigma_g$ contra a alternativa unimodal, a distribuição nula assintótica de $\mathcal{W}$ é também dada por (3.15) com coeficientes de correlação, sob H_0 , dados em (3.17).

3.4.4 Hipótese do tipo *loop*

Inicialmente vamos supor $g = 5$ e que as dispersões são constantes, isto é, $\sigma_i = \sigma$. Testemos a hipótese $H_0 : \mu_1 = \dots = \mu_5$ contra $H_1 : \mu_1 \leq [\mu_2, \mu_3, \mu_4] \leq \mu_5$ com pelo menos uma desigualdade estrita. Essas hipóteses são equivalentes a $H_0 : \Pi_2 \mu = 0$ contra $H_1 : \Pi_2 \mu \geq 0$, onde $\mu = (\mu_1, \dots, \mu_5)^t$, e

$$\Pi_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Temos então que

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \theta(k) \begin{pmatrix} L_1 + L_0 & -L_0 \\ -L_0 & L_2 + L_0 \end{pmatrix}$$

onde $L_0 = \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1})$, L_1 uma matriz quadrada de ordem 3 com w_1^{-1} em todas suas posições e L_2 uma matriz similar a L_1 com w_5^{-1} em todas suas posições. Nota-se portanto que a matriz $\Sigma = \Pi_2 V_\beta^{-1} \Pi_2^t$ é singular não satisfazendo assim a condição (positiva definida) requerida nos cálculos das probabilidades de nível $Q(g-1, \ell; \Sigma)$'s definidas na Seção 2.5. No entanto, Singh e Wright (1993) (vide Apêndice A.5) deduzem os níveis de probabilidade até $g = 5$ para o caso de ordem do tipo *loop* e apresentam aproximações para $g \geq 6$. É possível mostrar, supondo $\sigma_i = \sigma$, que a distribuição nula assintótica de $\mathcal{W}$ toma a forma dada em (3.15) com $P(\ell, g; a)$ sendo substituída por $P_L(\ell, g; a)$ definida no Apêndice A.5.

De maneira geral os coeficientes assumem as seguintes formas:

$$\rho_{ij} = \left\{ \frac{a_{i+1} a_{j+1}}{(a_1 + a_{i+1})(a_1 + a_{j+1})} \right\}^{\frac{1}{2}} \quad i, j = 1, \dots, g-2; \quad i \neq j$$

$$\rho_{ij} = - \left\{ \frac{a_1 a_g}{(a_1 + a_{i+1})(a_{i+1} + a_g)} \right\}^{\frac{1}{2}} \quad i = 1, \dots, g-2; \quad j = g-2+i,$$

$$\rho_{ij} = \left\{ \frac{a_{i-g+3} a_{j-g+3}}{(a_{i-g+3} + a_g)(a_{j-g+3} + a_g)} \right\}^{\frac{1}{2}} \quad i = g-1, \dots, 2(g-2);$$

$j = g, \dots, 2(g-2)$, $\rho_{ii} = 1$ e $\rho_{ij} = 0$ para os demais casos.

3.5 Modelo de regressão do valor extremo

Como é de maior interesse prático, estudamos nesta seção a aplicação da teoria de testes do tipo *one-sided* à distribuição do valor extremo. Sendo um caso particular da distribuição log-gama generalizada, a distribuição do valor extremo satisfaz todas as condições de regularidade impostas ao modelo (3.11). Sejam então Y_{ij} , $i = 1, \dots, g$ e $j = 1, \dots, r_i$, variáveis aleatórias independentes tais que

$$Y_{ij} = x_i^t \beta + \sigma_i Z_{ij}, \quad (3.18)$$

onde Z_{ij} tem distribuição do valor extremo padrão com função de densidade de probabilidade dada por $\exp(z - e^z)$, $-\infty < z < \infty$. Nessas condições, a função log-verossimilhança é reduzida para a forma

$$L(\theta) \propto - \sum_{i=1}^g [r_i \log \sigma_i - \sum_{j=1}^{r_i} \{z_{ij} + \exp(z_{ij})\}], \quad (3.19)$$

e a matriz de informação de Fisher pode ser expressa como

$$F = \begin{bmatrix} X^t W X & (1 - \gamma) X^t W \\ (1 - \gamma) W X & \left\{ \frac{\pi^2}{6} + (1 - \gamma)^2 \right\} W \end{bmatrix},$$

onde $\gamma = 0.5772 \dots$ é a constante de Euler, X é uma matriz $g \times p$ com linhas $x_i^t = (x_{i1}, \dots, x_{ip})$, $W = \text{diag}\{w_1, w_2, \dots, w_g\}$ e $w_i = r_i \sigma_i^{-2}$, $i = 1, 2, \dots, g$. A

partir das condições (1), (2), (3) e (4) e de condições usuais de regularidade (vide Apêndice B), temos que quando $n \rightarrow \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_{\theta}^{-1}),$$

e, em particular, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_{\beta}^{-1})$ e $\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N(0, V_{\sigma}^{-1})$, onde

$$\begin{aligned} V_{\beta}^{-1} &= (X^t W_a X)^{-1} \{I_p + (1 - \gamma)^2 X^t W_a V_{\sigma}^{-1} W_a X (X^t W_a X)^{-1}\} \text{ e} \\ V_{\sigma}^{-1} &= [W_a^{1/2} \{ \frac{\pi^2}{6} I_p + (1 - \gamma)^2 (I_p - H_a) \} W_a^{1/2}]^{-1}, \end{aligned}$$

com $H_a = W_a^{1/2} X (X^t W_a X)^{-1} X^t W_a^{1/2}$. Para o caso de g populações, X reduz a uma matriz identidade de ordem g e $E(Y_{ij}) = \mu_i - \gamma \sigma_i$, onde μ_i são os parâmetros de locação, $i = 1, 2, \dots, g$. Como nF^{-1} é um estimador consistente de V_{θ}^{-1} , as variâncias assintóticas de $\hat{\mu}_i$ e $\hat{\sigma}_i$ ficam dadas por $\text{Var}(\hat{\mu}_i) = \sigma_i^2 r_i^{-1} \{1 + (1 - \gamma)^2 (6/\pi^2)\}$ e $\text{Var}(\hat{\sigma}_i) = \sigma_i^2 r_i^{-1} (6/\pi^2)$, $i = 1, \dots, g$. Ainda temos que $\text{Cov}(\hat{\mu}_i, \hat{\sigma}_j) = -(1 - \gamma) \text{Var}(\hat{\sigma}_j)$, $i = j$, e $\text{Cov}(\hat{\mu}_i, \hat{\sigma}_j) = \text{Cov}(\hat{\mu}_i, \hat{\mu}_j) = \text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j) = 0$, para os demais casos.

Quando $\sigma_i = \sigma$ é desconhecida, a matriz de informação de Fisher reduz para

$$F = \frac{1}{\sigma^2} \begin{bmatrix} X^t W X & (1 - \gamma) X^t w \\ (1 - \gamma) w^t X & n \{ \frac{\pi^2}{6} + (1 - \gamma)^2 \} \end{bmatrix},$$

onde $w = (r_1, \dots, r_g)^t$ e $W = \text{diag}(r_1, \dots, r_g)$. Consequentemente, podemos mostrar que $V_{\beta}^{-1} = \sigma^2 (X^t W_a X)^{-1} \{I_p + \frac{(1-\gamma)^2}{\sigma^2} X^t w_a V_{\sigma}^{-1} w_a^t X (X^t W_a X)^{-1}\}$ onde $W_a = \text{diag}\{a_1, \dots, a_g\}$, $w_a = \{a_1, \dots, a_g\}^t$ e

$$V_{\sigma}^{-1} = \sigma^2 \{ \frac{\pi^2}{6} + (1 - \gamma)^2 (1 - w_a^t X (X^t W_a X)^{-1} X^t w_a) \}^{-1}.$$

A seguir algumas aplicações do modelo de regressão do valor extremo são oferecidas.

3.5.1 Modelo de regressão de retas paralelas

Consideremos agora o modelo de regressão linear

$$Y_{ijk} = \alpha_i + \delta x_{ij} + \sigma Z_{ijk}, \quad (3.20)$$

onde $i = 1, \dots, g$, $j = 1, \dots, d$, $k = 1, \dots, r_{ij}$ e Z_{ijk} tem distribuição do valor extremo padrão. Esse tipo de modelo tem sido largamente aplicado em Farmacologia para estudar a relação entre potências de diferentes drogas, especialmente para respostas binárias (vide Finney, 1971; Morgan, 1992; Paula and Sen, 1995). Temos assumido aqui o mesmo número de doses para cada droga a fim de simplificar a distribuição nula assintótica de $\mathcal{W}$. A quantidade x_{ij} denota a j -ésima dose (ou log dose) correspondente à i -ésima droga. A hipótese de homogeneidade das potências é equivalente à hipótese $H_0 : \alpha_1 = \dots = \alpha_g$. Daí podemos estar interessados em testar H_0 contra hipóteses de potências ordenadas (ou interceptos ordenados).

Para o modelo dado em (3.20), o vetor paramétrico é $\beta = (\alpha_1, \dots, \alpha_g, \delta)^t$, a matriz modelo é $X = (X_1^t, \dots, X_g^t)^t$, onde X_i é a matriz $d \times (g+1)$ cuja j -ésima linha tem 1 na i -ésima posição, x_{ij} na última posição e zeros nas posições restantes. Em particular, sob as condições (1)-(4), segue de Bradley e Gart (1962) que $\frac{\sqrt{n}}{\sigma}(\alpha - \hat{\alpha}) \xrightarrow{d} N(0, V_\alpha^{-1})$ onde os elementos da matriz V_α^{-1} de dimensão $g \times g$, são expressos da seguinte forma:

$$v_{ii} = a_i^{-1} + T(\bar{m} + \epsilon_i)^2 + 6\{1 + T\bar{m}(\bar{m} + \epsilon_i)\}^2/\pi^2$$

e

$$v_{ii'} = T(\bar{m} + \epsilon_i)(\bar{m} + \epsilon_{i'}) + 6\{1 + T\bar{m}(\bar{m} + \epsilon_i)\}\{1 + T\bar{m}(\bar{m} + \epsilon_{i'})\}/\pi^2$$

para $1 \leq i' \neq i \leq g$, $a_{i.} = a_{i1} + \dots + a_{ig}$, $\epsilon_i = m_i - \bar{m}$, $m_i = \sum_{t=1}^g \frac{a_{it}}{a_{i.}} x_{it}$, $T = \left[\sum_{i=1}^g \left\{ \sum_{t=1}^d a_{it} x_{it}^2 - (\bar{m} + \epsilon_i)^2 a_{i.} \right\} \right]^{-1}$ e $\bar{m} = \sum \sum_{i,t} a_{it} x_{it}$.

Portanto, para testar hipóteses do tipo $H_0 : \Pi_2 \alpha = 0$ contra $H_1 : \Pi_2 \alpha \geq 0$, para o caso de ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop*, com pelo menos uma desigualdade estrita sob H_1 , onde Π_2 é uma matriz de constantes $q \times g$ ($q \leq g$) de posto completo, os coeficientes de correlação associados à matriz $\Sigma = \Pi_2 V_\alpha^{-1} \Pi_2'$ dependem unicamente das quantidades $\bar{m}$ e T , os quais não dependem dos grupos, e de $a_{i.}$ e ϵ_i , para $i = 1, 2, \dots, g$. Em particular, para a estrutura balanceada $\epsilon_i = 0$ para todo $i = 1, 2, \dots, g$ (vide Paula, 1996), as componentes da matriz V_α^{-1} tomam a forma

$$v_{ii} - v = a_{i.}^{-1} \text{ e } v_{ii'} - v = 0,$$

onde $v = T\bar{m}^2 + 6\{1 + T\bar{m}^2\}^2/\pi^2$. Os coeficientes de correlação associados com a matriz Σ tomam as mesmas formas dadas em (3.14), (3.16) e (3.17) (com a_i sendo substituído por $a_{i.}$), para os casos de ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop*, respectivamente. Portanto, a distribuição nula assintótica de $\mathcal{W}$ é dada nesses casos por

$$\lim_{n \rightarrow \infty} P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a_{.}) P_r\{\chi_\ell^2 \geq c\},$$

onde $a_{.} = (a_{1.}, \dots, a_{g.})^t$ são os pesos. Um estimador consistente de $a_{i.}$ é $r_{i.}/n$. Portanto, se o experimento é conduzido de tal forma que $\bar{x}_1 = \dots = \bar{x}_g$, onde $\bar{x}_i = \sum_{t=1}^d \frac{r_{it}}{r_{i.}} x_{it}$ é um estimador consistente de m_i , a distribuição nula assintótica de $\mathcal{W}$ para testar H_0 contra uma alternativa de ordem simples, ordem simples em forma de árvore, ordem unimodal ou ordem do tipo *loop* pode ser aproximada pela distribuição qui-quadrado barra. Porém, mesmo quando a quantidade de variação entre $\bar{x}_1, \dots, \bar{x}_g$ é moderada, essa aproximação ainda

é recomendada. Isso será discutido na Seção 3.6.

3.5.2 Hipóteses de ordem para os percentis

Para a distribuição do valor extremo com parâmetros de locação e escala dados por μ e σ , respectivamente, o p -ésimo percentil pode ser expresso analiticamente como $y_p = \mu + \sigma \log[-\log(1 - p)]$ (vide Lawless, 1982). Desse modo, podemos estar interessados em testar hipóteses com restrições de ordem relacionadas com y_{pi} , $i = 1, \dots, g$. Por exemplo, em g populações, podemos querer testar a hipótese $H_0 : y_{p1} = \dots = y_{pg}$, $0 < p < 1$, contra alguma hipótese ordenada. As estimativas de máxima verossimilhança para (μ_i, σ_i) , sob H_0 , satisfazem as equações $\mu_i + \sigma_i \log[-\log(1 - p)] = \mu_{i+1} + \sigma_{i+1} \log[-\log(1 - p)]$, $i = 1, \dots, g - 1$, a qual leva em particular à solução $\mu_1 = \dots = \mu_g$ e $\sigma_1 = \dots = \sigma_g$.

Como ilustração, consideremos aqui o caso de ordem simples e $g = 5$ que corresponde às hipóteses $H_0 : \Pi_2 \theta = 0$ contra $H_1 : \Pi_2 \theta \geq 0$ com pelo menos uma desigualdade estrita. Nesse caso, temos $\theta = (\mu_1, \dots, \mu_5, \sigma_1, \dots, \sigma_5)^t$

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -f_p & f_p \end{pmatrix} \text{ e}$$

$$\Pi_2 V_\theta^{-1} \Pi_2^t = [\gamma_1 + 2\gamma_2 f_p + \gamma_3 f_p^2](S + S_0),$$

onde

$$\gamma_1 = 1 + 6(1 - \gamma)^2/\pi^2, \quad \gamma_2 = -6(1 - \gamma)/\pi^2, \quad \gamma_3 = 6/\pi^2,$$

$f_p = \log[-\log(1 - p)]$, $S_0 = \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e

$$S = \begin{pmatrix} w_1^{-1} & -w_2^{-1} & 0 & 0 \\ -w_2^{-1} & w_2^{-1} & -w_3^{-1} & 0 \\ 0 & -w_3^{-1} & w_3^{-1} & -w_4^{-1} \\ 0 & 0 & -w_4^{-1} & w_4^{-1} \end{pmatrix},$$

onde novamente para simplicidade de notação denotaremos $w_i = a_i \sigma_i^{-2}$.

Sob H_0 e com a suposição $\sigma_i = \sigma$, temos por exemplo que

$$\begin{aligned} \rho_{34} &= - \left\{ \frac{a_3 a_5}{(a_4 + a_3)(a_4 + a_5)} \right\}^{1/2} \\ \rho_{23} &= - \left\{ \frac{a_2 a_4}{(a_3 + a_2)(a_3 + a_4)} \right\}^{1/2}. \end{aligned}$$

Os demais ρ 's são obtidos similarmente. Para testar $H_0 : \Pi_2 \theta = 0$ contra ordem simples em forma de árvore, temos

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -f_p & 0 & f_p & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & -f_p & 0 & 0 & f_p & 0 \\ -1 & 0 & 0 & 0 & 1 & -f_p & 0 & 0 & 0 & f_p \end{pmatrix} \text{ e}$$

$$\Pi_2 V_\theta^{-1} \Pi_2' = [\gamma_1 + 2\gamma_2 f_p + \gamma_3 f_p^2] w_1^{-1} (A + A_0),$$

onde $A_0 = w_1 \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e A é uma matriz com 1 em todas suas posições. De $\Pi_2 V_\theta^{-1} \Pi_2'$, obtemos

$$\begin{aligned} \rho_{23} &= \left\{ \frac{a_3 a_4}{(a_1 + a_3)(a_1 + a_4)} \right\}^{1/2} \text{ e} \\ \rho_{34} &= \left\{ \frac{a_4 a_5}{(a_1 + a_4)(a_1 + a_5)} \right\}^{1/2}. \end{aligned}$$

Os demais ρ 's são obtidos de forma similar.

Em geral, podemos mostrar para os casos de ordem simples e ordem simples em forma de árvore, que os coeficientes de correlação associados à matriz Σ não dependem dos percentis, sob H_0 . Eles tomam as mesmas formas dadas em (3.14) e (3.16), respectivamente. Consequentemente, a distribuição nula assintótica de $\mathcal{W}$ é também uma distribuição qui-quadrado barra.

Para testar $H_0 : \Pi_2 \theta = 0$ contra uma alternativa unimodal, temos

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -f_p & f_p & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & f_p & -f_p & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & f_p & -f_p \end{pmatrix} \text{ e}$$

$$\Pi_2 V_\theta^{-1} \Pi_2' = [\gamma_1 + 2\gamma_2 f_p + \gamma_3 f_p^2](U + U_0),$$

onde $U_0 = \text{diag}(w_2^{-1}, w_3^{-1}, w_4^{-1}, w_5^{-1})$ e

$$U = \begin{pmatrix} w_1^{-1} & -w_2^{-1} & 0 & 0 \\ -w_2^{-1} & w_2^{-1} & w_3^{-1} & 0 \\ 0 & w_3^{-1} & w_3^{-1} & -w_4^{-1} \\ 0 & 0 & -w_4^{-1} & w_4^{-1} \end{pmatrix}.$$

Sob H_0 e com a suposição $\sigma_i = \sigma$, temos por exemplo que

$$\rho_{34} = - \left\{ \frac{a_3 a_5}{(a_4 + a_3)(a_4 + a_5)} \right\}^{1/2}$$

$$\rho_{23} = \left\{ \frac{a_2 a_4}{(a_3 + a_2)(a_3 + a_4)} \right\}^{1/2}.$$

Os demais ρ 's são obtidos similarmente. Para o caso de teste de hipóteses do tipo *loop*, obtemos resultados similares. Podemos mostrar também para o caso de teste de hipóteses unimodal que os coeficientes de correlação associados à matriz Σ não dependem dos percentis, sob H_0 . Eles tomam as mesmas formas dadas em (3.17) e portanto a distribuição nula assintótica de $\mathcal{W}$ é também a

distribuição qui-quadrado barra como definida em (3.15). É importante observar aqui, que todos os resultados obtidos para teste de hipóteses de ordem para os percentis, também valem para o caso da distribuição log-gama generalizada.

3.5.3 Hipótese nula restrita

Como foi discutido na Seção 2.4 a determinação da distribuição nula assintótica menos favorável de $\mathcal{W}$ para testar hipóteses do tipo $H_0 : h_1(\theta) = 0, h_2(\theta) \geq 0$ contra $H_1 - H_0$ onde $H_1 : \theta \in R^{p+q}$, pode tornar-se complexa quando a matriz de correlação associada a $\Sigma = \Pi_2(\theta)V_\theta^{-1}\Pi_2(\theta)^t$ depender de θ sob H_0 . Embora, sob condições gerais de regularidade, a distribuição nula assintótica menos favorável seja uma mistura de distribuições do tipo qui-quadrado, a determinação da situação menos favorável irá depender do desenvolvimento de algoritmos de busca, em geral de alto custo computacional. No entanto, há pelo menos uma subclasse de hipóteses para os parâmetros em (3.18) em que a determinação da distribuição nula menos favorável não é tão complexa. Para ilustrar, suponha a seguinte classe de hipóteses :

$$H_0 : \Pi_1\beta = 0, \Pi_2\beta \geq 0 \text{ contra } H_1 - H_0, \quad (3.21)$$

onde $H_1 : \beta \in R^p$, Π_1 e Π_2 são matrizes conhecidas (não dependem de θ) de ordem $k_1 \times p$ e $(k - k_1) \times p$, respectivamente, e de posto completo. Assim, segue de Kodde e Palm (1986) e de Gourieroux e Monfort (1995) que a distribuição nula menos favorável para testar (3.21) é dada por

$$\lim_{n \rightarrow \infty} P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^q Q(q, \ell; \Sigma) P_r\{\chi_{k-\ell}^2 \geq c\}, \quad (3.22)$$

onde $c \geq 0$, $q = k - k_1$ e

$$\Sigma = \Pi_2 V_\beta^{-1} \Pi_2^t - \Pi_2 V_\beta^{-1} \Pi_1^t [\Pi_1 V_\beta^{-1} \Pi_1^t]^{-1} \Pi_1 V_\beta^{-1} \Pi_2^t.$$

Isso ocorre porque os coeficientes de correlação associados a Σ irão depender apenas de $(\sigma_1, \dots, \sigma_g)^t$ através de V_β^{-1} . Esse é um caso particular de Σ desconhecida em que a situação menos favorável ocorre quando $\Pi_1\beta = 0$ e $\Pi_2\beta = 0$. O problema que permanece é a determinação dos valores de $(\sigma_1, \dots, \sigma_g)^t$ que levam à situação menos favorável no lado direito de (3.22). Uma solução seria substituir $\sigma_1, \dots, \sigma_g$ por estimativas consistentes.

Resta, portanto, o estudo de situações de interesse prático em que a matriz de correlação associada a Σ não depende de θ sob a hipótese $\Pi_1\beta = 0$ e $\Pi_2\beta = 0$. Mostra-se facilmente para o caso de g populações que se temos em H_0 uma ordem simples, ordem simples em forma de árvore, ordem unimodal ou ordem do tipo *loop* em $\mu_1, \dots, \mu_g$ supondo $\sigma_i = \sigma$ ou em $\sigma_1, \dots, \sigma_g$ supondo $\mu_i = \mu$, as correlações associadas à matriz Σ dependem, sob $\mu_1 = \dots = \mu_g$ ou sob $\sigma_1 = \dots = \sigma_g$, dos pesos $a_1, \dots, a_g$, sendo a distribuição nula menos favorável de $\mathcal{W}$ uma qui-quadrado barra dada por

$$\lim_{n \rightarrow \infty} P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a) P_r\{\chi_{g-\ell}^2 \geq c\}. \quad (3.23)$$

Esse resultado vale também quando testarmos ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop* para os percentis $y_{p_1}, \dots, y_{p_g}$ e para os interceptos $\alpha_1, \dots, \alpha_g$ no modelo de retas paralelas sob a condição $\epsilon_i = 0$, para todo i .

3.6 Estudo da robustez da distribuição $\bar{\chi}^2$

Nesta seção estudaremos a aproximação da distribuição $\bar{\chi}_g^2$ para pesos iguais para o modelo de retas paralelas descrito na Seção 3.5.1 quando o parâmetro

de dispersão é fixo porém não constante. Esse modelo inclui como casos particulares os modelos para g populações envolvendo os parâmetros $(\mu_1, \sigma_1), \dots, (\mu_g, \sigma_g)$.

Note que m_i definido na Seção 3.5.1 faz o papel de uma média ponderada dos valores das doses no i -ésimo grupo. Logo, se definirmos $\epsilon_i = m_i - \bar{m}$, temos que ϵ_i pode ser interpretado como um tipo de afastamento do i -ésimo grupo com relação à média global. Um valor alto para $|\epsilon_i|$ indica que a dosagem média aplicada no i -ésimo grupo é maior do que a dosagem média utilizada no experimento. Portanto, uma forma de balanceamento na aplicação das doses seria assumir que $|\epsilon_1| = \dots = |\epsilon_g|$. Note que os elementos da matriz V_α^{-1} tal que $\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, V_\alpha^{-1})$ podem ser expressos, quando $\sigma_1^2, \dots, \sigma_g^2$ são fixados, por

$$v_{ii} = \sigma_i^2 a_{i.}^{-1} + \frac{(\bar{m} + \epsilon_i)^2}{T}$$

e

$$v_{ii'} = \frac{(m_i + \epsilon_i)(\bar{m} + \epsilon_{i'})}{T},$$

onde $i \neq i' = 1, \dots, g$ e

$$T = \sum_{i=1}^g \left\{ \sigma_i^2 \sum_{j=1}^d a_{ij} x_{ij}^2 - a_{i.}^{-1} \left(\sum_{j=1}^d a_{ij} x_{ij} \right)^2 \right\}.$$

Se assumimos que as variâncias são homogêneas, $\sigma_1^2 = \dots = \sigma_g^2 = \sigma^2$, temos que

$$\lim_{\epsilon_i \rightarrow 0} v_{ii} = \frac{\sigma^2}{a_{i.}} + \frac{\bar{m}}{T} \text{ e } \lim_{\epsilon_i \rightarrow 0} v_{ii'} = \frac{\bar{m}}{T},$$

onde $i \neq i' = 1, \dots, g$. Daí, mostra-se facilmente que as correlações associadas à matriz $\Sigma = \Pi_2 V_\alpha^{-1} \Pi_2^t$ para testar $H_0 : \alpha_1 = \dots = \alpha_g$ contra uma alternativa de ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop* coincidem com as correlações do caso de g populações com a_i

sendo substituído por a_i . Portanto, em particular para os casos de ordem simples e ordem simples em forma de árvore, se os pesos são iguais, $a_1 = \dots = a_g$, a distribuição nula assintótica de $\mathcal{W}$ fica, respectivamente, dada por

$$\lim_{n \rightarrow \infty} P_r \{ \mathcal{W} \geq c \} = \sum_{\ell=1}^g P_s(\ell, g) P_r \{ \chi_\ell^2 \geq c \}$$

e

$$\lim_{n \rightarrow \infty} P_r \{ \mathcal{W} \geq c \} = \sum_{\ell=1}^g P_T(\ell, g) P_r \{ \chi_\ell^2 \geq c \},$$

onde $P_s(\ell, g)$ e $P_T(\ell, g)$, $\ell = 1, \dots, g$, são as probabilidades de nível para pesos iguais (vide Apêndice A).

Seja $e_{01}(\alpha)$ o valor crítico para um nível de significância α de uma distribuição qui-quadrado barra com pesos iguais. Definimos

$$C_{01}(t; \Sigma) = \sum_{\ell=0}^{g-1} Q(g-1, \ell; \Sigma) P_r \{ \chi_\ell^2 \geq t \},$$

onde $t = e_{01}(\alpha)$. O estudo da distribuição de $C_{01}(t; \Sigma)$, à medida que as quantidades $|\epsilon_i|$, σ_i^2 e a_i variam, $i = 1, \dots, g$, pode indicar situações em que a distribuição qui-quadrado barra com pesos iguais seria recomendada para aproximar a distribuição nula assintótica da estatística $\mathcal{W}$. Esse tipo de estudo foi inicialmente proposto por Robertson e Wright (1983) para estudar a aproximação da distribuição $\bar{\chi}_g^2$ com pesos $P(\ell, g; a)$ pela mesma distribuição com pesos $P_s(\ell, g)$. A conclusão dos mesmos é recomendar a aproximação quando a razão entre o maior peso e o menor peso não for superior a 3.5. A aproximação não se mostrou muito satisfatória quando os pesos tomam a forma $\cup$ ou a forma $\cap$. Resultados similares foram obtidos por Wright e Tran (1985) para o caso de hipótese simples em forma de árvore e por Paula (1996) para alguns modelos lineares generalizados.

Para medir a variação dos ϵ'_i s, definimos a quantidade $\epsilon_{\max} = \max\{|\epsilon_1|, \dots, |\epsilon_g|\}$ e estudamos o comportamento de $C_{01}(\epsilon_{01}(\alpha); \Sigma)$ à medida que $\epsilon_{\max}$ varia para valores fixos de $(a_1, \dots, a_g)$, $(\sigma_1^2, \dots, \sigma_g^2)$ e α . Adotamos $g = 4$, uma vez que para $g \geq 5$ não há formas fechadas para os coeficientes $Q(g-1, \ell; \Sigma)$'s, o que aumentaria substancialmente o custo computacional. Para cada caso geramos 10000 conjuntos de valores para $(\epsilon_1, \epsilon_2, \epsilon_3)$, supondo ϵ_i normalmente distribuído com média zero e variância 1. Note que $\sum_{i=0}^g \epsilon_i = 0$, logo ϵ_4 sai da relação $\epsilon_4 = -\epsilon_1 - \epsilon_2 - \epsilon_3$. Para cada conjunto gerado computamos $\epsilon_{\max}$ e então a probabilidade $C_{01}(\epsilon_{01}(\alpha); \Sigma)$. Os 10000 valores obtidos para C_{01} foram classificados de acordo com a distribuição de frequências dada na terceira coluna das Tabelas C1-C16 (vide Apêndice C), cujos resultados são resumidos na Tabela 3.1.

Tabela 3.1
Desempenho das aproximações.

Caso	Pesos	Dispersões			
		Iguais	Diferentes	Tipo- $\cap$	Tipo- $\cup$
Ordem simples	Iguais	*	*	-	+
	Diferentes	*	*	-	+
	Tipo- $\cap$	+	+	*	+
	Tipo- $\cup$	-	-	-	*
Ordem Simples em forma de árvore	Iguais	*	-	-	*
	Diferentes	*	*	*	+
	Tipo- $\cap$	*	*	*	+
	Tipo- $\cup$	*	-	-	*

- : subestima, * : satisfatório, + : superestima.

Como podemos observar pela Tabela 3.1 para o caso de ordem simples, dos 16 casos estudados envolvendo variações para os pesos $a_1, \dots, a_4$ e para as

dispersões $\sigma_1^2, \dots, \sigma_4^2$, apenas 6 casos podem ser considerados satisfatórios. Esses casos incluem todas as situações em que os pesos e os parâmetros de escala têm a mesma forma (4 casos) e ainda as situações de (pesos iguais, dispersões diferentes) e (dispersões iguais, pesos diferentes). Para os demais casos, embora a aproximação ainda seja razoável, sempre há uma tendência da probabilidade $C_{01}(t; \Sigma)$ ser subestimada ou superestimada pela aproximação $\sum_{\ell=1}^g P_s(\ell, g) P_r\{\chi_\ell^2 \geq t\}$. Assim, podemos recomendar o uso da aproximação de pesos iguais sempre que a razão entre o maior e o menor tamanho amostral não exceder 2.0 e a razão entre a maior e a menor dispersão não for superior a 3.0. Já para o caso de ordem simples em forma de árvore há um número maior de casos satisfatórios (10 casos) dentre os 16 estudados, o que mostra uma melhor adequação da aproximação $\sum_{\ell=1}^g P_s(\ell, g) P_r\{\chi_\ell^2 \geq t\}$ para a probabilidade $C_{01}(t; \Sigma)$. Aqui, as recomendações feitas para o caso de ordem simples são também estendidas para todos os casos satisfatórios.

3.7 Aplicações

3.7.1 Desempenho de turbinas

A Tabela 3.2 descreve os resultados de um experimento planejado para comparar o desempenho de cinco tipos diferentes de instrumentos de apoio para turbinas de alta velocidade (McCool, 1979).

Tabela 3.2
*Tempo de falhas de turbinas**.

Componentes				
Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
3.03	3.19	3.46	5.88	6.43
5.53	4.26	5.22	6.74	9.97
5.60	4.47	5.69	6.90	10.39
9.30	4.53	6.54	6.98	13.55
9.92	4.67	9.16	7.21	14.45
12.51	4.69	9.40	8.14	14.72
12.95	5.78	10.19	8.59	16.81
15.21	6.79	10.71	9.80	18.39
16.04	9.37	12.58	12.28	20.84
16.84	12.75	13.41	25.46	21.51

* Fonte: Lawless, 1982, p. 201

Foi observado o tempo de falha para dez réplicas de cada tipo. Seja Y_{ij} o logaritmo do tempo de falha correspondente à j -ésima réplica da i -ésima componente, $i = 1, \dots, 5$ e $j = 1, \dots, 10$. Suponhamos que Y_{ij} segue uma distribuição do valor extremo com parâmetro de locação μ_i e parâmetro de escala σ_i . Nosso interesse aqui é ilustrar a aplicação da metodologia desenvolvida testando algumas hipóteses ordenadas relacionadas com os parâmetros (μ_i, σ_i) . A Tabela 3.3 apresenta as estimativas de máxima verossimilhança (desvio padrão assintótico) para esses parâmetros. A estatística da razão de verossimilhança para testar, por exemplo, a hipótese H_0 contra H_1 envolvendo (μ_i, σ_i) pode ser expressa na forma

$$\mathcal{W} = 2 \sum_{i=1}^g [-r_i \log(\hat{\sigma}_{i1}/\hat{\sigma}_{i0}) + \sum_{j=1}^{r_i} \{(\hat{z}_{ij1} - \hat{z}_{ij0}) - (e^{\hat{z}_{ij1}} - e^{\hat{z}_{ij0}})\}],$$

onde $(\hat{\sigma}_{i0}, \hat{z}_{ij0})$ e $(\hat{\sigma}_{i1}, \hat{z}_{ij1})$ denotam as estimativas de máxima verossimilhança para (σ_i, z_{ij}) sob H_0 e H_1 , respectivamente.

Tabela 3.3

Estimativas de máxima verossimilhança (desvio padrão assintótico) para os parâmetros de locação e escala e o 10o. percentil.

Componente	Parâmetros		
	Locação	Escala	10o.Percentil
Tipo I	2.490(0.207)	0.386(0.153)	1.620(0.454)
Tipo II	1.926(0.219)	0.431(0.162)	0.956(0.480)
Tipo III	2.271(0.188)	0.319(0.139)	1.552(0.413)
Tipo IV	2.410(0.239)	0.516(0.177)	1.250(0.525)
Tipo V	2.794(0.174)	0.274(0.129)	2.170(0.382)

Inicialmente, vamos testar a hipótese $H_0 : \sigma_1 = \dots = \sigma_5$ contra H_1 : pelo dois σ_i 's diferentes. A estatística da razão de verossimilhança $\mathcal{W} = 4.34$ ($P \cong 0.456$), indica pela não rejeição de H_0 . A partir dessa conclusão desejamos testar, como sugerido pelos dados, a hipótese $H_0 : \mu_1 = \mu_4, \mu_2 = \mu_3, \mu_5 \geq [\mu_1, \dots, \mu_4]$ contra $H_1 - H_0$, onde $H_1 : \mu_i \in R, i = 1, \dots, 5$. As estimativas de máxima verossimilhança sob H_0 e H_1 são, respectivamente, $\hat{\mu}_{10} = \hat{\mu}_{40} = 2.483(0.1509)$, $\hat{\mu}_{20} = \hat{\mu}_{30} = 2.115(0.1509)$, $\hat{\mu}_{50} = 2.753(0.2134)$, $\hat{\sigma} = 0.411(0.071)$ e $\hat{\mu}_{11} = 2.484(0.2114)$, $\hat{\mu}_{21} = 1.941(0.2114)$, $\hat{\mu}_{31} = 2.242(0.2114)$, $\hat{\mu}_{41} = 2.491(0.2114)$, $\hat{\mu}_{51} = 2.754(0.2114)$ e $\hat{\sigma} = 0.403(0.070)$, o qual leva para o valor da estatística $\mathcal{W} = 2.693$. Como mencionado antes, a distribuição nula assintótica de $\mathcal{W}$ é dada por (3.23) com $g - \ell$ substituído por $7 - \ell$ uma vez que temos duas restrições por igualdade em H_0 . Além disso, temos um experimento balanceado ($r_1 = \dots = r_5$) e portanto podemos substituir os $P(\ell, g; a)$'s em (3.23) pelas probabilidades $P_T(\ell, g)$'s de pesos iguais correspondentes ao caso de ordem simples em forma de árvore (vide Robertsom et al., 1988). Podemos calcular o nível descritivo correspondente a $\mathcal{W} = 2.693$, aproximadamente, como

$$P \cong \sum_{\ell=1}^5 P_T(\ell, 5) P_r\{\chi_{7-\ell}^2 \geq c\}$$

$$\begin{aligned}
&= P_T(5, 5)Pr(\chi_2^2 \geq 2.693) + P_T(4, 5)Pr(\chi_3^2 \geq 2.693) \\
&\quad + P_T(3, 5)Pr(\chi_4^2 \geq 2.693) + P_T(2, 5)Pr(\chi_5^2 \geq 2.693) \\
&\quad + P_T(1, 5)Pr(\chi_6^2 \geq 2.693) \\
&= 0.2000 \times Pr\{\chi_2^2 \geq 2.693\} + 0.41226 \times Pr\{\chi_3^2 \geq 2.693\} \\
&\quad + 0.29022 \times Pr\{\chi_4^2 \geq 2.693\} + 0.08774 \times Pr\{\chi_5^2 \geq 2.693\} \\
&\quad + 0.00978 \times Pr\{\chi_6^2 \geq 2.693\} \\
&= 0.485
\end{aligned}$$

indicando forte evidência em favor de H_0 .

3.7.2 Experimento com virus

A Tabela 3.4 contém parte dos resultados de um experimento que envolveu uma preparação padrão do virus da *poliomielitis* e quatro preparações novas com o virus (Finney, 1978, p.441). Cinco camundongos machos e cinco fêmeas foram inoculados com cinco doses da preparação usual (padrão); o mesmo número de animais foi inoculado com duas doses de cada preparação nova. Com o objetivo de aplicar os resultados apresentados na Seção 3.5.1, onde é assumida uma estrutura balanceada, temos considerado somente as doses 2 e 4 para todas as preparações. A resposta considerada é o número de dias decorridos antes de cada camundongo ficar doente. Dois camundongos, que ficaram aparentemente bem após dezesseis dias, receberam um valor arbitrário para a resposta, 20 dias.

Finney apresenta uma interessante discussão sobre possíveis métodos de análise e particularmente sugere o uso da transformação $Y = \log(\text{No. de dias} - 0.5)$.

Como essa transformação leva o tempo (No. de dias) para a reta real, parece ser razoável supor que a resposta segue uma distribuição do valor extremo. Assim, um possível modelo de regressão seria dado por

$$Y_{ijk\ell} = \alpha_i + \delta x_{ij} + \gamma_\ell + \sigma Z_{ijk\ell},$$

para $i, k = 1, \dots, 5$ e $\ell, j = 1, 2$, onde σ é assumido desconhecido, $Z_{ijk\ell}$ tem distribuição do valor extremo padrão, x_{ij} denota a j -ésima dose correspondente à i -ésima preparação (1:padrão), γ_2 representa o efeito sexo (macho com relação à fêmea) e $\gamma_1 = 0$. Um teste preliminar de $H_0 : \gamma_2 = 0$ contra $H_1 : \gamma_2 \neq 0$ fornece $P = 0.833$, indicando pela não rejeição da hipótese nula.

Table 3.4*Resposta de camundongos à inoculação com o vírus da poliomyelitis.¹*

Preparação (Number)	Dilution (as $\log_{10}$)	Sexo	Número de dias
1	2	M	3, 3, 3, 3, 4
		F	3, 3, 3, 3, 4
	4	M	4, 5, 5, 5, 6
		F	4, 6, 6, 6, 8
2	2	M	3, 4, 5, 5, 6
		F	3, 3, 3, 3, 5
	4	M	5, 6, 7, 8, 9
		F	5, 5, 6, 7, 20
3	2	M	3, 3, 3, 3, 4
		F	3, 3, 3, 3, 3
	4	M	4, 6, 6, 6, 7
		F	5, 6, 6, 6, 6
4	2	M	3, 4, 5, 5, 5
		F	4, 4, 4, 4, 4
	4	M	6, 6, 7, 8, 20
		F	6, 6, 7, 10, 10
5	2	M	3, 3, 3, 3, 3
		F	3, 3, 3, 3, 4
	4	M	4, 4, 5, 6, 6
		F	4, 4, 5, 5, 5

Assumimos então o modelo de regressão de retas paralelas

$$Y_{ijk} = \alpha_i + \delta x_{ij} + \sigma Z_{ijk},$$

onde $k = 1, 2, \dots, 10$. Ao ajustar o modelo através, por exemplo, do programa dado em Aitkin et al., (1989), obtemos as estimativas de máxima verossimilhança (desvio padrão assintótico) $\hat{\alpha}_1 = 0.185(0.111)$, $\hat{\alpha}_2 = 0.696(0.125)$,

$\hat{\alpha}_3 = 0.167(0.114)$, $\hat{\alpha}_4 = 0.711(0.126)$, $\hat{\alpha}_5 = 0.094(0.106)$, $\hat{\delta} = 0.388(0.032)$ e $\hat{\sigma} = 0.302(0.024)$. O desvio $D(y, \hat{\mu}) = \sum_{i=1}^5 \sum_{j=1}^2 \sum_{k=1}^{10} (\exp \hat{z}_{ij} - \hat{z}_{ij} - 1)$ (vide Aitkin, 1989) o qual pode ser usado como indicador da qualidade do ajuste, deu $D(y, \hat{\mu}) = 92.89$ (94 graus de liberdade) indicando que o ajuste não é inadequado.

Como não temos qualquer tipo de informação a priori sobre restrições para as potências das drogas, a aplicação de um teste restrito será feito assumindo na hipótese nula alguma tendência que faça sentido no problema. Entre as possíveis hipóteses nulas restritas podemos estar interessados, por exemplo, em testar $H_0 : \alpha_1 \leq [\alpha_2, \dots, \alpha_5]$ contra uma hipótese não restrita. Desejamos verificar nesse caso, se todas as preparações são pelo menos tão eficientes quanto a preparação padrão. Como o interesse se concentra na hipótese nula, similarmente ao teste de qualidade do ajuste, devemos ter cuidado ao especificarmos o nível de significância do teste. A estatística da razão de verossimilhança para testar H_0 contra uma hipótese restrita pode ser expressa como

$$\mathcal{W} = 2 \sum_{i=1}^g [-r_i \log(\hat{\sigma}_{i1}/\hat{\sigma}_{i0}) + \sum_{j=1}^{r_i} \{(\hat{z}_{ij1} - \hat{z}_{ij0}) - (e^{\hat{z}_{ij1}} - e^{\hat{z}_{ij0}})\}],$$

onde $(\hat{\sigma}_{i0}, \hat{z}_{ij0})$ e $(\hat{\sigma}_{i1}, \hat{z}_{ij1})$ denotam as estimativas de máxima verossimilhança para (σ_i, z_{ij}) sob H_0 e H_1 , respectivamente.

As estimativas de máxima verossimilhança sob a hipótese nula foram obtidas pelo procedimento descrito em McDonald e Diamond (1983). Essas estimativas são : $\hat{\alpha}_1 = \hat{\alpha}_5 = 0.138(0.098)$, $\hat{\alpha}_2 = 0.689(0.125)$, $\hat{\alpha}_3 = 0.162(0.114)$, $\hat{\alpha}_4 = 0.704(0.127)$, $\hat{\delta} = 0.390(0.032)$ e $\hat{\sigma} = 0.303(0.024)$. Portanto, a estatística da razão da verossimilhança fica igual a $\mathcal{W} = 0.300$. A distribuição nula

assintótica de $\mathcal{W}$ pode ser bem aproximada pela distribuição qui-quadrado barra dada em (3.23) com $P(\ell, 5; a)$ substituído por $P_T(\ell, 5)$, devido à estrutura balanceada, isto é, $r_i = \dots = r_5 = 20$. Usando os valores tabelados para $P_T(\ell, 5)$ dados no Apêndice de Robertsom et al.(1988), encontramos o nível descritivo

$$P \cong 0 + 0.41226 \times P_r\{\chi_1^2 \geq 0.3\} + 0.29022 \times P_r\{\chi_2^2 \geq .3\} \\ + 0.08774 \times P_r\{\chi_3^2 \geq 0.3\} + 0.00978 \times P_r\{\chi_4^2 \geq .3\} = 0.584,$$

que indica na direção da hipótese nula. Conforme anotado por Finney (1978), a situação em que há uma maior quantidade de doses para a preparação padrão é recomendada quando pouco é conhecido acerca das potências das preparações novas. Nesse caso, segue da Seção 3.5.3 que a distribuição nula assintótica de $\mathcal{W}$ é dada por (3.22), uma vez que os coeficientes de correlação associados a Σ não dependem de $\alpha_1, \dots, \alpha_5$. No entanto, a dificuldade de computação permanece devido à necessidade agora do cálculo do nível de probabilidade $Q(4, 4; \Sigma)$ para o qual não há forma fechada.

Capítulo 4

Hipóteses restritas no modelo de regressão estrutural

4.1 Introdução

O principal objetivo deste capítulo é fazer inferências acerca do parâmetro β_i , também conhecido como coeficiente angular, em modelos de regressão estrutural com erros nas variáveis. Consideramos g populações, tal que para a população i temos uma variável aleatória bivariada (y_i, x_i) com

$$y_{ij} = \alpha_i + \beta_i x_{ij}, \quad (4.1)$$

$j = 1, \dots, r_i$ e $i = 1, \dots, g$. Os parâmetros α_i e β_i , $i = 1, \dots, g$, podem ser estimados e x_{ij} e y_{ij} não são observados diretamente. Podemos observar unicamente

$$Y_{ij} = y_{ij} + e_{ij} \text{ e} \quad (4.2)$$

$$X_{ij} = x_{ij} + u_{ij}, \quad (4.3)$$

com a suposição usual

$$\begin{pmatrix} e_{ij} \\ u_{ij} \\ x_{ij} \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 0 \\ \mu_{x_i} \end{pmatrix}; \begin{pmatrix} \sigma_{e_i}^2 & 0 & 0 \\ 0 & \sigma_{u_i}^2 & 0 \\ 0 & 0 & \sigma_{x_i}^2 \end{pmatrix}\right), \quad (4.4)$$

$j = 1, \dots, r_i$ e $i = 1, \dots, g$. A idéia principal implícita nas equações (4.1), (4.2) e (4.3) é que (x_{ij}, y_{ij}) não é observado e a estimação tem que ser baseada em (Y_{ij}, X_{ij}) , o qual é observado, $j = 1, \dots, r_i$ e $i = 1, \dots, g$. Sabe-se que quando ambos $\sigma_{e_i}^2$ e $\sigma_{u_i}^2$ são desconhecidos β_i não é identificável. A suposição comum para contornar esse problema é tomar a razão de variâncias $\lambda_{ei} = \frac{\sigma_{e_i}^2}{\sigma_{u_i}^2}$ ou o coeficiente $k_{xi} = \frac{\sigma_{x_i}^2}{\sigma_{x_i}^2 + \sigma_{u_i}^2}$ como conhecidos. Notamos que podemos tomar $\lambda_{ei} = 1$, $i = 1, \dots, g$, sem perda de generalidade. Porém, k_{xi} é tipicamente conhecido como o coeficiente de confiabilidade (Fuller, 1987). É claro que supor k_{xi} conhecido é equivalente a supor λ_{xi} conhecido, onde $\lambda_{xi} = \frac{\sigma_{x_i}^2}{\sigma_{u_i}^2}$.

Neste capítulo discutimos a parametrização ortogonal para ambas as suposições e sua estimação por máxima verossimilhança. Apresentamos também o comportamento de $\frac{I_{Fi}}{n}$, quando $n \rightarrow \infty$, com I_{Fi} sendo a matriz de informação sob o modelo estrutural. Apresentamos uma revisão da estimação por máxima verossimilhança para o modelo especificado pelas suposições (4.1)-(4.4). Discutimos também a distribuição nula assintótica de três estatísticas assintoticamente equivalentes sob hipóteses gerais do tipo *one-sided* para ambos os casos, λ_{xi} e λ_{ei} conhecidas, $i = 1, \dots, g$. Situações particulares onde a distribuição nula assintótica reduz para a conhecida distribuição qui-quadrado barra são apresentadas. Finalmente, estudos de simulação ilustram o comportamento dos testes estatísticos sob a hipótese nula (vide também Rojas, Paula e Bolfarine, 1995).

4.2 Parametrização ortogonal

Notamos que o modelo (4.1)-(4.4) pode ser escrito como

$$Z_{ij} = g_{ij} + \epsilon_{ij}, \quad (4.5)$$

onde

$$Z_{ij} = \begin{pmatrix} Y_{ij} \\ X_{ij} \end{pmatrix}, \quad g_{ij} = g(x_{ij}) = \begin{pmatrix} \alpha_i + \beta_i x_{ij} \\ x_{ij} \end{pmatrix} \quad \text{e} \quad \epsilon_{ij} = \begin{pmatrix} e_{ij} \\ u_{ij} \end{pmatrix},$$

$i = 1, \dots, g$ e $j = 1, \dots, r_i$. Assim, Z_{ij} , $i = 1, \dots, g$ e $j = 1, \dots, r_i$, são independentes e identicamente distribuídos com

$$Z_{ij} \sim N_2(\mu_i; \Sigma_i),$$

onde

$$\mu_i = E(Z_{ij}) = \begin{pmatrix} \mu_{Y_i} \\ \mu_{X_i} \end{pmatrix} = \begin{pmatrix} \alpha_i + \beta_i \mu_{x_i} \\ \mu_{x_i} \end{pmatrix} \quad (4.6)$$

e

$$\begin{aligned} \Sigma_i &= \text{Cov}[Z_{ij}] \\ &= \begin{cases} \begin{pmatrix} \lambda_{xi} \beta_i^2 \sigma_{ui}^2 + \sigma_{\epsilon i}^2 & \lambda_{xi} \beta_i \sigma_{ui}^2 \\ \lambda_{xi} \beta_i \sigma_{ui}^2 & (\lambda_{xi} + 1) \sigma_{ui}^2 \end{pmatrix}, & \text{se } \lambda_{xi} \text{ é conhecida} \\ \begin{pmatrix} \beta_i^2 \sigma_{xi}^2 + \lambda_{\epsilon i} \sigma_{ui}^2 & \beta_i \sigma_{xi}^2 \\ \beta_i \sigma_{xi}^2 & \sigma_{xi}^2 + \sigma_{ui}^2 \end{pmatrix}, & \text{se } \lambda_{\epsilon i} \text{ é conhecida.} \end{cases} \end{aligned} \quad (4.7)$$

Também pode ser mostrado que

$$|\Sigma_i| = \begin{cases} (\lambda_{xi} \beta_i^2 \sigma_{ui}^2 + (\lambda_{xi} + 1) \sigma_{\epsilon i}^2) \sigma_{ui}^2, & \text{se } \lambda_{xi} \text{ é conhecida} \\ (\lambda_{\epsilon i} \sigma_{ui}^2 + (\beta_i^2 + \lambda_{\epsilon i}) \sigma_{xi}^2) \sigma_{ui}^2, & \text{se } \lambda_{\epsilon i} \text{ é conhecida.} \end{cases} \quad (4.8)$$

Quando

$$\theta_i = \begin{cases} (\alpha_i, \mu_{xi}, \sigma_{\epsilon i}^2, \sigma_{ui}^2, \beta_i)^t, & \text{se } \lambda_{xi} \text{ é conhecida} \\ (\alpha_i, \mu_{xi}, \sigma_{xi}^2, \sigma_{ui}^2, \beta_i)^t, & \text{se } \lambda_{\epsilon i} \text{ é conhecida} \end{cases} \quad (4.9)$$

é o vetor de parâmetros, temos que a função log-verossimilhança pode ser expressa na forma

$$L(\theta) \propto -\frac{1}{2} \sum_{i=1}^g r_i \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^g \sum_{j=1}^{r_i} (Z_{ij} - \mu_i)^t \Sigma_i^{-1} (Z_{ij} - \mu_i), \quad (4.10)$$

onde μ_i e Σ_i são dados em (4.6) e (4.7), respectivamente. Podemos mostrar que a matriz de informação correspondente à log-verossimilhança (4.10) não é diagonal e portanto não é simples de ser obtida (Arellano-Valle e Bolfarine, 1995). Isso dificulta em grandes amostras fazer inferências para β_i , particularmente testes estatísticos. Uma maneira para resolver essa dificuldade é considerar uma transformação ortogonal de θ_i , isto é, transformar θ_i em $\phi_i = (\phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i}, \beta_i)^t$ tal que ϕ_{ij} , $i = 1, 2, 3, 4$, sejam ortogonais a β_i , $i = 1, \dots, g$. Quando λ_{ei} é conhecida a parametrização ortogonal é dada em Wong (1989) e quando λ_{xi} é conhecida a solução é dada em Bolfarine e Cordani (1993).

A solução apresentada em Wong (1989) e Bolfarine e Cordani (1993) pode ser escrita como

$$\begin{aligned} \phi_{1i} &= \alpha_i + \beta_i \mu_{xi}, \\ \phi_{2i} &= \mu_{xi}, \\ \phi_{4i} &= \sigma_{ui}^2 \text{ e} \end{aligned} \quad (4.11)$$

$$\phi_{3i} = \begin{cases} \lambda_{xi} \beta_i^2 \sigma_{ui}^2 + (\lambda_{xi} + 1) \sigma_{ei}^2, & \text{se } \lambda_{xi} \text{ é conhecida} \\ (\beta_i^2 + \lambda_{ei}) \sigma_{xi}^2 + \lambda_{ei} \sigma_{ui}^2, & \text{se } \lambda_{ei} \text{ é conhecida,} \end{cases} \quad (4.12)$$

$i = 1, \dots, g$. Considerando a parametrização $\phi_i = (\phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i}, \beta_i)^t$ dada acima, temos que

$$\mu_i = \mu_i(\phi) = \begin{pmatrix} \phi_{1i} \\ \phi_{2i} \end{pmatrix}$$

e

$$\begin{aligned}
\Sigma_i &= \Sigma_i(\phi) \\
&= \begin{cases} (\lambda_{xi} + 1)^{-1} \begin{pmatrix} \phi_{3i} + (\lambda_{xi}\beta_i)^2\phi_{4i} & (\lambda_{xi} + 1)\lambda_{xi}\beta_i\phi_{4i} \\ (\lambda_{xi} + 1)\lambda_{xi}\beta_i\phi_{4i} & (\lambda_{xi} + 1)^2\phi_{4i} \end{pmatrix}, & \text{se } \lambda_{xi} \text{ é conhecida} \\ (\beta_i^2 + \lambda_{ei})^{-1} \begin{pmatrix} \beta_i^2\phi_{3i} + \lambda_{ei}^2\phi_{4i} & \beta_i(\phi_{3i} - \lambda_{ei}\phi_{4i}) \\ \beta_i(\phi_{3i} - \lambda_{ei}\phi_{4i}) & \phi_{3i} + \beta_i^2\phi_{4i} \end{pmatrix}, & \text{se } \lambda_{ei} \text{ é conhecida,} \end{cases} \quad (4.13)
\end{aligned}$$

$i = 1, \dots, g$. Notemos que, em ambos os casos

$$|\Sigma_i| = \phi_{3i}\phi_{4i}, \quad (4.14)$$

$i = 1, \dots, g$.

4.3 Estimadores de máxima verossimilhança e matriz de informação

Diferenciando a log-verossimilhança (4.10) dada em termos dos parâmetros ortogonais ϕ e igualando as derivadas a zero, obtemos os estimadores de máxima verossimilhança os quais podem ser escritos, quando λ_{xi} é conhecida como

$$\begin{aligned}
\hat{\phi}_{3i} &= (\lambda_{xi} + 1)S_{YYi} - 2(\lambda_{xi}\hat{\beta}_i)S_{YXi} + (\lambda_{xi}\hat{\beta}_i^2)\hat{\phi}_{4i}, \\
\hat{\phi}_{4i} &= \frac{S_{XXi}}{\lambda_{xi} + 1}, \\
\hat{\beta}_i &= \left(\frac{\lambda_{xi} + 1}{\lambda_{xi}}\right)\frac{S_{YXi}}{S_{XXi}}.
\end{aligned}$$

$i = 1, \dots, g$, onde $S_{XXi} = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/n_i$, $S_{YYi} = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/n_i$, $S_{YXi} = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(X_{ij} - \bar{X}_i)/n_i$, $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ e $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$, $i = 1, \dots, g$.

Substituindo $\hat{\phi}_{4i}$ e $\hat{\beta}_i$ em $\hat{\phi}_{3i}$, temos que

$$\hat{\phi}_{3i} = (\lambda_{xi} + 1)S_{YY.Xi},$$

onde

$$S_{YY.Xi} = S_{YYi} - S_{XXi}^{-1}S_{YXi}^2 = S_{YYi}(1 - r_{YXi}^2)$$

e

$$r_{YXi} = \frac{S_{YXi}}{(S_{YYi}S_{XXi})^{1/2}},$$

$i = 1, \dots, g$. Quando λ_{ei} é conhecida segue que

$$\begin{aligned}\hat{\phi}_{3i} &= \frac{\hat{\beta}_i^2 S_{YYi} + 2\lambda_{ei}\hat{\beta}_i S_{YXi} + \lambda_{ei}^2 S_{XXi}}{\hat{\beta}_i^2 + \lambda_{ei}}, \\ \hat{\phi}_{4i} &= \frac{S_{YYi} - 2\hat{\beta}_i S_{YXi} + \hat{\beta}_i^2 S_{XXi}}{\hat{\beta}_i^2 + \lambda_{ei}} \text{ e} \\ \hat{\beta}_i &= \left(\frac{S_{YYi} - \lambda_{ei} S_{XXi}}{2S_{YXi}} \right) + \left\{ \left(\frac{S_{YYi} - \lambda_{ei} S_{XXi}}{2S_{YXi}} \right)^2 + \lambda_{ei} \right\}^{1/2}.\end{aligned}$$

Podemos mostrar que o estimador de máxima verossimilhança de $\phi_{Si} = (\phi_{3i}, \phi_{4i}, \beta_i)^t$ é dado pela solução da equação

$$\Sigma_i(\hat{\phi}_{Si}) = \Sigma_i,$$

onde $\Sigma_i(\hat{\phi}_{Si})$ é dada em (4.13), com ϕ_{Si} substituída por $\hat{\phi}_{Si}$, $i = 1, \dots, g$. Os outros estimadores podem também ter expressões alternativas como, por exemplo

$$\hat{\phi}_{3i} = \begin{cases} (\lambda_{xi} + 1)S_{YYi} - (\lambda_{xi}\hat{\beta}_i)^2\hat{\phi}_{4i}, & \text{se } \lambda_{xi} \text{ é conhecida} \\ \lambda_{ei}S_{YYi} - \hat{\beta}_i S_{YXi}, & \text{se } \lambda_{ei} \text{ é conhecida.} \end{cases}$$

Notemos também que

$$\hat{\phi}_{3i}\hat{\phi}_{4i} = |\hat{\Sigma}_i| = S_{YY.Xi}S_{XXi}, \quad (4.15)$$

$i = 1, \dots, g$. Não é difícil mostrar que a matriz de informação correspondente à subpopulação i pode ser escrita como (Arellano-Valle e Bolfarine, 1995)

$$K_i = \begin{pmatrix} K_{Li} & 0 \\ 0 & K_{Si} \end{pmatrix},$$

isto é, K_i é uma matriz bloco diagonal com

$$K_{Li} = n_i \Sigma_i^{-1}$$

e

$$K_{Si} = n_i \begin{pmatrix} \frac{1}{2\phi_{3i}^2} & 0 & 0 \\ 0 & \frac{1}{2\phi_{4i}^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_{\beta i}^2} \end{pmatrix},$$

onde

$$\sigma_{\beta i}^2 = \begin{cases} \phi_{3i}/\lambda_{xi}^2\phi_{4i}, & \text{se } \lambda_{xi} \text{ é conhecida} \\ [(\beta_i^2 + \lambda_{ei})/(\phi_{3i} - \lambda_{ei}\phi_{4i})]^2\phi_{3i}\phi_{4i}, & \text{se } \lambda_{ei} \text{ é conhecida,} \end{cases}$$

$i = 1, \dots, g$. Suponhamos agora que, $n_i/n \rightarrow a_i > 0$, $i = 1, \dots, g$. Segue então que

$$K_i \longrightarrow J_i = a_i \begin{pmatrix} \Sigma_i^{-1} & 0 & 0 & 0 \\ 0 & \frac{1}{2\phi_{3i}^2} & 0 & 0 \\ 0 & 0 & \frac{1}{2\phi_{4i}^2} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_{\beta i}^2} \end{pmatrix}.$$

$i = 1, \dots, g$. É facilmente visto que a matriz J_i é positiva definida, $i = 1, \dots, g$. Assim, para o caso de g populações, sob condições de regularidade (vide Bradley e Gart, 1962) temos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta^{-1}),$$

com $\theta^t = (\theta_1, \dots, \theta_g)$, $\theta_i^t = (\phi_i, \beta_i)$ e $V_\theta = \text{diag}(J_1, \dots, J_g)$, tal que, em particular

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta^{-1}),$$

com $\beta = (\beta_1, \dots, \beta_g)^t$ e

$$V_\beta^{-1} = \text{diag}\left(\frac{\sigma_{\beta_1}^2}{a_1}, \dots, \frac{\sigma_{\beta_g}^2}{a_g}\right),$$

onde a notação $\text{diag}(d_1, \dots, d_g)$ é definida para uma matriz diagonal de ordem $g \times g$ com elementos na diagonal dados por $d_1, \dots, d_g$.

4.4 Hipóteses restritas

Conforme descrito na Seção 2.2 vamos supor $h(\theta) = (h_1(\theta), \dots, h_k(\theta))^t$, onde $k \leq 5g$, e suponhamos $\Pi(\theta) = \{(\partial/\partial\theta)h(\theta)\}$ contínua e de posto completo. Então, quando $n \rightarrow \infty$, segue que

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \xrightarrow{d} N(0, \Pi(\theta)V_\theta^{-1}\Pi(\theta)^t). \quad (4.16)$$

Vamos supor o caso I em que $H_0 : h_1(\theta) = 0, \dots, h_k(\theta) = 0$ contra $H_1 : h_1(\theta) \neq 0, h_2(\theta) \geq 0$, com pelo menos uma desigualdade estrita em H_1 quando $k_1 = 0$. Supondo que $L(\theta)$ dado em (4.10) é uma função côncava em θ e que as condições de Gourieroux e Monfort (1995) valem para a normal bivariada Z_{ij} , $i = 1, 2, \dots, g$ e $j = 1, 2, \dots, r_i$, temos a distribuição nula assintótica

$$\lim_{n \rightarrow \infty} P_r[\mathcal{W} \geq c] = \sum_{\ell=0}^q Q(q, \ell; \Sigma) P_r[\chi_\ell^2 \geq c],$$

onde $\Sigma = \Pi(\theta)V_\theta^{-1}\Pi(\theta)^t$. Discutimos a seguir alguns casos em que a distribuição nula assintótica acima reduz à qui-quadrado barra.

4.5 Coeficiente de confiabilidade conhecido

Consideraremos inicialmente o caso em que $k_{xi} = \sigma_{xi}^2 / (\sigma_{xi}^2 + \sigma_{ui}^2)$ é conhecido, $i = 1, \dots, g$. Seja $\theta_i = (\phi_i, \beta_i)^t$, $I_\theta = \text{diag}(I_1, \dots, I_g)$ e suponhamos que $n_i/n \rightarrow a_i > 0$ quando $n \rightarrow \infty$, $i = 1, \dots, g$. Assim

$$V_\theta = \lim_{n \rightarrow \infty} \frac{I_\theta}{n} = \text{diag}(J_1, \dots, J_g)$$

é uma matriz positiva definida com

$$K_i \rightarrow J_i = a_i \text{diag}(\Sigma_i, D),$$

onde $J_i = \lim_{n \rightarrow \infty} \frac{I_i}{n}$ e

$$D = \text{diag}\left(\frac{1}{2\phi_{3i}^2}, \frac{1}{2\phi_{4i}^2}, \frac{\lambda_{xi}^2 \phi_{4i}}{\phi_{3i}}\right).$$

Novamente, podemos afirmar que sob condições de regularidade (vide Bradley e Gart, 1962)

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta^{-1})$$

e em particular

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta^{-1}),$$

com $\beta = (\beta_1, \dots, \beta_g)^t$.

Para esse caso, demonstramos que $V_\theta^{-1} = \text{diag}(J_1^{-1}, \dots, J_g^{-1})$, onde

$$J_i^{-1} = \frac{1}{a_i} \begin{pmatrix} \Sigma_i^{-1} & 0 & 0 & 0 \\ 0 & 2\phi_{3i}^2 & 0 & 0 \\ 0 & 0 & 2\phi_{4i}^2 & 0 \\ 0 & 0 & 0 & \frac{\phi_{3i}}{\lambda_{xi}^2 \phi_{4i}} \end{pmatrix}$$

e

$$V_{\beta}^{-1} = \text{diag}(J_1^{-1}(\beta), \dots, J_g^{-1}(\beta))$$

com

$$J_i^{-1}(\beta) = \frac{\phi_{3i}}{a_i \lambda_{xi}^2 \phi_{4i}}, \quad i = 1, \dots, g.$$

4.5.1 Ordem simples

Analisaremos inicialmente o caso $g = 4$. Para testar hipóteses em ordem simples, isto é, $H_0 : \beta_1 = \dots = \beta_4$ contra $H_1 : \beta_1 \leq \dots \leq \beta_4$, com pelo menos uma desigualdade estrita, as quais podem ser escritas como $H_0 : \Pi_2 \beta = 0$ contra $H_1 : \Pi_2 \beta \geq 0$, onde $\beta = (\beta_1, \dots, \beta_4)^t$ e

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix},$$

temos que

$$\Pi_2 V_{\beta}^{-1} \Pi_2' = S + S_0,$$

onde

$$S = \begin{pmatrix} s_1 & -s_2 & 0 \\ -s_2 & s_2 & -s_3 \\ 0 & -s_3 & s_3 \end{pmatrix}, \quad S_0 = \text{diag}(s_2, s_3, s_4)$$

e $s_k = \phi_{3k}/a_k \phi_{4k} \lambda_{xk}$ para $k = 1, 2, 3, 4$. De $\Pi_2 V_{\beta}^{-1} \Pi_2'$ obtemos

$$\rho_{12} = \rho_{21} = - \left\{ \left(1 + \frac{a_2 \phi_{33} \phi_{42} \lambda_{x2}^2}{a_3 \phi_{32} \phi_{43} \lambda_{x3}^2} \right) \left(1 + \frac{a_2 \phi_{31} \phi_{42} \lambda_{x2}^2}{a_1 \phi_{32} \phi_{41} \lambda_{x1}^2} \right) \right\}^{-1/2}$$

e

$$\rho_{23} = \rho_{32} = - \left\{ \left(1 + \frac{a_3 \phi_{34} \phi_{43} \lambda_{x3}^2}{a_3 \phi_{33} \phi_{44} \lambda_{x4}^2} \right) \left(1 + \frac{a_3 \phi_{32} \phi_{43} \lambda_{x3}^2}{a_2 \phi_{33} \phi_{42} \lambda_{x2}^2} \right) \right\}^{-1/2}.$$

De forma geral, ao testar $H_0 : \beta_1 = \dots = \beta_g$, contra $H_1 : \beta_1 \leq \dots \leq \beta_g$, com pelo menos uma desigualdade estrita sob H_1 , os coeficientes de correlação sob a hipótese nula tomam a forma

$$\rho_{ij} = - \left\{ \left(1 + \frac{a_j \phi_{3i} \phi_{4j} \lambda_{xj}^2}{a_i \phi_{3j} \phi_{4i} \lambda_{xi}^2} \right) \left(1 + \frac{a_j \phi_{3,i+2} \phi_{4j} \lambda_{xj}^2}{a_{i+2} \phi_{3j} \phi_{4,i+2} \lambda_{x,i+2}^2} \right) \right\}^{-1/2} \quad (4.17)$$

quando $|i - j| = 1$, $\rho_{ij} = 1$ se $i = j$ e $\rho_{ij} = 0$ se $|i - j| > 1$, $i = 1, \dots, g - 2$. Segue da expressão (4.17) que quando ϕ_{3i} e ϕ_{4i} são constantes nas g populações, isto é

$$\phi_{3i} = \phi_3 \quad \text{e} \quad \phi_{4i} = \phi_4, \quad (4.18)$$

$i = 1, \dots, g$, então as correlações ρ_{ij} dependem unicamente dos pesos $a_1^* = a_1 \lambda_{x1}^2, \dots, a_g^* = a_g \lambda_{xg}^2$ os quais são conhecidos. Para esse caso, a distribuição nula assintótica para o teste estatístico $\mathcal{W}$, para efeito de testar a homogeneidade das g médias contra ordem simples, reduz para a conhecida distribuição qui-quadrado barra

$$\lim_{n \rightarrow \infty} Pr[\mathcal{W} \geq c] = \sum_{\ell=1}^g P(\ell, g; a^*) Pr\{\chi_{\ell-1}^2 \geq c\}, \quad (4.19)$$

onde $P(\ell, s; a^*) = Q(s - 1, \ell - 1; \Sigma)$ e $a^* = (a_1^*, \dots, a_g^*)^t$ são os pesos. Em particular, se temos pesos iguais, $a_1^* = \dots = a_g^*$, os coeficientes nulos $P(\ell, s; a^*)$ tomam a forma recursiva (vide Apêndice A) $P_S(1, s) = 1/g$, $P_S(s, s) = 1/s!$ e

$$P_S(i, s) = \frac{1}{s} P_S(i - 1, s - 1) + \frac{g - 1}{s} P_S(i, s - 1),$$

$i = 2, \dots, g - 1$. Por outro lado, se pretendemos testar $H_0 : \beta_1 \leq \dots \leq \beta_g$ contra $H_1 : \beta_i \in R$, $i = 1, 2, \dots, g$ e se assumimos que $\phi_{3i} = \phi_3$ e $\phi_{4i} = \phi_4$ para todo i , segue dos resultados acima e da Seção 2.4 que a distribuição nula assintótica menos favorável é dada por

$$\lim_{n \rightarrow \infty} Pr\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a^*) Pr\{\chi_{g-\ell}^2 \geq c\}. \quad (4.20)$$

Caso não seja razoável supor (4.18), fica extremadamente complexo a determinação da distribuição nula menos favorável uma vez que os coeficientes de correlação associados à matriz $\Pi_2 V_\theta^{-1} \Pi_2^t$ dependerão dos parâmetros sob a hipótese nula.

4.5.2 Ordem simples na forma de árvore

Outra situação de importância prática, como já foi mencionado anteriormente, corresponde a testes de hipóteses em ordem simples em forma de árvore, ou seja, $H_0 : \beta_1 = \dots = \beta_g$ contra $H_1 : \beta_1 \leq [\beta_2, \dots, \beta_g]$, com pelo menos uma desigualdade estrita sob H_1 . Nesse caso, para $g = 4$, $\beta = (\beta_1, \dots, \beta_4)^t$ e

$$\Pi_2 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

segue que

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \frac{\phi_{31}}{a_1 \phi_{41} \lambda_{x1}^2} (A + D),$$

onde A é uma matriz 3×3 com 1 em todas suas posições e

$$D = \frac{a_1 \phi_{41} \lambda_{x1}^2}{\phi_{31}} \text{diag}(d_2, d_3, d_4),$$

onde $d_j = \frac{\phi_{3j}}{a_j \phi_{4j} \lambda_{xj}^2}$, $j = 2, 3, 4$. Daí obtemos

$$\begin{aligned} \rho_{12} = \rho_{21} &= \left\{ \left(1 + \frac{a_1 \phi_{32} \phi_{41} \lambda_{x1}^2}{a_2 \phi_{31} \phi_{42} \lambda_{x2}^2} \right) \left(1 + \frac{a_1 \phi_{33} \phi_{41} \lambda_{x1}^2}{a_3 \phi_{31} \phi_{43} \lambda_{x3}^2} \right) \right\}^{-1/2}, \\ \rho_{13} = \rho_{31} &= \left\{ \left(1 + \frac{a_1 \phi_{32} \phi_{41} \lambda_{x1}^2}{a_2 \phi_{31} \phi_{42} \lambda_{x2}^2} \right) \left(1 + \frac{a_1 \phi_{34} \phi_{41} \lambda_{x1}^2}{a_4 \phi_{31} \phi_{44} \lambda_{x4}^2} \right) \right\}^{-1/2} \text{ e} \\ \rho_{23} = \rho_{32} &= \left\{ \left(1 + \frac{a_1 \phi_{33} \phi_{41} \lambda_{x1}^2}{a_3 \phi_{31} \phi_{43} \lambda_{x3}^2} \right) \left(1 + \frac{a_1 \phi_{34} \phi_{41} \lambda_{x1}^2}{a_4 \phi_{31} \phi_{44} \lambda_{x4}^2} \right) \right\}^{-1/2}. \end{aligned}$$

A forma geral dos ρ_{ij} , sob H_0 é a seguinte:

$$\rho_{ij} = \left\{ \left(1 + \frac{a_1 \phi_{3,i+1} \phi_{41} \lambda_{x1}^2}{a_{i+1} \phi_{31} \phi_{4,i+1} \lambda_{x,i+1}^2} \right) \left(1 + \frac{a_1 \phi_{3,j+1} \phi_{41} \lambda_{x1}^2}{a_{j+1} \phi_{31} \phi_{4,j+1} \lambda_{x,j+1}^2} \right) \right\}^{-\frac{1}{2}} \quad (4.21)$$

para $i, j = 1, \dots, g-1$. Para testar a ordem simples em forma de árvore temos de (4.18) e (4.21) que quando ϕ_{3i} e ϕ_{4i} são constantes em todas as g populações, as correlações ρ_{ij} dependem unicamente dos pesos $a_1 \lambda_{x1}^2, \dots, a_g \lambda_{xg}^2$. Nesse caso, a distribuição nula assintótica de $\mathcal{W}$ é dada por

$$\lim_{n \rightarrow \infty} P_r \{ \mathcal{W} \geq c \} = \sum_{\ell=0}^{g-1} Q(g-1, \ell; \Sigma) P_r \{ \chi_\ell^2 \geq c \}, \quad (4.22)$$

onde $\Sigma = \Pi_2 V_\beta^{-1} \Pi_2'$ não depende dos parâmetros, sendo portanto conhecida o que leva a uma única distribuição. No entanto, não é possível reduzi-la à qui-quadrado barra. Levando em conta o fato de $\Pi_2 V_\beta^{-1} \Pi_2'$ ser conhecida sob a hipótese $H_0 : \beta_1 = \dots = \beta_g$, podemos mostrar usando o lema de Wolak (vide Seção 2.4) que a distribuição nula assintótica menos favorável para testarmos $H_0 : \beta_1 \leq [\beta_2, \dots, \beta_g]$ contra $H_1 : \beta_i \in R, i = 1, 2, \dots, g$, com a suposição (4.18), é dada por

$$\lim_{n \rightarrow \infty} P_r \{ \mathcal{W} \geq c \} = \sum_{\ell=0}^{g-1} Q(g-1, g-1-\ell; \Sigma) P_r \{ \chi_\ell^2 \geq c \}.$$

4.6 Razão de variâncias conhecida

Consideraremos nesta seção o caso em que a razão das variâncias $\lambda_{\epsilon i} = \frac{\sigma_{\epsilon i}^2}{\sigma_u^2}$ é conhecida, $i = 1, \dots, g$. Sem perda de generalidade, suponhamos que $\lambda_{\epsilon i} = 1$. Seja $\theta_i = (\phi_i, \beta_i)'$, $I_\theta = \text{diag}(I_1, \dots, I_g)$ e suponhamos que $n_i/n \rightarrow a_i > 0$ quando $n \rightarrow \infty, i = 1, \dots, g$. Assim

$$V_\theta = \lim_{n \rightarrow \infty} \frac{I_\theta}{n} = \text{diag}(J_1, \dots, J_g)$$

é uma matriz positiva definida com

$$K_i \rightarrow J_i = a_i \text{diag}(\Sigma_i, D),$$

onde $J_i = \lim_{n \rightarrow \infty} \frac{I_i}{n}$ e

$$D = \text{diag}\left(\frac{1}{2\phi_{3i}^2}, \frac{1}{2\phi_{4i}^2}, \frac{(\phi_{3i} - \phi_{4i})^2}{\phi_{3i}\phi_{4i}(\beta_i^2 + 1)^2}\right).$$

Para esse caso, $V_\theta^{-1} = \text{diag}(J_1^{-1}, \dots, J_g^{-1})$, onde

$$J_i^{-1} = \frac{1}{a_i} \begin{pmatrix} \Sigma_i^{-1} & 0 & 0 & 0 \\ 0 & 2\phi_{3i}^2 & 0 & 0 \\ 0 & 0 & 2\phi_{4i}^2 & 0 \\ 0 & 0 & 0 & \frac{(\beta_i^2 + 1)^2 \phi_{3i} \phi_{4i}}{(\phi_{3i} - \phi_{4i})^2} \end{pmatrix}.$$

Seja

$$V_\beta^{-1} = \text{diag}(J_1^{-1}(\beta), \dots, J_g^{-1}(\beta)),$$

onde

$$J_i^{-1}(\beta) = \frac{(\beta_i^2 + 1)^2 \phi_{3i} \phi_{4i}}{(\phi_{3i} - \phi_{4i})^2 a_i}, \quad i = 1, \dots, g.$$

4.6.1 Ordem simples

No caso de ordem simples e $g = 4$, isto é, $H_0 : \beta_1 = \dots = \beta_4$ contra $H_1 : \beta_1 \leq \dots \leq \beta_4$, com pelo menos uma desigualdade estrita sob H_1 , temos que $\Pi_2 V_\beta^{-1} \Pi_2^t$ tem a forma

$$\Pi_2 V_\beta^{-1} \Pi_2^t = S + S_0,$$

onde

$$S = \begin{pmatrix} s_1 & -s_2 & 0 \\ -s_2 & s_2 & -s_3 \\ 0 & -s_3 & s_3 \end{pmatrix}, \quad S_0 = \text{diag}(s_2, s_3, s_4)$$

e $s_k = (\beta_j^2 + 1)^2 \phi_{3k} \phi_{4k} / a_k (\phi_{3k} - \phi_{4k})^2$ para $k = 1, 2, 3, 4$. Aqui obtemos

$$\begin{aligned} \rho_{12} = \rho_{21} = & - \left\{ \left(1 + \frac{a_2 (\beta_1^2 + 1)^2 \phi_{31} \phi_{41} (\phi_{32} - \phi_{42})^2}{a_1 (\phi_{31} - \phi_{41})^2 (\beta_2^2 + 1)^2 \phi_{32} \phi_{42}} \right) \right. \\ & \left. \left(1 + \frac{a_2 (\beta_3^2 + 1)^2 \phi_{33} \phi_{43} (\phi_{32} - \phi_{42})^2}{a_3 (\phi_{33} - \phi_{43})^2 (\beta_2^2 + 1)^2 \phi_{32} \phi_{42}} \right) \right\}^{-1/2} e \end{aligned}$$

$$\begin{aligned} \rho_{23} = \rho_{32} = & - \left\{ \left(1 + \frac{a_3 (\beta_2^2 + 1)^2 \phi_{32} \phi_{42} (\phi_{33} - \phi_{43})^2}{a_2 (\phi_{32} - \phi_{42})^2 (\beta_3^2 + 1)^2 \phi_{33} \phi_{43}} \right) \right. \\ & \left. \left(1 + \frac{a_3 (\beta_4^2 + 1)^2 \phi_{34} \phi_{44} (\phi_{33} - \phi_{43})^2}{a_4 (\phi_{34} - \phi_{44})^2 (\beta_3^2 + 1)^2 \phi_{33} \phi_{43}} \right) \right\}^{-1/2}. \end{aligned}$$

Genericamente, para o caso de ordem simples, os coeficientes de correlação associados com a matriz $\Pi_2 V_\beta^{-1} \Pi_2^t$ são da forma

$$\begin{aligned} \rho_{ij} = & - \left\{ \left(1 + \frac{a_j (\beta_i^2 + 1)^2 \phi_{3i} \phi_{4i} (\phi_{3j} - \phi_{4j})^2}{a_i (\phi_{3i} - \phi_{4i})^2 (\beta_j^2 + 1)^2 \phi_{3j} \phi_{4j}} \right) \right. \\ & \left. \left(1 + \frac{a_j (\beta_{i+2}^2 + 1)^2 \phi_{3,i+2} \phi_{4,i+2} (\phi_{3j} - \phi_{4j})^2}{a_{i+2} (\phi_{3,i+2} - \phi_{4,i+2})^2 (\beta_j^2 + 1)^2 \phi_{3j} \phi_{4j}} \right) \right\}^{-1/2} \end{aligned}$$

se $|i - j| = 1$, $\rho_{ij} = 1$ se $i = j$ e $\rho_{ij} = 0$ se $|i - j| > 1$. Sob H_0 , esses coeficientes

de correlação tomam a forma

$$\rho_{ij} = - \left\{ \left(1 + \frac{a_j \phi_{3i} \phi_{4i} (\phi_{3j} - \phi_{4j})^2}{a_i (\phi_{3i} - \phi_{4i})^2 \phi_{3j} \phi_{4j}} \right) \left(1 + \frac{a_j \phi_{3,i+2} \phi_{4,i+2} (\phi_{3j} - \phi_{4j})^2}{a_{i+2} (\phi_{3,i+2} - \phi_{4,i+2})^2 \phi_{3j} \phi_{4j}} \right) \right\}^{-1/2} \quad (4.23)$$

se $|i - j| = 1$, $\rho_{ij} = 1$ se $i = j$ e $\rho_{ij} = 0$ se $|i - j| > 1$. Devemos notar que sob as suposições (4.18), as correlações sob H_0 dependem unicamente dos pesos $a_1, \dots, a_g$. Portanto, a distribuição nula assintótica de $\mathcal{W}$ é dada por (4.19). Para pesos iguais podemos utilizar $P_s(\ell, g)$ no lugar de $P(\ell, g; a)$. Se o interesse é agora testar $H_0 : \beta_1 \leq \dots \leq \beta_g$ contra $H_1 : \beta_i \in R$, $i = 1, 2, \dots, g$, sob (4.18), temos pelo lema de Wolak que a distribuição nula assintótica menos favorável é alcançada quando $\beta_1 = \dots = \beta_g$, sendo portanto dada por

$$\lim_{n \rightarrow \infty} P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a) P_r\{\chi_{g-\ell}^2 \geq c\}.$$

4.6.2 Ordem simples na forma de árvore

Discutimos agora o caso de ordem simples em forma de árvore com $g = 4$, isto é, $H_0 : \beta_1 = \dots = \beta_4$ contra $H_1 : \beta_1 \leq [\beta_2, \beta_3, \beta_4]$, com pelo menos uma desigualdade estrita sob H_1 . Neste caso, podemos mostrar que

$$\Pi_2 V_\beta^{-1} \Pi_2^t = \frac{(\beta_1^2 + 1)^2 \phi_{31} \phi_{41}}{a_1 (\phi_{31} - \phi_{41})^2} (A + D),$$

onde A é uma matriz 3×3 com 1 em todas suas posições e

$$D = \frac{a_1 (\phi_{31} - \phi_{41})^2}{\phi_{31} \phi_{41} (\beta_1^2 + 1)^2} \text{diag}(d_2, d_3, d_4),$$

onde $d_j = \frac{\phi_{3j} \phi_{4j} (\beta_j^2 + 1)^2}{a_j (\phi_{3j} - \phi_{4j})^2}$, $j = 2, 3, 4$. Usando os resultados anteriores podemos mostrar que para testar $H_0 : \beta_1 = \dots = \beta_g$ contra $H_1 : \beta_1 \leq [\beta_2, \dots, \beta_g]$, os

coeficientes de correlação tomam as formas

$$\rho_{ij} = \left\{ \left(1 + \frac{(\beta_{i+1} + 1)^2 \phi_{3,i+1} \phi_{4,i+1}}{a_{i+1} (\phi_{3,i+1} - \phi_{4,i+1})^2} \gamma \right) \left(1 + \frac{(\beta_{j+1}^2 + 1)^2 \phi_{3,j+1} \phi_{4,j+1}}{a_{j+1} (\phi_{3,j+1} - \phi_{4,j+1})^2} \gamma \right) \right\}^{-\frac{1}{2}}$$

se $|i - j| = 1$ e $\rho_{ij} = 1$ se $i = j$, com

$$\gamma = \frac{a_1 (\phi_{31} - \phi_{41})^2}{(\beta_1^2 + 1)^2 \phi_{31} \phi_{41}}.$$

Sob H_0 , os coeficientes de correlação tomam a forma

$$\begin{aligned} \rho_{ij} = & \left\{ \left(1 + \frac{\phi_{3,i+1} \phi_{4,i+1}}{a_{i+1} (\phi_{3,i+1} - \phi_{4,i+1})^2} \frac{a_1 (\phi_{3i} - \phi_{4i})^2}{\phi_{31} \phi_{41}} \right) \right. \\ & \left. \left(1 + \frac{\phi_{3,j+1} \phi_{4,j+1}}{a_{j+1} (\phi_{3,j+1} - \phi_{4,j+1})^2} \frac{a_1 (\phi_{3i} - \phi_{4i})^2}{\phi_{31} \phi_{41}} \right) \right\}^{-\frac{1}{2}} \end{aligned} \quad (4.24)$$

se $|i - j| = 1$ e $\rho_{ij} = 1$ se $i = j$. Mostra-se facilmente que sob (4.18) a distribuição nula assintótica de $\mathcal{W}$ não depende dos parâmetros, no entanto os coeficientes de correlação associados à matriz $\Pi_2 V_\beta^{-1} \Pi_2^t$ não tomam uma forma que leve à distribuição qui-quadrado barra. Para o caso mais geral de testarmos $H_0 : \beta_1 \leq [\beta_2, \dots, \beta_g]$ contra $H_1 : \beta_i \in R, i = 1, 2, \dots, g$, a distribuição nula assintótica menos favorável é alcançada sob (4.18) quando $\beta_1 = \dots = \beta_g$ sendo portanto única pelos resultados mencionados anteriormente.

4.7 Simulações

Para investigar o uso das probabilidades com pesos iguais $P_S(\ell, s)$ como uma aproximação para $Q(s-1, \ell-1; \Sigma)$, quando os pesos $a_1, \dots, a_g$ não são muito diferentes e as diferenças $\Delta\phi_3 = \max \phi_{3i} - \min \phi_{3i}$ e $\Delta\phi_4 = \max \phi_{4i} - \min \phi_{4i}$ não se afastem muito de zero, temos desenvolvido um estudo de simulação

cujos resultados são apresentados a seguir. É importante lembrar que para o caso de ordem simples quando $\phi_{3i} = \phi_3$ e $\phi_{4i} = \phi_4$, para todo i , temos uma qui-quadrado barra e em particular quando $a_1 = \dots = a_g$, sempre podemos determinar os coeficientes da distribuição nula assintótica de $\mathcal{W}$.

Sem perda de generalidade, assumimos $g = 4$ e geramos ϕ_{3i} e ϕ_{4i} como sendo variáveis aleatórias com distribuição exponencial. Para o caso da razão de variâncias conhecida, temos assumido $\lambda_{ei} = 1$, $i = 1, \dots, 4$. Quatro situações possíveis foram estudadas: pesos iguais, pesos diferentes, $w = (a, 3a, a, 3a)$, pesos- $\cup$, $w = (3a, a, a, 3a)$ e pesos- $\cap$, $w = (a, 3a, 3a, a)$. As Tabelas 4.1 e 4.2 apresentam os resultados para o caso da razão de variâncias conhecida, enquanto que as Tabelas 4.3 e 4.4 apresentam os resultados quando o coeficiente de confiabilidade é conhecido. Foi calculada a distribuição de frequências de

$$C_{01}(t; \Sigma) = \sum_{\ell=0}^3 Q(3, \ell; \Sigma) P_r \{ \chi_{\ell-1}^2 \geq t \},$$

para $t = c_{01}(0.05)$, valor crítico de 5% do caso de pesos iguais. 10000 conjuntos aleatórios de $(\phi_{31}, \dots, \phi_{34}, \phi_{41}, \dots, \phi_{44})$ foram gerados. Para cada valor de $\Delta\phi_3$ e $\Delta\phi_4$, o valor de $C_{01}(c_{01}(0.05); \Sigma)$ foi calculado. Essas 10000 probabilidades foram classificadas segundo a terceira coluna das Tabelas 4.1 - 4.4. Tipicamente encontramos os limites $0.017 \leq C_{01}(t; \Sigma) \leq 0.077$ e $0 \leq \Delta\phi_3, \Delta\phi_4 \leq 15$.

Da Tabela 4.1 podemos notar um desempenho satisfatório da aproximação de pesos iguais, particularmente para $\Delta\phi_3$ e $\Delta\phi_4$ próximos de zero. Não foi encontrada diferença substancial entre os casos de pesos iguais e pesos diferentes. Para pesos- $\cup$ (Tabela 4.2) as probabilidades subestimam, em geral, o valor verdadeiro da probabilidade. A maioria dos casos caem no intervalo $(0.017, 0.060)$. Para pesos- $\cap$ (Tabela 4.2), em contraste com o caso de pesos- $\cup$,

as probabilidades são superestimadas. A maioria dos casos estão entre 0.040 e 0.077. Finalmente, para o caso do coeficiente de confiabilidade conhecido (Tabelas 4.3 e 4.4), observamos tendências similares a aquelas encontradas nas Tabelas 4.1 e 4.2. Porém, como observado das tabelas, boas aproximações não estão necessariamente associadas com valores pequenos para $\Delta\phi_3$ e $\Delta\phi_4$.

Para resumir os resultados apresentados acima, podemos concluir que a aproximação de pesos iguais pode ser recomendada para os casos de pesos iguais e diferentes, sob as suposições de que ϕ_3 e ϕ_4 tenham uma distribuição exponencial. Essa conclusão parece valer igualmente para os dois casos em consideração, razão de variâncias e coeficiente de confiabilidade conhecidos. Porém, para os casos de pesos- $\cup$ e pesos- $\cap$ a aproximação não parece ser adequada.

Tabela 4.1. Distribuição de frequências para $C_{01}(\epsilon_{01}(0.05); \Sigma)$, para 10000 conjuntos gerados de $(\phi_{31}, \dots, \phi_{34}, \phi_{41}, \dots, \phi_{44})$ em cada caso e assumindo a razão de variâncias conhecida.

Modelo	$\Delta\phi_4$	Intervalo	$\Delta\phi_3$		
			0-1	1-3	> 3
Pesos iguais	0-1	0.01-0.04	0.1424	0.1909	0.2108
		0.04-0.06	0.6985	0.6117	0.5486
		0.06-0.08	0.1591 (597)	0.1973 (1566)	0.2405 (370)
	1-3	0.01-0.04	0.1928	0.2075	0.2207
		0.04-0.06	0.5991	0.5653	0.5506
		0.06-0.08	0.2081 (1504)	0.2272 (3644)	0.2287 (870)
	> 3	0.01-0.04	0.2283	0.2201	0.2090
		0.04-0.06	0.5598	0.5442	0.5819
		0.06-0.08	0.2120 (368)	0.2356 (904)	0.2090 (177)
	0-1	0.01-0.04	0.1505	0.1670	0.1761
		0.04-0.06	0.7330	0.6597	0.6132
		0.06-0.08	0.1165 (618)	0.1734 (1569)	0.2107 (318)
Pesos (a,3a,a,3a)	1-3	0.01-0.04	0.1879	0.2035	0.1801
		0.04-0.06	0.6398	0.5911	0.5748
		0.06-0.08	0.1723 (1538)	0.2054 (3661)	0.2451 (816)
	> 3	0.01-0.04	0.2101	0.1989	0.2017
		0.04-0.06	0.5710	0.5709	0.5708
		0.06-0.08	0.2188 (352)	0.2302 (895)	0.2275 (233)

Tabela 4.2. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Sigma)$, para 10000 conjuntos gerados de $(\phi_{31}, \dots, \phi_{34}, \phi_{41}, \dots, \phi_{44})$ em cada caso e assumindo a razão de variâncias conhecida.

Modelo	$\Delta\phi_4$	Intervalo	$\Delta\phi_3$		
			0-1	1-3	> 3
pesos-U	0-1	0.01-0.04	0.4819	0.4264	0.4299
		0.04-0.06	0.4680	0.5156	0.4627
		0.06-0.08	0.0501 (579)	0.0579 (1536)	0.1075 (335)
	1-3	0.01-0.04	0.4147	0.4194	0.3945
		0.04-0.06	0.5278	0.4906	0.4829
		0.06-0.08	0.0575 (1548)	0.0900 (3665)	0.1227 (905)
	> 3	0.01-0.04	0.3869	0.3865	0.4201
		0.04-0.06	0.5089	0.4994	0.4703
		0.06-0.08	0.1042 (336)	0.1140 (877)	0.1096 (219)
pesos- $\cap$	0-1	0.01-0.04	0.0363	0.0506	0.1373
		0.04-0.06	0.5310	0.5395	0.4741
		0.06-0.08	0.4327 (661)	0.4100 (1444)	0.3886 (386)
	1-3	0.01-0.04	0.0593	0.0741	0.1071
		0.04-0.06	0.5422	0.5192	0.4647
		0.06-0.08	0.3985 (1551)	0.4068 (3700)	0.4282 (850)
	> 3	0.01-0.04	0.1243	0.1016	0.1117
		0.04-0.06	0.5000	0.4584	0.4681
		0.06-0.08	0.3757 (354)	0.4400 (866)	0.4202 (188)

Tabela 4.3. Distribuição de frequências para $C_{01}(\epsilon_{01}(0.05); \Sigma)$, para 10000 conjuntos gerados de $(\phi_{31}, \dots, \phi_{34}, \phi_{41}, \dots, \phi_{44})$ em cada caso e assumindo o coeficiente de confiabilidade conhecido.

Modelo	$\Delta\phi_4$	Intervalo	$\Delta\phi_3$		
			0-1	1-3	> 3
Pesos Iguais	0-1	0.01-0.04	0.2233	0.2382	0.2151
		0.04-0.06	0.5047	0.5192	0.5587
		0.06-0.08	0.2720 (636)	0.2426 (1562)	0.2263 (358)
	1-3	0.01-0.04	0.2206	0.2398	0.2577
		0.04-0.06	0.5297	0.4860	0.4879
		0.06-0.08	0.2497 (1482)	0.2741 (3615)	0.2544 (908)
	> 3	0.01-0.04	0.2147	0.2645	0.2105
		0.04-0.06	0.4810	0.4915	0.4895
		0.06-0.08	0.3043 (368)	0.2440 (881)	0.3000 (190)
	0-1	0.01-0.04	0.2289	0.2136	0.2548
		0.04-0.06	0.5261	0.5366	0.4931
		0.06-0.08	0.2349 (664)	0.2498 (1517)	0.2521 (361)
Pesos (a,3a,a,3a)	1-3	0.01-0.04	0.2288	0.2321	0.2228
		0.04-0.06	0.5284	0.5268	0.5365
		0.06-0.08	0.2427 (1512)	0.2411 (3671)	0.2407 (835)
	> 3	0.01-0.04	0.2240	0.2347	0.2240
		0.04-0.06	0.5246	0.5306	0.4948
		0.06-0.08	0.2514 (366)	0.2347 (882)	0.2812 (192)

Tabela 4.4. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Sigma)$, para 10000 conjuntos gerados de $(\phi_{31}, \dots, \phi_{34}, \phi_{41}, \dots, \phi_{44})$ em cada caso e assumindo o coeficiente de confiabilidade conhecido.

Modelo	$\Delta\phi_4$	Intervalo	$\Delta\phi_3$		
			0-1	1-3	> 3
Pesos- $\cup$	0-1	0.01-0.04	0.4041	0.4241	0.4091
		0.04-0.06	0.4308	0.4415	0.4519
		0.06-0.08	0.1651 (636)	0.1344 (1495)	0.1390 (374)
	1-3	0.01-0.04	0.3793	0.3866	0.3961
		0.04-0.06	0.4764	0.4563	0.4611
		0.06-0.08	0.1443 (1587)	0.1571 (3629)	0.1429 (861)
	> 3	0.01-0.04	0.4392	0.3968	0.4118
		0.04-0.06	0.4332	0.4618	0.4363
		0.06-0.08	0.1276 (337)	0.1414 (877)	0.1520 (204)
	0-1	0.01-0.04	0.1186	0.1089	0.1408
		0.04-0.06	0.4876	0.4916	0.4648
		0.06-0.08	0.3937 (607)	0.3995 (1487)	0.3944 (355)
Pesos- $\cap$	1-3	0.01-0.04	0.1184	0.1367	0.1067
		0.04-0.06	0.4723	0.4647	0.4904
		0.06-0.08	0.4094 (1622)	0.3986 (3658)	0.4030 (881)
	> 3	0.01-0.04	0.1175	0.1140	0.1336
		0.04-0.06	0.4536	0.4870	0.4608
		0.06-0.08	0.4290 (366)	0.3990 (807)	0.4055 (217)

Capítulo 5

Conclusões

Os resultados obtidos no presente trabalho são variados. Inicialmente, apresentamos de forma unificada os principais resultados relativos ao estudo de teste de hipóteses restritas por desigualdades não-lineares. Definimos a estatística do teste e em particular, demonstramos que sob as condições de regularidade de Gourieroux e Monfort (1995) essa estatística tem distribuição nula assintótica que é uma mistura de distribuições do tipo qui-quadrado. A partir dessas condições de regularidade mostramos a equivalência assintótica da estatística da razão de verossimilhança com as estatísticas de Wald, de escore e de Kuhn-Tucker. Também definimos as probabilidades de nível, suas simplificações e aproximações.

Os resultados anteriores foram estendidos no Capítulo 3 para o modelo de regressão com distribuição log-gama generalizada. Considerando o modelo

$$Y_{ij} = x_i^t \beta + \sigma_i Z_{ij}, \quad i = 1, \dots, g; \quad j = 1, \dots, r_i \quad (5.1)$$

para dados não censurados, onde $x_i^t = (x_{i1}, \dots, x_{ip})$, $\beta = (\beta_1, \dots, \beta_p)^t$, $\sigma_i > 0$, com os erros Z_{ij} sendo independentes e igualmente distribuídos com distribuição log-gama generalizada, investigamos testes do tipo *one-sided* para β e σ_i , $i = 1, \dots, g$.

Como um dos nossos objetivos é procurar formas simplificadas para os pesos, mostramos que algumas estruturas para a matriz X fazem com que os pesos $Q(q, \ell; \Sigma)$ deixem de depender, sob a hipótese nula (através de Σ), dos parâmetros, particularmente para o caso de testarmos $H_0 : h(\theta) \geq 0$ contra $H_1 - H_0$, onde $H_1 : \theta \in R^p$. Seguindo Wolak (1991) isso somente seria possível em situações triviais sem interesse prático. Porém mostramos aqui que é possível $Q(q, \ell; \Sigma)$ reduzir a $P(\ell + 1, q + 1; w)$ em algumas situações de importância prática em modelos log-gama generalizados. Especificamente, mostramos para o caso de g populações, que se temos em H_0 uma ordem simples, ordem simples em forma de árvore, ordem unimodal ou ordem do tipo *loop* em $\mu_1, \dots, \mu_g$ supondo $\sigma_i = \sigma$ ou em $\sigma_1, \dots, \sigma_g$ supondo $\mu_i = \mu$, as correlações associadas à matriz Σ dependem, sob $\mu_1, \dots, \mu_g$ ou sob $\sigma_1, \dots, \sigma_g$ dos pesos $a_1, \dots, a_g$, sendo a distribuição nula assintótica menos favorável de $\mathcal{W}$ uma qui-quadrado barra dada por

$$\lim_{n \rightarrow \infty} P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g F(\ell, g; a) P_r\{\chi_{g-\ell}^2 \geq c\}.$$

De igual forma mostramos que esse resultado vale também quando testarmos ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop* para os percentís $y_{p1}, \dots, y_{pg}$ e para os interceptos $\alpha_1, \dots, \alpha_g$ no modelo de retas paralelas sob a condição $\epsilon_i = 0$, para todo i .

Além desses resultados, destacamos a extensão do modelo de regressão gaus-

siano para o modelo de regressão log-gama generalizado que não é da família exponencial e os parâmetros não são ortogonais.

O estudo da distribuição qui-quadrado barra no modelo de regressão de retas paralelas, indica que a distribuição qui-quadrado barra com pesos iguais, pode ser recomendada para aproximar a distribuição nula assintótica da estatística $\mathcal{W}$, para os casos de ordem simples e ordem simples em forma de árvore, sempre que a razão entre o maior e o menor tamanho amostral não exceder 2.0 e a razão entre a maior e a menor dispersão não for superior a 3.0.

A teoria descrita no Capítulo 2, particularmente as condições de Gourieroux e Monfort (1995) foi investigada para o modelo de regressão do valor extremo com censura do tipo II. No entanto, as condições parecem fortes dificultando, como mostramos a seguir, a conclusão das investigações que poderam ser retomadas futuramente. Vamos supor então o modelo (5.1) onde $\sigma_i = \sigma$ e Z_{ij} segue uma distribuição do valor extremo padrão. Sem perda de generalidade, suponha que $y_{i1} \leq y_{i2} \leq \dots \leq y_{ic_i}$ são as c_i menores observações de uma amostra aleatória de tamanho r_i . Então, a função de densidade conjunta de $y_{i1}, y_{i2}, \dots, y_{ic_i}$ (ver Lawless, 1982) fica dada por

$$\frac{r_i!}{(r_i - c_i)!} \left\{ \prod_{j=1}^{c_i} f(y_{ij}; \mu_i, \sigma) \right\} \{S(y_{ic_i})\}^{(r_i - c_i)},$$

onde $S(y) = \exp\{-\exp(\frac{y - \mu}{\sigma})\}$ é a função de sobrevivência de Y . Além disso, assumimos a estrutura de regressão $\mu_i = \mu(x_i) = x_i^t \beta$ onde $x_i = (x_{i1}, \dots, x_{ip})^t$ é o vetor $p \times 1$ de valores das variáveis explicativas e $\beta = (\beta_1, \dots, \beta_p)^t$ é o vetor $p \times 1$ de coeficientes de regressão. A função log-verossimilhança de $(\beta^t, \sigma)^t$

pode ser expressa na forma

$$L(\beta, \sigma) = -\log \sigma \sum_{i=1}^g c_i + \left[\sum_{j=1}^{c_i} z_{ij} - \left\{ (r_i - c_i) \exp(z_{ic_i}) + \sum_{j=1}^{c_i} \exp(z_{ij}) \right\} \right],$$

onde $z_{ij} = \frac{y_{ij} - \mu(x_i)}{\sigma}$. Segue de Pratt (1981) que $L(\beta, \sigma)$ é uma função côncava de $(\beta^t, \sigma)^t$. Esse fato garante a unicidade do estimador de máxima verossimilhança $(\hat{\beta}^t, \hat{\sigma})^t$. A função de escore para β e σ pode ser expressa, respectivamente, na forma

$$U(\beta) = -\frac{1}{\sigma} X^t (C + A) \mathbf{1} \quad \text{e} \quad U(\sigma) = -\frac{1}{\sigma} \sum_{i=1}^g (c_i + z_{i.} + b_i),$$

onde $C = \text{diag}(c_1, \dots, c_g)$, $A = \text{diag}(a_1, \dots, a_g)$, $z_{i.} = \sum_{j=1}^{c_i} z_{ij}$,

$$\begin{aligned} a_i &= (r_i - c_i) \exp(z_{ic_i}) + \sum_{j=1}^{c_i} \exp(z_{ij}) \quad \text{e} \\ b_i &= (r_i - c_i) z_{ic_i} \exp(z_{ic_i}) + \sum_{j=1}^{c_i} z_{ij} \exp(z_{ij}). \end{aligned}$$

Através de cálculos similares, obtemos a matriz de informação observada,

$$F_0 = -\frac{1}{\sigma} \begin{bmatrix} X^t A X & X^t (A + B - C) \mathbf{1} \\ \mathbf{1}^t (A + B - C) X & -\sum_{i=1}^g d_i \end{bmatrix},$$

onde X é uma matriz $g \times p$ com i -ésima linha dada por $x_i = (x_{i1}, \dots, x_{ip})^t$, $B = \text{diag}(b_1, \dots, b_g)$ e $d_i = c_i + 2(z_{i.} - b_i) - (r_i - c_i) z_{ic_i}^2 \exp(z_{ic_i}) - \sum_{j=1}^{c_i} z_{ij}^2 \exp(z_{ij})$.

A esperança de F_0 leva à matriz de informação de Fisher

$$F = \frac{1}{\sigma} \begin{bmatrix} X^t C X & X^t C (I + \Gamma) \mathbf{1} \\ \mathbf{1}^t (I + \Gamma) C^t X & \mathbf{1}^t (C + \Delta) \mathbf{1} \end{bmatrix},$$

onde $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_g)$, $\Delta = \text{diag}(\delta_1, \dots, \delta_g)$, $\gamma_i = -c_i^{-1} \sum_{j=1}^{c_i} E(z_{ij})$ e $\delta_i = (r_i - c_i) E\{z_{ic_i}^2\} + \sum_{j=1}^{c_i} E\{z_{ij}^2 \exp(z_{ij})\}$, $i = 1, \dots, g$.

Em vista da impossibilidade de provarmos a convergência de $\lim_n \frac{F_0}{n}$ ou de $\lim_n \frac{E(F_0)}{n}$ para uma matriz finita positiva definida, fomos impedidos de aplicar as condições gerais de regularidade de Gourieroux e Monfort (1995) ou a teoria geral desenvolvida por Wolak (1991). Daí segue também a impossibilidade de aplicarmos a teoria geral para teste de hipóteses restritas por desigualdades definida no Capítulo 2.

Hipóteses restritas para o parâmetro β_i em modelos de regressão estrutural foram consideradas no Capítulo 4. Foi discutida a parametrização ortogonal e sua estimação por máxima verossimilhança. Provamos que $\lim_n \frac{I_{F_i}}{n}$ é uma matriz finita positiva definida, com I_{F_i} sendo a matriz de informação sob o modelo estrutural. Assim, para o caso de g populações, sob as condições de regularidade estabelecidas por Bradley e Gart (1962) temos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta^{-1}),$$

onde $V_\theta = \lim_n \frac{I_{F_i}}{n}$. Como a log-verossimilhança $L(\theta)$ dada em (4.10) é côncava em θ e confirmado que as condições de regularidade de Gourieroux e Monfort (1995) valem para a normal bivariada Z_{ij} , $i = 1, \dots, g$ e $j = 1, \dots, r_i$ provamos que a distribuição nula assintótica da estatística da razão de verossimilhança fica dada por

$$\lim_n P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^g Q(q, \ell; \Sigma) P_r\{\chi_\ell^2 \geq c\},$$

onde $\Sigma = \Pi(\theta)V_\theta^{-1}\Pi(\theta)^t$.

Ao considerarmos os dois casos, coeficiente de confiabilidade conhecido e razão de variâncias conhecida, demonstramos sob a suposição (4.18) isto é, $\phi_{3i} = \phi_3$ e $\phi_{4i} = \phi_4$ ao testar $H_0 = \beta_1 = \dots = \beta_g$ contra $H_1 = \beta_1 \leq \dots \leq \beta_g$,

os coeficientes de correlação dependem unicamente dos pesos os quais são conhecidos. Para esse caso, provamos então que a distribuição nula assintótica para o teste estatístico $\mathcal{W}$, reduz para a conhecida distribuição qui-quadrado barra

$$\lim_n P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a^*) P_r\{\chi_{\ell-1}^2 \geq c\}.$$

Por outro lado, se pretendemos testar $H_0 = \beta_1 \leq \dots \leq \beta_g$ contra $H_1 : \beta_i \in R$, $i = 1, \dots, g$ e assumimos (4.18), temos que a distribuição nula assintótica menos favorável fica dada por

$$\lim_n P_r\{\mathcal{W} \geq c\} = \sum_{\ell=1}^g P(\ell, g; a^*) P_r\{\chi_{g-\ell}^2 \geq c\}.$$

Para o caso de ordem simples em forma de árvore, provamos que os coeficientes de correlação dependem unicamente dos pesos. Nesse caso, a distribuição nula assintótica de $\mathcal{W}$ fica dada por

$$\lim_n P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^{g-1} Q(g-1, \ell; \Sigma) P_r\{\chi_{\ell}^2 \geq c\},$$

onde $\Sigma = \Pi_2 V_{\beta}^{-1} \Pi_2^t$ não depende dos parâmetros, sendo portanto conhecida, o que leva a uma única distribuição. No entanto, não é possível reduzi-la à qui-quadrado barra.

Levando em conta o fato de $\Pi_2 V_{\beta}^{-1} \Pi_2^t$ ser conhecida sob a hipótese $H_0 = \beta_1 = \dots = \beta_g$, mostraamos usando o lema de Wolak (1991) que a distribuição nula assintótica menos favorável para testarmos $H_0 : \beta_1 \leq [\beta_2, \dots, \beta_g]$ contra $H_1 : \beta_i \in R$, $i = 1, \dots, g$ com a suposição (4.18), é dada por

$$\lim_n P_r\{\mathcal{W} \geq c\} = \sum_{\ell=0}^{g-1} Q(g-1, g-1-\ell; \Sigma) P_r\{\chi_{\ell}^2 \geq c\}.$$

O estudo de simulação para efeito de determinarmos os casos em que a distribuição qui-quadrado barra pode ser aproximada pela distribuição qui-quadrado barra com pesos iguais, indica que essa última pode ser recomendada para os casos de pesos iguais e diferentes, sob as suposições de que ϕ_3 e ϕ_4 tenham distribuição exponencial. Essa conclusão parece valer igualmente para os dois casos em consideração, razão de variâncias e coeficiente de confiabilidade conhecidos. Porém para os casos de pesos- $\cap$ e pesos- $\cup$ a aproximação não parece se adequar.

5.1 Futuras pesquisas

Os seguintes assuntos são relacionados com a presente tese, os quais poderam ser investigados em futuras pesquisas:

- Estender a teoria desenvolvida aqui para dados com algum tipo de censura. Em particular, dando continuidade ao presente trabalho, estamos desenvolvendo e aplicando esta teoria para dados com censura do tipo I.
- Estudos de poder podem ser desenvolvidos a fim de compararmos as estatísticas assintoticamente equivalentes.
- Também estamos considerando o modelo de calibração como uma outra aplicação da teoria desenvolvida aqui.
- Acreditamos que esta teoria possa ser desenvolvida e aplicada na família de distribuições elípticas.

Apêndice A

Aproximações para $P(\ell, k; w)$

Como foi mencionado na Seção 2.5, para uma classe de hipóteses restritas conhecidas como de quase-ordem $\preceq$, os coeficientes de correlação associados à matriz Σ tomam uma forma simplificada e não dependem dos parâmetros sob a hipótese nula. Nesses casos, a probabilidade de nível $Q(k, \ell; \Sigma)$ é denotada por $P(\ell + 1, k + 1; w)$, onde $w = (w_1, \dots, w_k)^t$ é um vetor de pesos que determinam os coeficientes de correlação de Σ . Várias aproximações têm sido desenvolvidas para $P(\ell, k; w)$ quando as hipóteses envolvidas são as médias de normais independentes com média comum e variâncias $w_1^{-1}, \dots, w_k^{-1}$. Se w_i depende de σ^2 conhecido, as estatísticas são conhecidas como qui-quadrado barra e denotadas por $\bar{\chi}_{01}^2$ e $\bar{\chi}_{12}^2$ dependendo se estamos no caso I ou caso II, respectivamente. Quando σ^2 é desconhecido, essas estatísticas são denotadas por $\bar{E}_{01}^2$ e $\bar{E}_{12}^2$, respectivamente. Apresentamos o cálculo e algumas aproximações para $P(\ell, k; w)$ para os casos de ordem simples, ordem simples na forma de árvore e ordem do tipo *loop*.

A.0.1 Um algoritmo geral para calcular $P(\ell, k; w)$

Uma aproximação recursiva para calcular as probabilidades de nível $P(\ell, k; w)$'s consiste de duas partes. Primeiro é calculado $P(k, k; w)$ e depois uma fórmula recursiva para $P(\ell, k; w)$ com $1 \leq \ell \leq k$ é obtida em função de $P(j, j; w)$ e $P(1, j; w)$ com $j \leq k$. De fato, o valor de $P(1, k; w)$ pode ser obtido da equação

$$\sum_{\ell=1}^k P(\ell, k; w) = 1.$$

Com efeito, sejam $Y_1, Y_2, \dots, Y_k$ variáveis aleatórias normais independentes com uma média comum e variâncias $w_1^{-1}, w_2^{-1}, \dots, w_k^{-1}$ e suponhamos que M é o número de conjuntos de nível na regressão isotônica de $Y = (Y_1, Y_2, \dots, Y_k)^t$ com pesos w_i e uma ordem especificada que denotamos por $\preceq$. A probabilidade $P_r(Y_1 \leq Y_2 \leq \dots \leq Y_k)$ é determinante nesses cálculos. Pela transformação $U_i = Y_{i+1} - Y_i$ para $i = 1, \dots, k-1$, essa probabilidade é transformada na probabilidade do quadrante, P_{k-1} , definida por

$$P_{k-1} = P_r[U_1 \geq 0, U_2 \geq 0, \dots, U_{k-1} \geq 0], \quad (\text{A.1})$$

na qual $U = (U_1, \dots, U_{k-1})^t$ tem distribuição normal multivariada com média zero e matriz de correlação na forma ($\rho_{ii} = 1$, $\rho_{ij} = 0$ se $|i - j| \geq 1$)

$$\rho_{i,i+1} = \rho_{i+1,i} = - \left[\frac{w_i w_{i+2}}{(w_i + w_{i+1})(w_{i+1} + w_{i+2})} \right], \quad i = 1, 2, \dots, k-2. \quad (\text{A.2})$$

Grande parte da literatura existente para a determinação das probabilidades do quadrante já foi apresentada na Seção 2.5. As expressões para os primeiros quatro P_{k-1} com pesos arbitrários são

$$P_1 = \frac{1}{2}$$

$$\begin{aligned}
P_2 &= \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12} \\
P_3 &= \frac{1}{8} + \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{13} + \sin^{-1} \rho_{23})
\end{aligned}$$

e

$$\begin{aligned}
P_4 &= \frac{1}{16} + \frac{1}{8\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{23} + \sin^{-1} \rho_{34}) \\
&\quad + \frac{1}{4\pi^2} \int_0^{-\rho_{34}} \int_0^{-\rho_{12}} [(1-u^2)(1-v^2) - \rho_{23}^2]^{-\frac{1}{2}} \partial u \partial v
\end{aligned} \tag{A.3}$$

(vide Abrahamsom 1964, Childs 1967). A integral em P_4 pode ser escrita como

$$\int_0^{-\rho_{12}} (1-v^2)^{-\frac{1}{2}} \sin^{-1} \left[\frac{-(1-v^2)^{\frac{1}{2}} \rho_{12}}{(1-v^2 - \rho_{23}^2)^{\frac{1}{2}}} \right] \partial v,$$

e o termo envolvido na integral pode ser obtido das tabelas de Abrahamsom. Ele tabela uma função $P(a, b, c)$ para valores não negativos de seus argumentos. Em termos de $P(a, b, c)$,

$$P_4 = P(-\rho_{12}, -\rho_{23}, -\rho_{34}) + \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{23} + \sin^{-1} \rho_{34}).$$

Não são conhecidas formas fechadas para P_{k-1} , com $k > 5$. Se os pesos são iguais, $w_1 = w_2 = \dots = w_k$, então P_{k-1} é a probabilidade dos elementos de uma amostra aleatória serem não decrescentes. Como cada permutação dos postos é igualmente provável, $P_{k-1} = 1/k!$, esse resultado vale para qualquer conjunto de variáveis aleatórias Y_i simétricas independentes para as quais ocorrem empates com probabilidade zero. Em particular, o resultado vale se as Y_i 's são variáveis contínuas independentes e identicamente distribuídas.

Para ilustrar o uso de P_{k-1} no cálculo de $P(k, k; w)$, consideremos a ordem do tipo *loop*. Antes porém adotemos a seguinte notação :

$$[i_1, i_2, \dots, i_\alpha] \preceq [j_1, j_2, \dots, j_\beta] \quad \text{se e somente se } i_s \preceq j_t,$$

para $s = 1, 2, \dots, \alpha$ e $t = 1, 2, \dots, \beta$. A ordem *loop* para $k = 4$ é caracterizada por $1 \preceq [2, 3] \preceq 4$, e μ é isotônico com relação a essa ordem se e somente se $\mu_1 \preceq [\mu_2, \mu_3] \preceq \mu_4$. Nesse caso

$$P(4, 4; w) = P(Y_1 \leq Y_2 \leq Y_3 \leq Y_4) + P(Y_1 \leq Y_3 \leq Y_2 \leq Y_4),$$

as duas últimas probabilidades são P_3 com $w = (w_1, w_2, w_3, w_4)$ e $w' = (w_1, w_3, w_2, w_4)$. As $P(\ell, 4; w)$'s para $\ell < k$ podem ser calculadas pela aplicação da fórmula recorrente dada a seguir.

Para uma ordem arbitrária, $P(k, k; w)$ é a soma de $P[Y_{\pi(1)} < Y_{\pi(2)} \cdots < Y_{\pi(k)}]$ sobre determinado π , permutações de $1, 2, \dots, k$. Porém, se μ é isotônica com relação a $\preceq$ o que implica que $\mu_i \preceq \mu_j$ e $\mu_j \preceq \mu_i$ para $1 \leq i \neq j \leq k$, então $[M = k] = \emptyset$ e $P(k, k; w) = 0$. A seguir desenvolvemos a fórmula recorrente.

Seja $1 \leq m \leq k$ e seja $B_1, B_2, \dots, B_m$ uma partição de X em conjuntos de nível (não necessariamente ordenados). Para um m fixado, seja $\mathcal{L}_{mk}$ a coleção de todas as decomposições de X em conjuntos de nível (não necessariamente ordenados). Para A um subconjunto não vazio de X , seja $W_A = \sum_{i \in A} w_i$, $C_A = \text{card}(A)$ e para $A = \{i_1, i_2, \dots, i_j\}$ com $1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq k$ e $j = C_A$, seja $w(A) = (w_{i_1}, w_{i_2}, \dots, w_{i_j})^t$. Para uma dada decomposição, definamos $\preceq'$ sobre $1, 2, \dots, m$ por $i \preceq' j$ se existe $s \in B_i$ e $t \in B_j$ com $s \preceq t$ e sobre cada B_i , $\preceq_i$ denota a restrição de $\preceq$ a B_i . Além disso, seja $P(m, m; W_{B_1}, W_{B_2}, \dots, W_{B_m})$ a probabilidade de m conjuntos de nível com ordem parcial $\preceq'$ e vetor de pesos $(W_{B_1}, W_{B_2}, \dots, W_{B_m})^t$ e seja $P(1, C_{B_i}; w(B_i))$ a probabilidade de um conjunto de nível com ordem parcial $\preceq_i$ e vetor de pesos $w(B_i)$. Antes de enunciar o teorema seguinte notamos que as demonstrações de teoremas, lemas e corolários enunciados nesta seção podem ser vistas em Robertson, Wright and Dykstra (1988).

Teorema A.1: para $m \in \{2, 3, \dots, k-1\}$,

$$P(m, k; w) = \sum_{(B_1, \dots, B_m) \in \mathcal{L}_{mk}} P(m, m; W_{B_1}, W_{B_2}, \dots, W_{B_m}) \prod_{i=1}^m P(1, C_{B_i}; w(B_i)). \quad (\text{A.4})$$

Embora (A.4) seja uma fórmula para calcular $P(\ell, k; w)$, pode ser tedioso seu uso sendo mais recomendado o cálculo de $P(k, k; w)$.

A.0.2 Ordem simples com pesos arbitrários

O caso especial de maior importância é aquele de ordem simples, isto é, $H_1 : \mu_1 \leq \mu_2 \leq \dots \leq \mu_k$. Denotemos por $P_s(\ell, k; w)$ as probabilidades de nível com ordem simples. Quando $k = 2$ temos

$$P_s(1, 2; w) = P_s(2, 2; w) = \frac{1}{2}.$$

Quando $k = 3$

$$P_s(3, 3; w) = P_2 = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12}$$

e usando (A.4)

$$\begin{aligned} P_s(2, 3; w) &= P_s(2, 2; w_1 + w_2, w_3) P_s(1, 2; w_1, w_2) \\ &\quad + P_s(2, 2; w_1, w_2 + w_3) P_s(1, 2; w_2, w_3) = \frac{1}{2} \end{aligned}$$

e

$$P_s(1, 3; w) = \frac{1}{4} - \frac{1}{2\pi} \sin^{-1} \rho_{12}.$$

Para $k = 4$, $P_s(4, 4; w) = P_3$ é dado por (A.3) e

$$\begin{aligned} P_s(3, 4; w) &= P_s(3, 3; w_1, w_2, w_3 + w_4) P_s(1, 2; w_3, w_4) \\ &\quad + P_s(3, 3; w_1, w_2 + w_3, w_4) P_s(1, 2; w_2, w_3) \\ &\quad + P_s(3, 3; w_1 + w_2, w_3, w_4) P_s(1, 2; w_1, w_2). \end{aligned}$$

O segundo termo em todos esses produtos é igual a $\frac{1}{2}$ e o primeiro termo é obtido pela fórmula para $P_s(3, 3; w)$. Por exemplo

$$P_s(3, 3; w_1, w_2, w_3 + w_4) = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho$$

com

$$\rho = - \left[\frac{w_1(w_3 + w_4)}{(w_1 + w_2)(w_2 + w_3 + w_4)} \right]^{\frac{1}{2}} = \frac{\rho_{12}}{[1 - \rho_{23}^2]^{\frac{1}{2}}}.$$

Como $\rho_{13} = 0$, ρ é o coeficiente de correlação parcial entre Y_1 e Y_2 dado Y_3 , o qual será denotado por $\rho_{12.3}$. Então

$$P_s(3, 4; w) = \frac{3}{8} + \frac{1}{4\pi} (\sin^{-1} \rho_{12.3} + \sin^{-1} \rho_{13.2} + \sin^{-1} \rho_{23.1}).$$

De novo aplicando (A.4), notamos que $\mathcal{L}_{24}$ consiste das decomposições $\{1\}$, $\{2, 3, 4\}$; $\{1, 2, 3\}$, $\{4\}$ e $\{1, 2\}$, $\{3, 4\}$, e lembrando que $P_s(2, 2; w) = P_s(1, 2; w) = \frac{1}{2}$, temos que

$$\begin{aligned} P_s(2, 4; w) &= \frac{1}{2} P_s(1, 3; w_2, w_3, w_4) + \frac{1}{2} P_s(1, 3; w_1, w_2, w_3) + \frac{1}{8} \\ &= \frac{3}{8} - \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{23}) \\ &= \frac{1}{2} - P_s(4, 4; w). \end{aligned}$$

De fato, isso implica que $P_s(1, 4; w) = \frac{1}{2} - P_s(3, 4; w)$. No caso de $k = 5$, $P_s(5, 5; w)$ é obtido pela aplicação de (A.3) com as tabelas de Abrahamsom ou integração numérica. Em Bartholomew (1959b) as seguintes expressões para $P_s(\ell, 5; w)$ são dadas:

$$\begin{aligned} P_s(4, 5; w) &= \frac{1}{4} + \frac{1}{8\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{34}) + \frac{1}{8\pi} (\sin^{-1} \rho_{12.3} + \sin^{-1} \rho_{23.1} \\ &\quad + \sin^{-1} \rho_{34.2} + \sin^{-1} \rho_{13.2} + \sin^{-1} \rho_{23.4} + \sin^{-1} \rho_{24.3}), \\ P_s(3, 5; w) &= \frac{3}{8} + \frac{1}{8\pi} (\sin^{-1} \rho_{12.34} + \sin^{-1} \rho_{13.24} + \sin^{-1} \rho_{14.23} + \sin^{-1} \rho_{23.14} \\ &\quad + \sin^{-1} \rho_{24.13} + \sin^{-1} \rho_{34.12}) - \frac{1}{8\pi} (\sin^{-1} \rho_{12} \end{aligned}$$

$$\begin{aligned}
& + \sin^{-1} \rho_{23} + \sin^{-1} \rho_{34}) - \frac{1}{4\pi^2} (\sin^{-1} \rho_{12} \sin^{-1} \rho_{34.12} \\
& + \sin^{-1} \rho_{23} \sin^{-1} \rho_{14.23} + \sin^{-1} \rho_{34} \sin^{-1} \rho_{12.34}), \\
P_s(2, 5; w) &= \frac{1}{2} - P_s(4, 5; w)
\end{aligned}$$

e

$$P_s(1, 5; w) = \frac{1}{2} - P_s(5, 5; w) - P_s(3, 5; w).$$

Para $2 \leq k \leq 5$ e ordem simples, a probabilidade de um número par de conjuntos de nível é a mesma de um número ímpar de conjuntos de nível, isto é

$$\sum_{\ell \text{ par}} P_s(\ell, k; w) = \sum_{\ell \text{ ímpar}} P_s(\ell, k; w) = \frac{1}{2}.$$

Isso também pode valer para k grande e pesos iguais. Baseados nessa observação e no cálculo de probabilidades de nível para várias ordens, tem sido conjecturado que esse resultado é verdadeiro para uma ordem arbitrária e $k \geq 2$ (vide Shapiro 1987a).

A.0.3 Ordem simples com pesos iguais

Para o caso de ordem simples, as expressões para as probabilidades de nível simplificam se os pesos são iguais. De fato, nesse caso é claro que a probabilidade de k níveis é $1/k!$. Se os pesos são iguais eliminamos w da notação, ou seja, as probabilidades de nível serão denotadas por $P_s(\ell, k)$. Como $P_s(\ell, k; w)$ não é modificada se os pesos são multiplicados por uma constante positiva, podemos assumir que $Y_1, Y_2, \dots, Y_k$, na definição das probabilidades de nível, são variáveis aleatórias independentes normais padrão. Será demonstrado que $P_s(\ell, k)$ tem distribuição livre na classe das variáveis aleatórias contínuas

independentes, identicamente distribuídas, isto é, a probabilidade de que a regressão isotônica de $Y_1, Y_2, \dots, Y_k$ com ordem simples e pesos iguais tenha ℓ conjuntos de nível não depende da distribuição de Y_i se assumimos que elas são independentes com uma distribuição contínua comum. Uma expressão para a função geradora de probabilidades de $\{P_s(\ell, k)\}$ será obtida e usada para derivar uma relação de recorrência para as probabilidades de nível com pesos iguais. A prova do lema dado a seguir é uma adaptação devida a Barton e Mallows (1961), da prova dada por Miles (1959) e como foi observado antes esta prova pode ser vista em Robertsom et al. (1988).

Com $g = (g_1, g_2, \dots, g_k)^t$ e $w = (w_1, w_2, \dots, w_k)^t$ vetores fixados de reais com $w_i > 0$, seja $r(\ell, k)$ o número de permutações π , de $(1, 2, \dots, k)$ para o qual um algoritmo, por exemplo o PAVA, aplicado a $(g_{\pi(1)}, g_{\pi(2)}, \dots, g_{\pi(k)})^t$ com vetor de pesos $(w_{\pi(1)}, w_{\pi(2)}, \dots, w_{\pi(k)})^t$ produz ℓ conjuntos de nível. O seguinte resultado originalmente devido a Miles demonstra que sobre uma condição simples em g e w , $r(\ell, k)$ não depende de g ou w e a função geradora de $r(\ell, k)$ tem uma forma simples.

Lema A.1: se g e w são tais que nenhum par de médias para os pesos de conjuntos diferentes de g_i com w_i são iguais, então a função geradora de $r(\ell, k)$ é dada por

$$\sum_{\ell=1}^k r(\ell, k) s^\ell = s(s+1) \dots (s+k-1). \quad (\text{A.5})$$

Teorema A.2: sejam $Y_1, Y_2, \dots, Y_k$ variáveis aleatórias contínuas, independentes identicamente distribuídas e seja $P_s(\ell, k)$ a probabilidade de que a regressão isotônica de $Y = (Y_1, Y_2, \dots, Y_k)^t$ com ordem simples e pesos iguais tenha ℓ conjuntos de nível. A função geradora de probabilidades de $\{P_s(\ell, k)\}$

é dada por

$$P_k(s) = \sum_{\ell=1}^k P_s(\ell, k) s^\ell = s(s+1) \dots (s+k-1)/k!. \quad (\text{A.6})$$

Corolário A.1: as probabilidades $P_s(\ell, k)$ satisfazem $P_s(1, k) = 1/k$, $P_s(k, k) = 1/k!$ e

$$P_s(\ell, k) = \frac{1}{k} P_s(\ell-1, k-1) + \frac{k-1}{k} P_s(\ell, k-1), \quad \ell = 2, 3, \dots, k-1. \quad (\text{A.7})$$

Corolário A.2: as probabilidades $P_s(\ell, k)$ são dadas por

$$P_s(\ell, k) = \frac{|s_k^\ell|}{k!} \quad \text{para } \ell = 1, 2, \dots, k.$$

Valores numéricos para $P_s(\ell, k)$ para $k \leq 20$ podem ser encontrados no Apêndice de Robertsom et al. (1988).

A.0.4 Ordem simples na forma de árvore

Para fazer comparações de vários tratamentos com um controle, a ordem simples em forma de árvore, $\mu_0 \leq [\mu_1, \mu_2, \dots, \mu_k]$, é de importância prática. As probabilidades de nível para ordem simples em forma de árvore são denotadas por $P_T(\ell, k+1; w)$. Usando o método discutido na seção anterior, $P_T(\ell, k+1; w)$ pode ser calculado ao considerar $P(Y_{\pi(0)}, Y_{\pi(1)}, \dots, Y_{\pi(k)})$ sobre as $k!$ permutações π , de $(0, 1, \dots, k)$, as quais satisfazem $\pi(0) = 0$. A aproximação dada a seguir é menos complicada.

Sejam $Y_0, Y_1, \dots, Y_k$ variáveis aleatórias independentes normais com média comum e variâncias $w_0^{-1}, w_1^{-1}, \dots, w_k^{-1}$, respectivamente. Para $i = 1, 2, \dots, k$, seja $U_i = Y_i - Y_0$ e notemos que

$$P_T(k+1, k+1; w) = P(U_1 \geq 0, U_2 \geq 0, \dots, U_k \geq 0).$$

Essa probabilidade é a probabilidade do quadrante P_k discutida anteriormente, mas nesse caso a matriz de correlação satisfaz ($\rho_{ii} = 1$),

$$\rho_{ij} = \left[\frac{w_i w_j}{(w_0 + w_i)(w_0 + w_j)} \right]^{\frac{1}{2}}, \quad 1 \leq i \neq j \leq k. \quad (\text{A.8})$$

Assim temos que

$$\begin{aligned} P_1 &= \frac{1}{2} \\ P_2 &= \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12} \end{aligned}$$

e

$$P_3 = \frac{1}{8} + \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{13} + \sin^{-1} \rho_{23}). \quad (\text{A.9})$$

Para $k > 3$, os P'_k s são menos complexos do que para o caso de ordem simples, uma vez que $\rho_{ij} = \lambda_i \lambda_j$ para $i \neq j$ com $\lambda_i = [w_i / (w_0 + w_i)]^{\frac{1}{2}}$ e aplicando o resultado de Dunnett e Sobel (1955) pode-se mostrar que

$$P_k = \int_{-\infty}^{\infty} \left[\prod_{i=1}^k \Phi \left\{ \frac{\lambda_i x}{(1 - \lambda_i^2)^{\frac{1}{2}}} \right\} \right] \phi(x) dx, \quad (\text{A.10})$$

onde Φ e ϕ são a função de distribuição e a função de densidade, respectivamente, de uma variável normal padrão.

De fato, para $k = 2$, $P_T(1, 2; w) = P_T(2, 2; w) = \frac{1}{2}$. Para $k = 3$, $P_T(3, 3; w) = P_2$,

$$\begin{aligned} P_T(2, 3; w) &= P_T(2, 2; w_0 + w_1, w_2) P_T(1, 2; w_0, w_1) \\ &\quad + P_T(2, 2; w_0 + w_2, w_1) P_T(1, 2; w_0, w_2) \\ &= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} \text{ e} \\ P_T(1, 3; w) &= \frac{1}{2} - P_T(3, 3; w). \end{aligned}$$

Para $k = 4$, $P_T(4, 4; w) = P_3$ e Bartholomew (1961a) tem mostrado que

$$\begin{aligned} P_T(3, 4; w) &= \frac{3}{8} + \frac{1}{4\pi}(\sin^{-1} \rho_{12.3} + \sin^{-1} \rho_{13.2} + \sin^{-1} \rho_{23.1}), \\ P_T(2, 4; w) &= \frac{1}{2} - P_T(4, 4; w) \text{ e} \\ P_T(1, 4; w) &= \frac{1}{2} - P_T(3, 4; w). \end{aligned}$$

Se os pesos são iguais, as probabilidades de nível para ordem simples em forma de árvore são menos complexas. Denotemos essas probabilidades por $P_T(\ell, k+1)$. A decomposição $\mathcal{L}_{\ell, k+1}$ consiste de $\{0\} \cup A$, com A um subconjunto de $\{1, 2, \dots, k\}$ de cardinalidade $k+1-\ell$ e $\ell-1$ subconjuntos unitários. Cada termo em (A.4) é o mesmo e assumindo que o peso comum é 1 temos que

$$P_T(\ell, k+1) = \binom{k}{\ell-1} P_T(1, k-\ell+2) P_T(\ell, \ell; (k-\ell+2), 1, \dots, 1). \quad (\text{A.11})$$

O último termo em (A.11) foi tabelado por Ruben (1954) para ℓ e k com $0 \leq k \leq 49$ e $k-9 \leq \ell \leq k+1$. Usando integração numérica ou as tabelas de Ruben, $P_T(\ell, k+1)$ são facilmente obtidas.

A.0.5 Ordem do tipo *loop*

A restrição $\mu_1 \leq [\mu_2, \mu_3, \dots, \mu_{k-1}] \leq \mu_k$ é chamada de ordem do tipo *loop*. Para $k = 3$ temos o caso de ordem simples. As probabilidades de nível denotadas por $P_L(\ell, k; w)$, serão obtidas a seguir para $k = 4$. Usando (A.3) calculamos as probabilidades do quadrante,

$$\begin{aligned} P_L(4, 4; w) &= P(Y_1 < Y_2 < Y_3 < Y_4) + P(Y_1 < Y_3 < Y_2 < Y_4) \\ &= \frac{1}{4\pi} \left(\frac{\pi}{2} - \theta_1 - \theta_3 + \frac{\pi}{2} - \theta_2 - \theta_4 \right), \end{aligned}$$

onde

$$\begin{aligned}\theta_1 &= \sin^{-1} \left[\frac{w_1 w_3}{(w_1 + w_2)(w_2 + w_3)} \right]^{\frac{1}{2}}, \\ \theta_2 &= \sin^{-1} \left[\frac{w_2 w_4}{(w_2 + w_3)(w_3 + w_4)} \right]^{\frac{1}{2}}\end{aligned}$$

e θ_3 é θ_1 , θ_4 é θ_2 com w_2 e w_3 trocados. Usando

$$\sin\left(\frac{\pi}{4} - \theta_1 - \theta_3\right) = \cos(\theta_1 + \theta_3) = \cos(\theta_1)\cos(\theta_3) - \sin(\theta_1)\sin(\theta_3),$$

temos que

$$P_L(4, 4; w) = \frac{1}{4\pi} (\sin^{-1} \tau_{21.13} + \sin^{-1} \tau_{24.43})$$

com

$$\tau_{21.13} = \left[\frac{w_2 w_3}{(w_1 + w_2)(w_1 + w_3)} \right]^{\frac{1}{2}} \quad \text{e} \quad \tau_{24.43} = \left[\frac{w_2 w_3}{(w_2 + w_4)(w_3 + w_4)} \right]^{\frac{1}{2}}.$$

Listando os elementos $\mathcal{L}_{34}$ e aplicando (A.4) e (A.3) obtemos

$$P_L(3, 4; w) = \frac{1}{2} + \frac{1}{4\pi} (\sin^{-1} \tau_{[12]3.34} + \sin^{-1} \tau_{[13]2.24} + \sin^{-1} \tau_{[24]3.31} + \sin^{-1} \tau_{[34]2.21})$$

com

$$\tau_{[rs]t.tu} = - \left[\frac{(w_r + w_s)w_u}{(w_r + w_s + w_t)(w_t + w_u)} \right]^{\frac{1}{2}}.$$

Os elementos de $\mathcal{L}_{24}$ são $\{1, 2\}, \{3, 4\}; \{1, 3\}, \{2, 4\}; \{1, 2, 3\}, \{4\}$ e $\{1\}, \{2, 3, 4\}$.

Os dois primeiros têm probabilidade $1/8$ cada um. A terceira decomposição tem probabilidade $\frac{1}{2} P_T(1, 3; w_1, w_2, w_3)$, isto é, a metade da probabilidade para um conjunto de nível com ordem simples em forma de árvore e $k = 2$. A probabilidade da quarta decomposição é a metade da probabilidade de um conjunto de nível com ordem simples em forma de árvore invertido. Ordem simples em forma de árvore invertido é o negativo da ordem simples em forma de árvore e

$$P_w(X | -\mathcal{H}) = -P_w(-X | \mathcal{H}).$$

Então , a probabilidade da quarta decomposição é a metade da probabilidade de um conjunto de nível com ordem simples em forma de árvore e $k = 2$ e o vetor de pesos $w = (w_4, w_2, w_3)^t$, isto é, $\frac{1}{2}P_T(1, 3; w_4, w_2, w_3)$. Agora é imediato mostrar que $P_L(2, 4; w) = \frac{1}{2} - P_L(4, 4; w)$, do qual segue que $P_L(1, 4; w) = \frac{1}{2} - P_L(3, 4; w)$. Se os pesos são iguais, então

$$\begin{aligned} P_L(4, 4) &= \frac{1}{12}, \\ P_L(3, 4) &= \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \frac{1}{\sqrt{3}} = 0.3041, \\ P_L(2, 4) &= \frac{5}{12} \text{ e} \\ P_L(1, 4) &= \frac{1}{\pi} \sin^{-1} \frac{1}{\sqrt{3}} = 0.1959. \end{aligned}$$

Agora apresentamos em detalhe o cálculo das probabilidades de nível para o caso de cinco populações e pesos arbitrários (vide Singh e Wright, 1993).

Temos seis subcasos nos quais cinco conjuntos de nível ordenados podem ocorrer. Eles são os seguintes: (a) $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$; (b) $\{1\}, \{2\}, \{4\}, \{3\}, \{5\}$; (c) $\{1\}, \{3\}, \{2\}, \{4\}, \{5\}$; (d) $\{1\}, \{3\}, \{4\}, \{2\}, \{5\}$; (e) $\{1\}, \{4\}, \{2\}, \{3\}, \{5\}$; e (f) $\{1\}, \{4\}, \{3\}, \{2\}, \{5\}$. No caso (a) a contribuição para as probabilidades de nível, $P_L(5, 5; w)$ é dada por

$$\frac{1}{16} + \frac{1}{8\pi} f(\rho) + \frac{1}{4\pi^2} \int_0^{-\rho_{34}} (1-x^2)^{-\frac{1}{2}} \sin^{-1} \left[-\frac{(1-x^2)^{\frac{1}{2}} \rho_{12}}{(1-x^2-\rho_{23}^2)^{\frac{1}{2}}} \right] \partial x \quad (\text{A.12})$$

onde

$$\begin{aligned} f(\rho) &= (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{23} + \sin^{-1} \rho_{34}) \text{ e} \\ \rho_{i,i+1} &= \rho_{i+1,i} = - \left[\frac{w_i w_{i+1}}{(w_i + w_{i+1})(w_{i+1} + w_{i+2})} \right]^{\frac{1}{2}}, \end{aligned} \quad (\text{A.13})$$

para $i = 1, 2, 3$. A contribuição para as probabilidades de nível do subcaso (b) pode ser obtida do subcaso anterior trocando w_3 e w_4 ; e a contribuição dos

restantes quatro subcasos podem ser obtidos de forma similar. A probabilidade de nível $P_L(5, 5; w)$ é a soma dessas seis probabilidades.

Temos doze subcasos nos quais quatro conjuntos de nível ordenados podem ocorrer e eles são divididos em seis pares de casos os quais são convenientes para obter $P_L(4, 5; w)$: (a) (i) $\{1, 2\}, \{3\}, \{4\}, \{5\}$ e (ii) $\{1, 2\}, \{4\}, \{3\}, \{5\}$; (b) (i) $\{1, 3\}, \{2\}, \{4\}, \{5\}$ e (ii) $\{1, 3\}, \{4\}, \{2\}, \{5\}$; (c) (i) $\{1, 4\}, \{2\}, \{3\}, \{5\}$ e (ii) $\{1, 4\}, \{3\}, \{2\}, \{5\}$; (d) (i) $\{1\}, \{2\}, \{3\}, \{4, 5\}$ e (ii) $\{1\}, \{3\}, \{2\}, \{4, 5\}$; (e) (i) $\{1\}, \{2\}, \{4\}, \{3, 5\}$ e (ii) $\{1\}, \{4\}, \{2\}, \{3, 5\}$; e (f) (i) $\{1\}, \{3\}, \{4\}, \{2, 5\}$ e (ii) $\{1\}, \{4\}, \{3\}, \{2, 5\}$. A probabilidade correspondente ao caso (a) é dada por

$$P_s(1, 2; w_1, w_2)[P_s(4, 4; w_1 + w_2, w_3, w_4, w_5) + P_s(4, 4; w_1 + w_2, w_4, w_3, w_5)] \\ = \frac{1}{8\pi} \left[\left(\frac{\pi}{2} - \theta_1 - \theta_3 \right) + \left(\frac{\pi}{2} - \theta_2 - \theta_4 \right) \right],$$

onde

$$\theta_1 = \sin^{-1} \left[\frac{(w_1 + w_2)w_4}{(w_1 + w_2 + w_3)(w_3 + w_4)} \right]^{\frac{1}{2}}, \\ \theta_2 = \sin^{-1} \left[\frac{w_3w_5}{(w_4 + w_5)(w_3 + w_4)} \right]^{\frac{1}{2}}$$

e $\theta_3 = \theta_1$ e $\theta_4 = \theta_2$ com w_3 e w_4 trocados. Usando $\sin(\frac{\pi}{4} - \theta_1 - \theta_3) = \cos \theta_1 \cos \theta_3 - \sin \theta_1 \sin \theta_3$ a probabilidade anterior pode ser escrita como

$$\frac{1}{8\pi} (\sin_{\tau_{3(1,2),(1,2)4}}^{-1} + \sin_{\tau_{35,54}}^{-1}),$$

onde

$$\tau_{r(u,v),(u,v)s} = \left[\frac{w_r w_s}{(w_u + w_v + w_r)(w_u + w_v + w_s)} \right]^{\frac{1}{2}} \quad (\text{A.14})$$

e

$$\tau_{ru,us} = \left[\frac{w_r w_s}{(w_u + w_r)(w_u + w_s)} \right]^{\frac{1}{2}}. \quad (\text{A.15})$$

As probabilidades dos restantes cinco pares de casos podem ser obtidas similarmemente. Então

$$\begin{aligned}
 P_L(4, 5; w) = & \frac{1}{8\pi} [\sin_{\tau_{3(1,2),(1,2)4}}^{-1} + \sin_{\tau_{35,54}}^{-1} + \sin_{\tau_{2(1,3),(1,3)4}}^{-1} \\
 & + \sin_{\tau_{25,54}}^{-1} + \sin_{\tau_{2(1,4),(1,4)3}}^{-1} + \sin_{\tau_{25,53}}^{-1} \\
 & + \sin_{\tau_{3(5,4),(5,4)2}}^{-1} + \sin_{\tau_{31,12}}^{-1} + \sin_{\tau_{4(5,3),(5,3)2}}^{-1} \\
 & + \sin_{\tau_{41,12}}^{-1} + \sin_{\tau_{4(5,2),(5,2)3}}^{-1} + \sin_{\tau_{41,13}}^{-1}].
 \end{aligned} \tag{A.16}$$

Os seguintes são os doze subcasos nos quais três conjuntos de nível podem ocorrer: (a) $\{1, 2, 3\}, \{4\}, \{5\}$; (b) $\{1, 2, 4\}, \{3\}, \{5\}$; (c) $\{1, 3, 4\}, \{2\}, \{5\}$; (d) $\{1\}, \{2\}, \{3, 4, 5\}$; (e) $\{1\}, \{3\}, \{2, 4, 5\}$; (f) $\{1\}, \{4\}, \{2, 3, 5\}$; (g) $\{1, 3\}, \{2\}, \{4, 5\}$; (h) $\{1, 4\}, \{2\}, \{3, 5\}$; (i) $\{1, 2\}, \{3\}, \{4, 5\}$; (j) $\{1, 4\}, \{3\}, \{2, 5\}$; (k) $\{1, 2\}, \{4\}, \{3, 5\}$; e (l) $\{1, 3\}, \{4\}, \{2, 5\}$.

A contribuição do subcaso (a) para as probabilidades de nível $P_L(3, 5; w)$ é dada por

$$\begin{aligned}
 & P_T(1, 3; w_1, w_2, w_3) P_s(3, 3; w_1 + w_2 + w_3, w_4, w_5) \\
 &= \left(\frac{1}{2} - P_T(1, 3; w_1, w_2, w_3) \right) \left(\frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12}^s \right) \\
 &= \left(\frac{1}{4} - \frac{1}{2\pi} \sin^{-1} \rho_{12}^T \right) \left(\frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12}^s \right),
 \end{aligned}$$

onde

$$\begin{aligned}
 \rho_{12}^T &= \left[\frac{w_2 w_3}{(w_1 + w_2)(w_1 + w_3)} \right]^{\frac{1}{2}} \\
 \rho_{12}^s &= - \left[\frac{w_1^* w_3^*}{(w_1^* + w_2^*)(w_2^* + w_3^*)} \right]^{\frac{1}{2}}
 \end{aligned} \tag{A.17}$$

e $w_1^* = w_1 + w_2 + w_3$, $w_2^* = w_4$ e $w_3^* = w_5$. As contribuições para as probabilidades de nível dos subcasos (b) e (c) são obtidas de (a) trocando w_3 e

w_4 , w_2 e w_4 , respectivamente. Similarmente, as contribuições de (d), (e) e (f) são obtidas de (a), (b) e (c) com as seguintes substituições: $w_1 = w_5$, $w_2 = w_4$, $w_3 = w_3$ e $w_5 = w_1$. A contribuição para $P_L(3, 5; w)$ do subcaso (g) é dada por

$$P_s(1, 2; w_1, w_3)P(1, 2; w_4, w_5)P_s(3, 3; w_1 + w_3, w_2, w_4 + w_5) \\ = \frac{1}{4} \left(\frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_{12}^s \right),$$

onde ρ_{12}^s é dada por (A.17) com as substituições $w_1^* = w_1 + w_3$, $w_2^* = w_2$ e $w_3^* = w_4 + w_5$. As contribuições para a probabilidade dos subcasos de (h) até (l) são obtidas de forma similar. Assim, a probabilidade de nível $P_L(3, 5; w)$ é igual à soma de todas essas contribuições. Como $\frac{1}{2} = P_L(1, 5; w) + P_L(3, 5; w) + P_L(5, 5; w) = P_L(2, 5; w) + P_L(4, 5; w)$ obtemos

$$P_L(2, 5; w) = \frac{1}{2} - P_L(4, 5; w) \text{ e} \\ P_L(1, 5; w) = \frac{1}{2} - P_L(3, 5; w) - P_L(5, 5; w).$$

A.1 Aproximações para as probabilidades de nível

Devido à complexidade das distribuições dos testes restritos, aproximações para essas distribuições têm sido de considerável interesse. Mesmo que as probabilidades de nível sejam conhecidas, o nível descritivo correspondente a um valor observado de um desses testes pode ser complexo de calcular. Para o caso de tamanhos amostrais iguais, Bartholomew (1959a, 1959b) propôs uma aproximação através de dois momentos da distribuição qui-quadrado para a distribuição nula de $\bar{\chi}_{01}^2$, e Sasabuchi e Kulatunga (1985) desenvolveram uma

aproximação através do segundo e terceiro momentos da distribuição beta para $\bar{E}_{01}^2$. Esse trabalho foi estendido por Singh e Wright (1986) para ser aplicado a $\bar{\chi}_{12}^2$ e $\bar{E}_{12}^2$ e os coeficientes necessários para aplicar essas aproximações para ordem simples e ordem simples em forma de árvore podem ser encontrados em Robertsom et al. (1988).

Siskind (1976) considerando o caso de ordem simples, aproximou a distribuição nula de $\bar{\chi}_{01}^2$ usando uma expansão em séries truncadas de Taylor no ponto $\rho_{i,i+1} = \frac{1}{2}$, o qual é o valor para pesos iguais de $\rho_{i,i+1}$. Essa aproximação para $P(\ell, k; w)$ depende dos pesos somente através da matriz R de correlações parciais e tem a vantagem de ser conservadora, mas tem sido desenvolvida unicamente para $\bar{\chi}_{01}^2$ com ordem simples e $k \leq 8$. A aproximação de Siskind parece ser mais fácil para obter valores muito menores do que a aproximação baseada em padrões de grandes e pequenos pesos e o fato da aproximação de Siskind ser válida unicamente quando $k \leq 8$ é uma desvantagem.

Siskind(1976) e Grove(1980) conjecturaram que a distribuição nula com pesos iguais para os testes restritos envolvidos, dão uma razoável aproximação para os casos de tamanhos amostrais desiguais se os tamanhos amostrais não são muito diferentes. Essa conjectura foi estabelecida em Robertson e Wright(1983), Wright e Tran(1985), Lucas et al. (1989) e Singh e Wright(1993) para os casos de ordem simples, ordem simples em forma de árvore, ordem unimodal e ordem do tipo *loop*, respectivamente.

Chase (1974), em um artigo fundamental para o estudo dessas aproximações, considerou o caso de ordem simples. Ele propõe interpolação entre os valores críticos para pesos iguais e aqueles da distribuição limite obtida ao deixar a

precisão ir para infinito. Chase encontrou uma boa aproximação para o caso no qual estamos interessados em comparar dosagens crescentes de uma droga com uma dose controle. Indexando a droga controle como um, desenvolveu a aproximação para o caso $w_2 = \dots = w_k$ e $w_1/w_2 > 1$. Primeiro, ele encontrou uma fórmula recorrente para $P(\ell, k; \infty)$, o valor limite para $P(\ell, k; w)$ quando $w_1 \rightarrow 0$ e o valor fixado para $w_2, \dots, w_k$. Chase tabulou valores críticos para as distribuições $\bar{\chi}^2$ e $\bar{E}^2$ associadas com esses valores limite. Para valores de w_1/w_2 entre um e ∞ ele encontrou uma interpolação linear em $(w_1/w_2)^{-\frac{1}{2}}$ que fornece boas aproximações. Por exemplo, se C_∞ denota o valor crítico apropriado calculado dessas probabilidades de nível limites e se C_1 denota o valor crítico apropriado de pesos iguais, então o valor crítico apropriado é dado por $C = C_\infty + (w_1/w_2)^{-\frac{1}{2}}(C_1 - C_\infty)$. Chase oferece avaliações de sua aproximação, exatas para $k \leq 4$ e pelo método de Monte Carlo para $k > 4$ e conclui que seus resultados são completamente satisfatórios. Sua aplicação ao caso de ordem simples confirma essa conclusão.

Robertson e Wright (1982b) obtiveram uma fórmula recursiva para as probabilidades de nível limites e generalizaram o trabalho de Chase ao admitirem precisão crescente em qualquer uma das amostras.

Seguindo as idéias de Chase para o caso de ordem simples, Robertson e Wright (1983) obtêm uma aproximação baseada em modelos de grandes e pequenos pesos. Resultados similares para ordem simples em forma de árvore são dados em Robertson e Wright (1985) e Wright e Tran (1985), ordem unimodal em Lucas et al. (1989) e ordem do tipo *loop* em Singh e Wright (1993).

Robertson e Wright (1982a) derivam limites superiores e inferiores para as

distribuições $\bar{\chi}^2$ e $\bar{E}^2$ e provam que em geral, esses limites não podem ser melhorados. O limite inferior é aproximado usando um conjunto de pesos em forma de $\cup$ com valores $(a, \epsilon, \dots, \epsilon, b)$, onde $a, b > 0$ e $\epsilon \rightarrow 0$. O limite superior é aproximado usando uma $\cap$ -configuração para os pesos na forma de $(\epsilon^H, \epsilon^{H-1}, \dots, \epsilon, 1, \epsilon, \dots, \epsilon^{H-1}, \epsilon^H)$ e fazendo $\epsilon \rightarrow 0$. Isto é, para k ímpar, onde $k = 2H + 1$. Ajustes óbvios são feitos para o caso de k par. Esses limites podem diferir substancialmente. Por exemplo, para $k = 5$ o valor crítico de 5% para pesos iguais é 5.049. Os limites superior e inferior para $P_w(\bar{\chi}_5^2 \geq 5.049)$ são 0.013 e 0.096, respectivamente. A partir da fórmula recorrente de Chase (1974), Robertsom e Wright (1982b) obtiveram a função geradora para $P(\ell, k; \infty)$,

$$\Phi_{\infty, k}(k) = \sum_{\ell=1}^k P(\ell, k; \infty) s^\ell = s \left(\frac{\frac{1}{2}s + k - \frac{1}{2}}{k-1} \right),$$

a qual produz uma fórmula recorrente e precisa para as probabilidades de nível limites, isto é

$$\begin{aligned} P(1, k+1; \infty) &= \frac{1}{2}(2k-1)P(1, k; \infty)/k = (2k)!/\{2^{2k}(k!)^2\}, \\ P(k+1, k+1; \infty) &= \frac{1}{2}P(k, k; \infty)/k = 2^{-k}/k! \text{ e} \\ P(\ell, k+1; \infty) &= \frac{1}{2k}\{P(\ell, k; \infty) + (2k-1)P(\ell, k; \infty)\}. \end{aligned}$$

Uma rotina simples pode ser usada para gerar as $P(\ell, k; \infty)$'s para qualquer k razoável. Para $k \leq 12$, elas são tabeladas por Robertson e Wright (1982b). Eles também notaram que os valores limite de $P(\ell, k; w)$ para a situação na qual $w_1 = \dots = w_{k-1}$ e $\frac{w_k}{w_1} \rightarrow \infty$ também são dadas como $P(\ell, k; \infty)$.

Na aproximação de $P(\ell, k; w)$ para o caso de ordem simples baseada em padrões de grandes e pequenos pesos desenvolvida por Robertsom e Wright

(1983), destacamos os seguintes comentários: primeiro são obtidos os valores limite para $P(\ell, k; \infty)$ para cada situação na qual os w_i assumem, em particular, dois valores, w_i permanece constante e outro aproxima-se a ∞ . Essa aproximação é uma interpolação entre o caso limite apropriado e o caso de pesos iguais. Robertsom e Wright (1983) sugerem uma aproximação que é uma interpolação entre os valores críticos baseados em coeficientes limite e coeficientes para o caso de pesos iguais. Depois de extensivo estudo, a sugestão é chamar de pesos grandes se eles excedem

$$\bar{w} = 0.65 \min(w_1, \dots, w_k) + 0.35 \max(w_1, \dots, w_k)$$

e interpolar na base de λ igual à raiz cúbica do quociente entre a média de pesos pequenos e a média de pesos grandes. Essa aproximação também foi testada em conjuntos de pesos com $\cup$ e $\cap$ -configuração. Para $w = (1, \epsilon, \epsilon, \epsilon, 1)^t$ e $w = (\epsilon, \sqrt{\epsilon}, 1, \sqrt{\epsilon}, \epsilon)^t$ e $a_{0.05}(w)$, os valores críticos aproximados são dados na Tabela 1 de Robertsom e Wright (1983). A aproximação desempenhou-se de forma satisfatória para conjuntos de pesos com $\cup$ -configuração.

Wright e Tran (1985) consideram aproximações para as probabilidades de nível necessárias para implementar o teste da razão de verossimilhança para ordem simples em forma de árvore e pesos desiguais. Seguindo o trabalho de Robertsom e Wright (1983) para ordem simples, Wright e Tran (1985) estudaram os coeficientes ou probabilidades de nível de pesos iguais como uma aproximação e desenvolveram uma aproximação baseada em modelos de grandes e pequenos pesos para a ordem simples em forma de árvore. A precisão da aproximação de pesos iguais foi considerada para casos particulares de conjuntos de pesos como também para conjuntos de pesos gerados aleatoriamente. Similarmente ao caso de ordem simples, eles demonstraram que para muitas situações de interesse a aproximação por pesos iguais é adequada. Porém, para

grandes variações nos pesos e determinados modelos de grandes e pequenos pesos essa aproximação pode conter considerável erro. Seguindo as idéias de Chase (1974), Wright e Tran obtêm as probabilidades de nível e a partir desses resultados propõem uma aproximação de pesos iguais para conjuntos de pesos com grande quantidade de variação.

Concluindo, Wright e Tran mostram que se existe moderada variação nos pesos a aproximação por pesos iguais é adequada, porém para tais conjuntos de pesos a aproximação baseada em Chase é mais precisa. Quando a variação no conjunto de pesos cresce, a melhora na aproximação baseada em modelos é maior do que aquela conseguida através de pesos iguais. É importante observar que em tais casos, isto é, quando os tamanhos amostrais são completamente diferentes, os efeitos de não-normalidade e heterocedasticidade são mais sérios. Se a suposição de normalidade é questionada, podemos usar um teste de postos. Nesse caso, a distribuição nula aproximada também depende dos $P(\ell, k; w)$'s [vide Robertsom e Wright 1985].

Apêndice B

Condições de regularidade de Bradley e Gart

B.1 Introdução

Seja $f_i(x_i, \theta)$, $i = 1, 2, \dots, g$, a função de densidade de probabilidade de X_i , (discreta ou contínua) onde X_i é um vetor aleatório com valores sobre uma região R_i a qual não depende de $\theta = (\theta_1, \dots, \theta_p)^t$ vetor de parâmetros desconhecidos definidos sobre um intervalo p -dimensional Θ . Não é necessário que cada f_i dependa de todos $\theta_1, \dots, \theta_p$. Sejam $x_{i1}, x_{i2}, \dots, x_{ir_i}$, r_i vetores independentes de observações de X_i , $i = 1, 2, \dots, g$. A função de verossimilhança conjunta é então

$$\phi = \prod_{i=1}^g \prod_{j=1}^{r_i} f_i(x_{ij}, \theta) \quad (\text{B.1})$$

e as equações normais são

$$\frac{\partial \log \phi}{\partial \theta_r} = 0, \quad r = 1, 2, \dots, p. \quad (\text{B.2})$$

Suponhamos que

- I(i) Para quase todo x_i em R_i e para todo $\theta \in \Theta$, as derivadas parciais

$$\frac{\partial \log f_i}{\partial \theta_r}, \quad \frac{\partial^2 \log f_i}{\partial \theta_r \partial \theta_s} \text{ e } \frac{\partial^3 \log f_i}{\partial \theta_r \partial \theta_s \partial \theta_t}$$

existem para $r, s, t = 1, 2, \dots, p$; $i = 1, 2, \dots, g$.

- I(ii) Para todas as densidades f_i , para quase todo x_i em R_i e para todo $\theta \in \Theta$

$$\left| \frac{\partial \log f_i}{\partial \theta_r} \right| < F_{ir}(x_i), \quad \left| \frac{\partial^2 \log f_i}{\partial \theta_r \partial \theta_s} \right| < F_{irs}(x_i) \text{ e } \left| \frac{\partial^3 \log f_i}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| < H_{irst}(x_i)$$

onde $F_{ir}(x_i)$ e $F_{irs}(x_i)$ são integráveis sobre R_i , M_i é uma constante positiva e $\int_{R_i} H_{irst}(x_i) f_i \partial x_i < M_i$, para $i = 1, 2, \dots, g$; $r, s, t = 1, 2, \dots, p$. Similarmente para todo f_i que seja função de probabilidade, $\sum_{x_i \in R_i} \partial f_i / \partial \theta_r$ e $\sum_{x_i \in R_i} \partial^2 f_i / \partial \theta_r \partial \theta_s$ convergem uniformemente para todo $\theta \in \Theta$ e

$$\left| \frac{\partial^3 \log f_i}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| < H_{irst}(x_i) \text{ onde } \sum_{x_i \in R_i} H_{irs}(x_i) f_i < M_i,$$

para todo $\theta \in \Theta$ e $r, s, t = 1, 2, \dots, p$; $i = 1, 2, \dots, g$. Essas suposições permitem trocar a ordem dos somatórios ou da diferenciação e integração.

- I(iii) Para todo $\theta \in \Theta$, a matriz $J = [J_{rs}(\theta)]$ com

$$J_{rs}(\theta) = \sum_{i=1}^g \mu_i \int_{R_i} \frac{\partial \log f_i}{\partial \theta_r} \frac{\partial \log f_i}{\partial \theta_s} f_i \partial x_i \quad (\text{B.3})$$

é positiva definida com determinante finito.

Para aquelas f_i que sejam funções de probabilidades discreta, definimos $\mu_i = r_i/n$ onde $n = \sum_{i=1}^g r_i$ e em (B.3) substituímos $\int_{R_i}$ por $\sum_{x_i \in R_i}$ e omitimos o diferencial.

B.2 Consistência

Teorema B.1: dada a função de verossimilhança (B.1) e as suposições I, se $\hat{\theta}$ representa a solução para as equações normais (B.2) e se θ_0 representa o verdadeiro valor de θ , então quando $r_i \rightarrow \infty$ e $\mu_i = r_i/n$ seja constante com $n = \sum_{i=1}^g r_i$, temos que $\hat{\theta}$ é um estimador consistente de θ_0 .

Prova: usando uma expansão de Taylor temos

$$\begin{aligned} \frac{\partial \log f}{\partial \theta_r} &= \left(\frac{\partial \log f}{\partial \theta_r} \right) (\theta_0) + \sum_{s=1}^p (\hat{\theta}_s - \theta_{0s}) \left(\frac{\partial^2 \log f}{\partial \theta_r \partial \theta_s} \right) (\theta_0) \\ &\quad + \frac{1}{2} \sum_{s,t=1}^p (\hat{\theta}_s - \theta_{0s})(\hat{\theta}_t - \theta_{0t}) \left(\frac{\partial^3 \log f}{\partial \theta_r \partial \theta_s \partial \theta_t} \right) (\theta^*). \end{aligned}$$

Daqui segue que se $\Gamma = \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{r_i} \partial \log f_{ij} / \partial \theta_r$, então

$$\begin{aligned} \Gamma &= \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{r_i} \left(\frac{\partial \log f_{ij}}{\partial \theta_r} \right) (\theta_0) + \frac{1}{n} \sum_{s=1}^p \sum_{i=1}^g \sum_{j=1}^{r_i} (\hat{\theta}_s - \theta_{0s}) \left(\frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s} \right) (\theta_0) \\ &\quad + \frac{1}{2n} \sum_{s,t=1}^p \sum_{i=1}^g \sum_{j=1}^{r_i} (\hat{\theta}_s - \theta_{0s})(\hat{\theta}_t - \theta_{0t}) \left(\frac{\partial^3 \log f_{ij}}{\partial \theta_r \partial \theta_s \partial \theta_t} \right) (\theta^*). \end{aligned}$$

Seja $\delta_s = \hat{\theta}_s - \theta_{0s}$ e $\delta_t = \hat{\theta}_t - \theta_{0t}$, então a equação anterior fica dada por

$$L_r(\theta) = L_r(\theta_0) - \sum_{s=1}^p L_{rs}(\theta_0) \delta_s + \frac{1}{2} \sum_{s,t=1}^p L_{rst} \delta_s \delta_t, \quad (\text{B.4})$$

onde

$$\begin{aligned} L_r(\theta) &= \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{r_i} \frac{\partial \log f_{ij}}{\partial \theta_r} = \sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial \log f_{ij}}{\partial \theta_r} \\ L_{rs}(\theta) &= -\frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{r_i} \frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s} = -\sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s} \\ L_{rst} &= \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{r_i} \frac{\partial^3 \log f_{ij}}{\partial \theta_r \partial \theta_s \partial \theta_t} (\theta^*) = \sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial^3 \log f_{ij}}{\partial \theta_r \partial \theta_s \partial \theta_t} (\theta^*). \end{aligned}$$

Agora, como $E(\partial \log f_{ij} / \partial \theta_r) = 0$, $r = 1, 2, \dots, p$, pela desigualdade de Chebyshev segue que $\frac{\partial \log f_{ij}}{\partial \theta_r} \xrightarrow{p} 0$ e da linearidade da esperança matemática temos que

$$\sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial \log f_{ij}}{\partial \theta_r} \xrightarrow{p} 0.$$

Assim, concluímos depois da aplicação do teorema de Slutsky, que

$$L_r(\theta_0) \xrightarrow{p} 0.$$

A seguir vemos que

$$\begin{aligned} L_{rs}(\theta_0) &= - \sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s} = - \sum_{i=1}^g \sum_{j=1}^{r_i} \frac{1}{r_i} \mu_i \frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s} \\ &= - \sum_{j=1}^{r_i} \sum_{i=1}^g \frac{1}{r_i} \mu_i \frac{\partial^2 \log f_{ij}}{\partial \theta_r \partial \theta_s}. \end{aligned}$$

Agora se $Z_j = \sum_{i=1}^g \mu_i (-\partial^2 \log f_{ij} / \partial \theta_r \partial \theta_s)$, então $E(Z_j) = J_{rs}(\theta_0)$ é uma matriz positiva definida com determinante finito. Assim, aplicando o teorema de Khintchine obtemos

$$\sum_{j=1}^{r_i} \frac{Z_j}{r_i} \xrightarrow{p} J_{rs}(\theta_0),$$

isto é

$$L_{rs}(\theta_0) \xrightarrow{p} J_{rs}(\theta_0).$$

Também, podemos concluir que

$$|L_{rst}| < \sum_{i=1}^g \mu_i M_i$$

depois de aplicar a esperança na seguinte relação :

$$|L_{rst}| \leq \sum_{i=1}^g \sum_{j=1}^{r_i} \frac{1}{r_i} \left| \frac{\partial^3 \log f_{ij}}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| < \sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} H_{irst}(x_i).$$

O resto da prova da consistência segue diretamente de Ghanda (1954). A prova dos Teoremas B.2 e B.3 também podem ser vistas em Ghanda (1954).

Teorema B.2: seja $\hat{\theta}$ um estimador consistente de θ e uma solução de (B.2).

Sob as suposições I , a matriz

$$\left[\frac{\partial^2 \log \phi}{\partial \theta_r \partial \theta_s}(\hat{\theta}) \right]$$

é positiva definida com probabilidade aproximando-se à um, quando $r_i = \mu_i n \rightarrow \infty$.

Teorema B.3: de todas as possíveis soluções para (B.2), sujeitas às condições I , uma e somente uma tende em probabilidade para o verdadeiro vetor paramétrico θ_0 .

B.3 Normalidade assintótica

Teorema B.4: sob as condições I , quando $\hat{\theta}$ é o vetor de estimadores de máxima verossimilhança, θ_0 é o vetor dos verdadeiros valores paramétricos, $n \rightarrow \infty$ e $r_i = \mu_i n$ para $i = 1, 2, \dots, g$, temos que $\sqrt{n}(\hat{\theta} - \theta_0)$ tem distribuição assintoticamente normal multivariada com média zero e matriz de variância-covariância J^{-1} , onde

$$J = \left[\sum_{i=1}^g \mu_i E \left(\frac{\partial \log f_i}{\partial \theta_r} \frac{\partial \log f_i}{\partial \theta_s} \right) (\theta_0) \right] = - \left[\sum_{i=1}^g \mu_i E \left(\frac{\partial^2 \log f_i}{\partial \theta_r \partial \theta_s} \right) (\theta_0) \right].$$

Prova : se substituirmos $\hat{\theta}$ por θ em (B.4), notamos que $L_r(\hat{\theta}) = 0$, e

$$L_r(\theta_0) = \sum_{s=1}^p (\hat{\theta}_s - \theta_{0s}) L_{rs}(\theta_0) - \frac{1}{2} \sum_{s,t=1}^p (\hat{\theta}_s - \theta_{0s})(\hat{\theta}_t - \theta_{0t}) L_{rst}, \quad (\text{B.5})$$

$r = 1, 2, \dots, p$, ou equivalentemente em notação matricial

$$(L + G)(\hat{\theta}^t - \theta_{0t}) = \left[\sum_{i=1}^g \mu_i \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{\partial \log f_{ij}}{\partial \theta_r}(\theta_0) \right]^t \quad (\text{B.6})$$

onde $L = [L_{rs}(\hat{\theta}_0)]$ e $G = [-\frac{1}{2} \sum_{t=1}^k (\hat{\theta}_t - \theta_{0t}) L_{rst}]$, L e G matrizes quadradas de ordem p . Temos visto que $L \xrightarrow{p} J$ positiva definida, $G \xrightarrow{p} 0$ pois $\hat{\theta} \xrightarrow{p} \theta_0$ e $|L_{rst}|$ é limitada. Então para r_i grande, $L + G$ pode ser invertida e pela aplicação do teorema de Slutsky temos que $L + G \xrightarrow{p} J$ e $(L + G)^{-1} \xrightarrow{p} J^{-1}$. Portanto de (B.7) segue que

$$\sqrt{n}(\hat{\theta}^t - \theta_{0t}) = (L + G)^{-1} \left[\sum_{i=1}^g \sqrt{\mu_i} \sum_{j=1}^{r_i} \frac{1}{\sqrt{r_i}} \frac{\partial \log f_{ij}}{\partial \theta_r}(\theta_0) \right]^t$$

e temos da forma multivariada do teorema do limite central que

$$\left[\sum_{i=1}^g \sqrt{\mu_i} \sum_{j=1}^{r_i} \frac{1}{\sqrt{r_i}} \frac{\partial \log f_{ij}}{\partial \theta_r}(\theta_0) \right]^t$$

tem distribuição normal multivariada com média zero e matriz de variância-covariância J . Assim, concluímos essa prova estabelecendo que quando $r_i = \mu_i n$ ($i = 1, 2, \dots, g$), e $n \rightarrow \infty$ a distribuição assintótica de $\sqrt{n}(\hat{\theta} - \theta_0)$ é normal multivariada com média zero e matriz de variância-covariância J^{-1} .

Apêndice C

Resultados da Simulação

Tabela C.1. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos iguais, dispersões iguais e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.001	0.115	0.094	0.075
	0.045-0.055	0.980	0.608	0.333	0.175
	0.055-0.070	0.019	0.260	0.432	0.525
	0.070-0.100	0.000 (2182)	0.017 (6960)	0.14 (818)	0.225 (40)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.136	0.530	0.615
	0.045-0.055	1.000	0.794	0.423	0.333
	0.055-0.100	0.000 (2276)	0.070 (6846)	0.047 (839)	0.051 (39)

Tabela C.2. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos $a = (0.15, 0.35, 0.15, 0.35)$ dispersões iguais e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.083	0.053	0.000
	0.045-0.055	0.994	0.640	0.287	0.132
	0.055-0.070	0.006	0.266	0.530	0.553
	0.070-0.100	0.000 (2193)	0.011 (6935)	0.131 (834)	0.316 (38)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.005	0.260	0.436
	0.045-0.055	0.943	0.796	0.660	0.487
	0.055-0.100	0.057	0.199	0.079	0.077
		(2291)	(6813)	(857)	(39)

Tabela C.3. Distribuição de frequências para $C_{01}(c_{01}(0.05); \Delta)$, para pesos- $\cap$, dispersões iguais e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.000	0.005	0.000
	0.045-0.055	0.063	0.397	0.265	0.256
	0.055-0.070	0.937	0.560	0.501	0.410
	0.070-0.100	0.000 (2240)	0.043 (6928)	0.230 (793)	0.333 (39)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.000	0.147	0.333
	0.045-0.055	0.944	0.781	0.722	0.583
	0.055-0.100	0.056	0.218	0.131	0.083
		(2266)	(6851)	(835)	(48)

Tabela C.4. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos-U, dispersões iguais e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.913	0.565	0.256	0.195
	0.045-0.055	0.087	0.347	0.376	0.317
	0.055-0.070	0.000	0.087	0.319	0.415
	0.070-0.100	0.000	0.001	0.048	0.073
		(2271)	(6880)	(808)	(41)
Ordem simples em forma de árvore	0.030-0.045	0.175	0.535	0.710	0.722
	0.045-0.055	0.825	0.448	0.253	0.278
	0.055-0.100	0.000	0.017	0.037	0.000
		(2203)	(6943)	(818)	(36)

Tabela C.5. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos iguais, dispersões $v = (1, 3, 1, 3)$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.113	0.099	0.000
	0.045-0.055	0.977	0.604	0.335	0.292
	0.055-0.070	0.023	0.268	0.428	0.500
	0.070-0.100	0.000	0.015	0.139	0.208
		(2289)	(6833)	(830)	(48)
Ordem simples em forma de árvore	0.030-0.045	0.631	0.603	0.717	0.750
	0.045-0.055	0.370	0.396	0.266	0.231
	0.055-0.100	0.000	0.002	0.018	0.019
		(2260)	(6833)	(855)	(52)

Tabela C.6. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos $a = (0.15, 0.35, 0.15, 0.35)$, dispersões $v = (1, 3, 1, 3)$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.034	0.089	0.027
	0.045-0.055	1.000	0.793	0.414	0.270
	0.055-0.070	0.000	0.172	0.462	0.595
	0.070-0.100	0.000 (2237)	0.001 (6862)	0.035 (864)	0.108 (37)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.074	0.455	0.537
	0.045-0.055	1.000	0.919	0.520	0.439
	0.055-0.100	0.000	0.008	0.026	0.024
		(2203)	(6934)	(822)	(41)

Tabela C.7. Distribuição de frequências para $C_{01}(c_{01}(0.05); \Delta)$, para pesos $\cap$, dispersões $v = (1, 3, 1, 3)$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.000	0.005	0.050
	0.045-0.055	0.151	0.480	0.386	0.225
	0.055-0.070	0.849	0.516	0.540	0.550
	0.070-0.100	0.000 (2182)	0.005 (6960)	0.069 (818)	0.175 (40)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.084	0.417	0.564
	0.045-0.055	1.000	0.910	0.565	0.410
	0.055-0.100	0.000	0.006	0.018	0.026
		(2276)	(6846)	(839)	(39)

Tabela C.8. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos- $\cup$, dispersões $v = (1, 3, 1, 3)$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.979	0.654	0.395	0.184
	0.045-0.055	0.021	0.321	0.409	0.447
	0.055-0.070	0.000	0.025	0.197	0.290
	0.070-0.100	0.000 (2193)	0.000 (6935)	0.000 (834)	0.079 (38)
Ordem simples em forma de árvore	0.030-0.045	1.000	0.819	0.829	0.821
	0.045-0.055	0.000	0.181	0.166	0.180
	0.055-0.100	0.000	0.000	0.006	0.000
		(2291)	(6813)	(857)	(39)

Tabela C.9. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos iguais, dispersões- $\cap$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	1.000	0.745	0.337	0.333
	0.045-0.055	0.000	0.239	0.446	0.357
	0.055-0.070	0.000	0.017	0.212	0.286
	0.070-0.100	0.000 (2184)	0.000 (6969)	0.005 (805)	0.024 (42)
Ordem simples em forma de árvore	0.030-0.045	0.631	0.619	0.771	0.771
	0.045-0.055	0.369	0.381	0.216	0.229
	0.055-0.100	0.000	0.001	0.012	0.000
		(2214)	(6933)	(818)	(35)

Tabela C.10. Distribuição de frequências para $C_{01}(c_{01}(0.05); \Delta)$, para pesos $a = (0.15, 0.35, 0.15, 0.35)$, dispersões- $\cap$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	1.000	0.724	0.213	0.077
	0.045-0.055	0.000	0.260	0.510	0.333
	0.055-0.070	0.000	0.016	0.276	0.538
	0.070-0.100	0.000 (2189)	0.000 (6940)	0.001 (832)	0.051 (39)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.165	0.634	0.750
	0.045-0.055	1.000	0.832	0.349	0.250
	0.055-0.100	0.000	0.003	0.017	0.000
		(2192)	(6939)	(837)	(32)

Tabela C.11. Distribuição de frequências para $C_{01}(c_{01}(0.05); \Delta)$, para pesos- $\cap$, dispersões- $\cap$ e 10000 conjuntos gerado ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.002	0.144	0.117	0.079
	0.045-0.055	0.999	0.739	0.429	0.263
	0.055-0.070	0.000	0.117	0.425	0.632
	0.070-0.100	0.000 (2387)	0.000 (6731)	0.028 (844)	0.026 (38)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.085	0.519	0.691
	0.045-0.055	1.000	0.905	0.463	0.310
	0.055-0.100	0.000	0.010	0.018	0.000
		(2211)	(6903)	(844)	(42)

Tabela C.12. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos- $\cup$, dispersões- $\cap$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.020-0.045	1.000	0.950	0.581	0.156
	0.045-0.055	0.000	0.050	0.357	0.594
	0.055-0.070	0.000	0.000	0.062	0.250
	0.070-0.100	0.000	0.000	0.000	0.000
		(2319)	(6829)	(820)	(32)
Ordem simples em forma de árvore	0.030-0.045	1.000	0.865	0.883	0.930
	0.045-0.055	0.000	0.135	0.117	0.070
	0.055-0.100	0.000	0.000	0.000	0.000
		(2292)	(6839)	(826)	(43)

Tabela C.13. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos iguais, dispersões- $\cup$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.000	0.000	0.000
	0.045-0.055	0.000	0.126	0.148	0.167
	0.055-0.070	1.000	0.852	0.672	0.563
	0.070-0.100	0.000	0.022	0.180	0.271
		(2289)	(6833)	(830)	(48)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.000	0.000	0.098
	0.045-0.055	0.627	0.716	0.847	0.732
	0.055-0.100	0.373	0.284	0.153	0.171
		(2203)	(6934)	(822)	(41)

Tabela C.14. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos $a = (0.15, 0.35, 0.15, 0.35)$, dispersões- $\cup$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.000	0.000	0.000
	0.045-0.055	0.000	0.131	0.084	0.075
	0.055-0.070	1.000	0.853	0.735	0.550
	0.070-0.100	0.000 (2182)	0.015 (6960)	0.181 (818)	0.375 (40)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.000	0.000	0.000
	0.045-0.055	0.000	0.056	0.577	0.763
	0.055-0.100	1.000	0.944	0.423	0.237
		(2246)	(6856)	(860)	(38)

Tabela C.15. Distribuição de frequências para $C_{01}(e_{01}(0.05); \Delta)$, para pesos- $\cap$, dispersões- $\cup$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.000	0.000	0.000
	0.045-0.055	0.000	0.002	0.025	0.051
	0.055-0.070	1.000	0.899	0.669	0.564
	0.070-0.100	0.000 (2240)	0.100 (6928)	0.306 (793)	0.385 (39)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.000	0.000	0.000
	0.045-0.055	0.000	0.106	0.538	0.750
	0.055-0.100	1.000	0.894	0.462	0.250
		(2266)	(6851)	(835)	(48)

Tabela C.16. Distribuição de frequências para $C_{01}(\epsilon_{01}(0.05); \Delta)$, para pesos- $\cup$, dispersões- $\cup$ e 10000 conjuntos gerados para ϵ .

Caso	Intervalo	ϵ_{max}			
		0-1	1-3	3-5	5-8
Ordem simples	0.030-0.045	0.000	0.002	0.022	0.000
	0.045-0.055	0.974	0.701	0.368	0.366
	0.055-0.070	0.026	0.295	0.553	0.561
	0.070-0.100	0.000 (2271)	0.002 (6880)	0.057 (808)	0.073 (41)
Ordem simples em forma de árvore	0.030-0.045	0.000	0.000	0.220	0.389
	0.045-0.055	1.000	0.961	0.725	0.583
	0.055-0.100	0.000	0.038	0.055	0.028
		(2203)	(6943)	(818)	(36)

Apêndice D

Programas

```

*****!
!                                     !
!                               PROGRAMA 1                               !
!                                     !
! Este programa refere-se ao ajuste da distribuicao do valor !
! extremo a cada grupo do exemplo dos tipos de turbina !
! ( ilustracao com t=t5 ) !
!                                     !
!*****!
!
$macro ajuste
$cal off=%k*log(t)
$print %k
$offset off
$cycle 10
$fit
$use weibull
$cal %l=%l-1
$endmac
!
$macro weibull
$cal %b=%k : z=%log(t)/10
: a=%fv*z : b=delta*z
: c=a-b : c=%cu(c)
: %k=c(10) : %k=1/%k
: %k = (%b + %k)/2
$endmac
!
$units 10
$data t1 t2 t3 t4 t5
$dinput 4
$cal delta=1
$yva delta
$err p
$cal %k=1.0 : %l=11 : t=t5
$while %l ajuste
$display e
$ext %pe
$cal %a=%pe(1)
!
! calculo e impressao dos parametros de escala, de locacao e do
! decimo percentil da distribuicao do valor extremo e dos res-
! pectivos desvios padroes assintoticos
!
$cal %s = 1/%k : %m = %exp(%a/%k) : %m = - %log(%m)
: %d = 0.1*(1 + (( 1 - 0.5772)**2)*6/((3.1416)**2))
: %d = %sqrt(%s*%d)
: %e = 0.1*6/((3.1416)**2)
: %e = %sqrt(%s*%e)
$print %m %d : %s %e
$cal %c = -2.250367
: %y = %m + %s*%c
: %v = %d**2 + (%e*%c)**2 - 2*%c*(1 - 0.5772)*(%e**2)
: %v = %sqrt(%v)
$print %y %v
$stop

```

```

*****
PROGRAMA 2
*****
Este programa refere-se ao ajuste da distribuicao do valor
extremo ao exemplo dos cinco tipos de turbina assumindo pa
rametros de locacao diferentes e mesmo parametro de escala
*****
$macro ajuste
$cal off=%k*log(t)
$print %k
$offset off
$cycle 10
$fit tipo
$use weibull
$cal %l=%l-1
$endmac
!
$macro weibull
$cal %b=%k : z=%log(t)/50
: a=%fv*z : b=delta*z
: c=a-b : c=%cu(c)
: %k=c(50) : %k=1/%k
: %k = (%b + %k)/2
$endmac
!
$units 50
$data t
$dinput 4
$factor tipo 5
$cal delta=1 : tipo=%gl(5,1)
$yva delta
$serr p
$cal %k=1.0 : %l=11
$while %l ajuste
$display e
$ext %pe
!
! calculo e impressao dos parametros de escala e de locacao da
! valor extremo e dos respectivos desvios padroes assintoticos
!
$macro trans
$cal %m = %exp(%a/%k) : %m = - %log(%m) :
: %d=0.1*(1+((1-0.5772)**2)*6/((3.1416)**2)) :
: %d=%sqrt(%s*%d) :
$print %m %d
$endmac
!
$cal %s=1/%k
$print %s
$cal %a=%pe(1) $use trans
$cal %a=%pe(1) + %pe(2) $use trans
$cal %a=%pe(1) + %pe(3) $use trans
$cal %a=%pe(1) + %pe(4) $use trans
$cal %a=%pe(1) + %pe(5) $use trans
$cal %e=(1/50)*6/((3.1416)**2)
: %e = %sqrt(%s*%e)
$print %s %e
$stop

```

```

*****
!                                     PROGRAMA 3
!
! Este programa refere-se ao ajuste da distribuicao do valor
! extremo aos dados do exemplo dos cinco tipos de turbina
!
! *****
!
! $macro ajuste
! $cal off=%k*log(t)
! $print %k
! $offset off
! $cycle 10
! $fit
! $use weibull
! $cal %l=%l-1
! $endmac
!
! $macro weibull
! $cal %b=%k : z=%log(t)/50
! : a=%fv*z : b=delta*z
! : c=a-b : c=%cu(c)
! : %k=c(50) : %k=1/%k
! : %k = (%b + %k)/2
! $endmac
!
! $units 50
! $data t
! $dinput 4
! $cal delta=1
! $yva delta
! $err p
! $cal %k=1.0 : %l=11
! $while %l ajuste
! $display e
! $ext %pe
! $cal %a=%pe(1)
!
! ! calculo e impressao dos parametros de
! ! escala e de locacao da valor extremo
! ! e dos respectivos desvios padroes
! ! assintoticos
!
! $cal %s = 1/%k : %m = %exp(%a/%k) : %m = - %log(%m)
! : %d = (1/50)*(1+((1-0.5772)**2)*6/((3.1416)**2))
! : %d = %sqrt(%s*%d)
! : %e = (1/50)*6/((3.1416)**2)
! : %e = %sqrt(%s*%e)
! $print %m %d : %s %e
! $stop

```

```

*****
PROGRAMA 4
*****
Este programa calcula as estimativas sob a hipotese nula res-
trita do exemplo que compara os cinco tipos de turbina
*****

$macro ajuste
$scal off=%k*%log(t)
$print %k
$offset off
$cycle 10
$fit tipo
$use weibull
$scal %l=%l-1
$endmac
!
$macro weibull
$scal %b=%k : z=%log(t)/50
: a=%fv*z : b=delta*z
: c=a-b : c=%cu(c)
: %k=c(50) : %k=1/%k
: %k = (%b + %k)/2
$endmac
!
$units 50
$data t
$dinput 4
$factor tipo 5
$scal delta=1 : tipo=%gl(5,1)
$factor tipo 3
$scal tipo = %if(%eq(tipo,4),1,tipo)
: tipo = %if(%eq(tipo,3),2,tipo)
: tipo = %if(%eq(tipo,5),3,tipo)
$yva delta
$serr p
$scal %k=1.0 : %l=11
$while %l ajuste
$display e
$ext %pe
!
! calculo e impressao dos parametros de escala e de locacao da
! valor extremo e dos respectivos desvios padroes assintoticos
!
$macro trans
$scal %m = %exp(%a/%k) : %m = - %log(%m)
: %d = (1/r)*(1+((1-0.5772)**2)*6/((3.1416)**2))
: %d = %sqrt(%s*%d)
$print %m %d
$endmac
!
$scal %s=1/%k
$print %s
$scal %a=%pe(1) $scal r=20 $use trans
$scal %a=%pe(1) + %pe(2) $scal r=20 $use trans
$scal %a=%pe(1) + %pe(3) $scal r=10 $use trans
$scal %e=(1/50)*6/((3.1416)**2)
: %e = %sqrt(%s*%e)
$print %s %e
$stop

```


PROGRAMA 5

```

*****
PROGRAMA 6
*****
Este programa calcula as estimativas do modelo de regressao
do valor extremo que compara quatro tipos de preparacao com
uma preparacao padrao (sem sexo)
*****
!
!
$macro ajuste
$cal off=%k*%log(t-0.5)
$print %k
$offset off
$cycle 10
$fit prepa + dose - %gm
$use weibull
$cal %l=%l-1
$endmac
!
$macro weibull
$cal %b=%k : z=%log(t-0.5)/100
: a=%fv*z : b=delta*z
: c=a-b : c=%cu(c)
: %k=c(100) : %k=1/%k
: %k = (%b + %k)/2
$endmac
!
$units 100
$data t dose
$dinput 4
$factor prepa 5
$cal delta=1 : prepa=%gl(5,5)
$yva delta
$err p
$cal %k=1.0 : %l=11
$while %l ajuste
$display e
$print %k
!
$stop

```

```
*****  
PROGRAMA 7  
*****  
Este programa calcula as estimativas sob a hipotese nula  
restrita do modelo de regressao do valor extremo correspon-  
dente ao exemplo em que sao comparados quatro preparacoes  
com uma preparacao padrao  
*****  
$macro ajuste  
$cal off=%k*log(t-0.5)  
$print %k  
$offset off  
$cycle 10  
$fit prepa + dose - %gm  
$use weibull  
$cal %l=%l-1  
$endmac  
!  
$macro weibull  
$cal %b=%k : z=log(t-0.5)/100  
: a=%fv*z : b=delta*z  
: c=a-b : c=%cu(c)  
: %k=c(100) : %k=1/%k  
: %k = (%b + %k)/2  
$endmac  
!  
$units 100  
$data t dose  
$dinput 4  
$factor prepa 5  
$cal delta=1 : prepa=%gl(5,5)  
: prepa = %if(%eq(prepa,5),1,prepa)  
: prepa = %if(%eq(prepa,3),1,prepa)  
: prepa = %if(%eq(prepa,4),3,prepa)  
$factor prepa 3  
$yva delta  
$err p  
$cal %k=1.0 : %l=11  
$while %l ajuste  
$display e  
$print %k  
!  
$stop
```

```
*****  
PROGRAMA 8  
*****  
Este programa calcula o valor da estatistica da razao de  
verossimilhanca para testar a homogeneidade dos parametros  
de escala da distribuicao do valor extremo no exemplo em  
que os cinco tipos de turbina sao comparados  
*****  
$units 10  
$data t1 t2 t3 t4 t5  
$dinput 4  
$scal t1=%log(t1) : t2=%log(t2) : t3=%log(t3)  
: t4=%log(t4) : t5=%log(t5)  
!  
$var 5 mu0 mul sig0 sigl  
$data mu0 $read 2.483 2.115 2.115 2.483 2.753  
$data mul $read 2.490 1.926 2.271 2.410 2.794  
$data sig0 $read 0.411 0.411 0.411 0.411 0.411  
$data sigl $read 0.386 0.431 0.319 0.516 0.274  
!  
$scal z1 = (t1 - mul(1))/sigl(1) : z2 = (t2 - mul(2))/sigl(2)  
: z3 = (t3 - mul(3))/sigl(3) : z4 = (t4 - mul(4))/sigl(4)  
: z5 = (t5 - mul(5))/sigl(5)  
: z10 = (t1 - mu0(1))/sig0(1) : z20 = (t2 - mu0(2))/sig0(2)  
: z30 = (t3 - mu0(3))/sig0(3) : z40 = (t4 - mu0(4))/sig0(4)  
: z50 = (t5 - mu0(5))/sig0(5)  
!  
$scal w = (z1 - z10) + (z2 - z20) + (z3 - z30) + (z4 - z40) +  
(z5 - z50) - (%exp(z1) - %exp(z10)) - (%exp(z2) - %exp(z20)) -  
(%exp(z3) - %exp(z30)) - (%exp(z4) - %exp(z40)) -  
(%exp(z5) - %exp(z50))  
: %w = %cu(w)  
: v = - %log(sigl/sig0) : %v = 10*%cu(v)  
: %w = 2*(%w + %v)  
!  
$print %w  
$stop
```

```

*****
PROGRAMA 9
*****
Este programa calcula o valor da estatística da razão de
verossimilhança do teste restrito correspondente ao exemplo
em que cinco tipos de turbina são comparados
*****
$units 10
$data t1 t2 t3 t4 t5
$dinput 4
$cal t1=%log(t1) : t2=%log(t2) : t3=%log(t3)
: t4=%log(t4) : t5=%log(t5)
!
$var 5 mu0 mul sig0 sig1
$data mu0 $read 2.483 2.115 2.115 2.483 2.753
$data mul $read 2.484 1.941 2.242 2.491 2.754
$data sig0 $read 0.411 0.411 0.411 0.411 0.411
$data sig1 $read 0.403 0.403 0.403 0.403 0.403
!
$cal z1 = (t1 - mul(1))/sig1(1) : z2 = (t2 - mul(2))/sig1(2)
: z3 = (t3 - mul(3))/sig1(3) : z4 = (t4 - mul(4))/sig1(4)
: z5 = (t5 - mul(5))/sig1(5)
: z10 = (t1 - mu0(1))/sig0(1) : z20 = (t2 - mu0(2))/sig0(2)
: z30 = (t3 - mu0(3))/sig0(3) : z40 = (t4 - mu0(4))/sig0(4)
: z50 = (t5 - mu0(5))/sig0(5)
!
$cal w = (z1 - z10) + (z2 - z20) + (z3 - z30) + (z4 - z40) +
(z5 - z50) - (%exp(z1) - %exp(z10)) - (%exp(z2) - %exp(z20))
- (%exp(z3) - %exp(z30)) - (%exp(z4) - %exp(z40)) -
(%exp(z5) - %exp(z50))
: %w = %cu(w)
: v = - %log(sig1/sig0) : %v = 10*%cu(v)
: %w = 2*(%w + %v)
!
$print %w
$stop

```

```

!*****!
!                                     PROGRAM 10                               !
!                                     !                                           !
! Este programa calcula o valor da estatística da razão de ve !
! rossimilhança correspondente ao teste restrito do exemplo !
! em que são comparados quatro preparações com uma prepara- !
! ção padrão !
!*****!
!
$units 100
$data t dose
$dinput 4
$factor prepa 5
$cal prepa=%gl(5,5)
: x1=%if(%eq(prepa,1),1,0)
: x2=%if(%eq(prepa,2),1,0)
: x3=%if(%eq(prepa,3),1,0)
: x4=%if(%eq(prepa,4),1,0)
: x5=%if(%eq(prepa,5),1,0)
: lp1 = 0.185*x1 + 0.696*x2 + 0.167*x3 + 0.711*x4 + 0.094*x5 +
:       0.388*dose
!
$cal prepa = %if(%eq(prepa,5),1,prepa)
: x1=%if(%eq(prepa,1),1,0)
: lp0 = 0.138*x1 + 0.689*x2 + 0.162*x3 + 0.704*x4 + 0.390*dose
!
$cal lt = %log(t) :
: e = (lt - lp1)/0.302 - (lt - lp0)/0.303 - %exp((lt - lp1)/0.302)
:   + %exp((lt - lp0)/0.303) - %log(0.302/0.303) : %w = 2*%cu(e)
$print %w
$stop

```

Bibliografia

- [1] Abrahamson, I.G. (1964). Orthant probabilities for the quadrivariate normal distribution. *Annals of Mathematical Statistics* 35, 1685-1703.
- [2] Aitkin, R. A.; Anderson, D.; Brian, F. e Hinde, J. (1989). *Statistical Modelling in Glim*. Oxford: Oxford Science Publications.
- [3] Arellano-valle e Bolfarine, H. (1995). A note on the simple structural regression model. *Annals of the Institute of Statistical Mathematics* (to appear).
- [4] Bacon, R.H. (1963). Approximations to multivariate normal orthant probabilities. *Annals of Mathematical Statistics* 34, 191-198.
- [5] Barlow, R. E.; Bartholomew, D. J.; Brummer, J. M. e Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: John Wiley.
- [6] Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika* 46, 36-48.
- [7] Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives. *Biometrika* 46, 328 - 335.
- [8] Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives (with discussion). *Journal of the Royal Statistical Society B* 23, 239-281.
- [9] Barton, D. E. e Mallows, C. L. (1961). The randomization bases of the problem of the amalgamation of weighted means. *Journal of the Royal Statistical Society B* 23, 423-433.
- [10] Bolfarine, H. e Cordani, L.K. (1993). Estimation of a structural linear regression model with a known reliability ratio. *Annals of the Institute of Statistical Mathematics*, 3, 531-540.

- [11] Bohrer, R. e Chow, W. (1978). Algorithm AS122. Weights for one sided multivariate inference. *Applied Statistics* 27, 100-104.
- [12] Bradley, R. A. e Gart, J. J. (1962). The asymptotic properties of the ML estimators when sampling from associated populations. *Biometrika* 49, 205-214.
- [13] Cadwell, J. H. (1951) The bivariate normal integral. *Biometrika* 38, 475-481.
- [14] Chanda, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika* 41, 56-61.
- [15] Chase, G. R. (1974). On testing for ordered alternatives with increased sample size for a control. *Biometrika* 61, 569-578.
- [16] Childs, D. R. (1967). Reduction of the multivariate normal integral to characteristic form. *Biometrika* 54, 293-300.
- [17] David, F. N. (1953). A note on the evaluation of the multivariate normal integral. *Biometrika* 40, 458-459.
- [18] Dun, O. J.; Kronmal, R. A. e Yee, W. J. (1968). Tables of multivariate t distribution. *Technical Report*, School of Public Health, UCLA, Los Angeles, CA.
- [19] Dunnett, C. W. e Sobel, M. (1955). Approximation to the probability integral and certain percentage points of a multivariate analogue of student's distribution. *Biometrika* 42, 258-260.
- [20] Fahrmeir, L. e Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13, 342-368.
- [21] Fahrmeir, L. e Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers* 35, 211-229.
- [22] Farewell, V. T. e Prentice, R. L. (1977). A study of distributional shape in life testing. *Technometrics* 19, 69-75.
- [23] Finney, D. J. (1971). *Probit Analysis*, Third Edition. Cambridge: Cambridge University Press.
- [24] Finney, D. J. (1978). *Statistical Methods in Biological Assay*, Third Edition. London: Griffin.

- [25] Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley.
- [26] Genz, A. e Kahaner, D. K. (1986). The numerical evaluation of certain multivariate normal integrals. *Journal of Computational and Applied Mathematics* 16, 255-258.
- [27] Gourieroux, C.; Holly, A. e Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50, 63 - 80.
- [28] Gourieroux, C. e Monfort, A. (1995). *Statistics and Econometric*. Volumes 1 e 2. Cambridge: Cambridge University Press.
- [29] Grove, D. M. (1980). A test of independence against a class of ordered alternatives in a $2 \times C$ contingency table. *Journal American Statistical Association* 75, 454-459.
- [30] Hager, H. W. e Bain, L. J. (1970). Inferential procedures for the generalized gamma distribution. *Journal American Statistical Association* 65, 1601-1609.
- [31] Harris, B. e Soms, A. P. (1980). The use of the tetrachoric series for evaluating multivariate normal probabilities. *Journal Multivariate Analysis* 10, 252-267.
- [32] Harter, H. L. (1967). Maximum likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples. *Technometrics* 13, 547-557.
- [33] Henery, R. J. (1981). An approximation to certain multivariate normal probabilities. *Journal of the Royal Statistical Society B* 43, 81-85.
- [34] Hillier, G. H. (1986). Joint tests for zero restrictions on nonnegative regression coefficients. *Biometrika* 73, 657-669.
- [35] Ihm, P. (1959). Numerical evaluation of certain multivariate normal integrals. *Sankhya* 21, 363-366.
- [36] John, S. (1959). On the evaluation of the probability integral of a multivariate normal distribution. *Sankhya* 21, 367-370.
- [37] John, S. (1966). On the evaluation of probabilities of convex polyhedra under multivariate normal and t-distributions. *Journal of the Royal Statistical Society B* 28, 366-369.
- [38] Kendall, M. G. (1941). Proof of relations connected with tetrachoric series and its generalization. *Biometrika* 32, 196-198.

- [39] Kodde, D. A. e Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54, 1243-1248.
- [40] Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* 50, 403 - 419.
- [41] Lawless, J. F. (1980). Inference in the generalized gamma and log-gamma distributions. *Technometrics* 22, 409-419.
- [42] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley.
- [43] Lee, C. C.; Robertson, T. e Wright, F. T. (1993). Bounds on distributions arising in order restricted inferences with restricted weights. *Biometrika* 80, 405-416.
- [44] Lucas, L. A.; Robertson, T. e Wright, F. T. (1989). Approximation of the level probabilities for a unimodal ordering. *Communications in Statistics Simulation and Computation* 18, 1401-1420.
- [45] Marsaglia, G. (1963). Expressing the normal distribution with covariance matrix $A + B$ in terms of one with covariance matrix A . *Biometrika* 50, 535-538.
- [46] McCool, J. I. (1979). Analysis of single classification experiments based on censored samples from the two-parameter Weibull distribution. *Journal of Statistical Planning and Inference* 3, 39-68.
- [47] McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models*, Second Edition, London: Chapman and Hall.
- [48] McDonald, J. M. e Diamond, I. (1983). Fiting generalized linear models with linear inequality constraints. *Glim Newsletter* 6, 29-36.
- [49] McFadden, J. A. (1955). Urn models of correlation and comparison with the multivariate normal integral. *Annals Mathematical Statistics* 26, 478-489.
- [50] McFadden, J. A. (1956). An aproximation for the symmetric, quadrivariate normal integral. *Biometrika* 43, 206-207.
- [51] McFadden, J. A. (1960). Two expansions for the quadrivariate normal integral. *Biometrika* 47, 325-333.
- [52] Miles, R. E. (1959). The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika* 46, 317-327.

- [53] Milton, R. C. (1972). Computer evaluation of the multivariate normal integral. *Technometrics* 14, 881-889.
- [54] Moran, P. A. P (1948). Rank correlation and product moment correlation. *Biometrika* 35, 203-206.
- [55] Moran, P. A. P (1984). The Monte Carlo evaluation of orthant probabilities for multivariate normal distributions. *Australian Journal Statistics* 26, 39-44.
- [56] Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*. London:Chapman and Hall.
- [57] Nuesch, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Annals Mathematical Statistics* 37, 113 -119.
- [58] Paula, G. A. e Rojas, O. V. (1995). On restricted hypotheses in extreme value regression models. *Technical Report* MAE-9511 IME-USP São Paulo. (revisado para o *Computational Statistics and Data Analysis*).
- [59] Paula, G. A. e Sen, P. K. (1994). Tests of ordered hypotheses in linkage in heredity. *Statistics and Probability Letters* 20, 395-400.
- [60] Paula, G. A. e Sen, P. K. (1995). One-sided tests in generalized linear models with parallel regression lines. *Biometrics* 51, 1494-1501.
- [61] Paula, G. A. (1996). On approximation of the level probabilities for testing ordered parallel regression lines. *Statistics and Probability Letters* (to appear).
- [62] Parr, V. B. e Webster, J. T. (1965). A method for discriminating between failure density functions used in reliability predictions. *Technometrics* 7, 1-10.
- [63] Pearson, K. (1901). Mathematical contributions to the theory of evolution VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A* 195, 1-47.
- [64] Perlman, M. D. (1969). One sided problems in multivariate analysis. *Annals Mathematical Statistics* 40, 549 - 567.
- [65] Plackett, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* 41, 351-360.
- [66] Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association* 76, 103-106.

- [67] Prentice, R. L. (1974). A log-gamma model and its maximum likelihood estimation. *Biometrika* 61, 539-544.
- [68] Robertson, T. e Wright, F. T. (1982a). Bounds on mixtures of distribution arising in order restricted inference. *American Statistician* 10, 302-306.
- [69] Robertson, T. e Wright, F. T. (1982b). Testing for ordered alternatives with increased precision in one of the samples. *Biometrika* 69, 579-586.
- [70] Robertson, T. e Wright, F. T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika* 70, 597-606.
- [71] Robertson, T. e Wright, F. T. (1985). One sided comparisons for treatments with control. *Canadian Journal Statistics* 13, 109-122.
- [72] Robertson, T.; Wright, F. T. e Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. New York: John Wiley.
- [73] Rojas, O. V.; Bolfarine, H. e Paula, G. A. (1995). On restricted hypotheses in structural regression models. *Technical Report* MAE-9525 IME-USP São Paulo. (submetido ao *Journal of Statistical Planning and Inference*).
- [74] Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika* 41, 200-227.
- [75] Sasabuchi, S. e Kulatunga, D. D. S. (1985). Some approximations for the null distributions of the $\bar{E}^2$ statistic used in order restricted inference. *Biometrika* 72, 476-480.
- [76] Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* 72, 133 - 144.
- [77] Shapiro, A. (1987). A conjecture related to chi-bar-squared distributions, *American Mathematical Monthly* 94, 46-47.
- [78] Silvapulle, M. J. (1991). On limited dependent variable models: maximum likelihood estimation and test of one sided hypothesis. *Econometric Theory* 7, 385- 395.
- [79] Silvapulle, M. J. (1994). On tests against one sided hypothesis in some generalized linear models. *Biometrics* 50, 853-858.

- [80] Singh, B. e Wright, F. T. (1986). Power series approximations to the null distribution of some chi-bar squared statistics. In Robertsom et al. (1988) *Advances in Order Restricted Statistical Inference*, Wiley.
- [81] Singh, B. e Wright, F. T. (1993). The level probabilities for a simple loop ordering. *Annals of the Institute of Statistical Mathematics* 45, 279-292.
- [82] Siskind, V. (1976). Approximate probability integrals and critical values for Bartholomew's test of ordered means. *Biometrika* 63, 647-654.
- [83] Sondhi, M. M. (1961). A note on the quadrivariate normal integral. *Biometrika* 48, 201-203.
- [84] Stacy, E. W. (1962). A generalization of the gamma distributions. *Annals Mathematical Statistics* 33, 1187-1192.
- [85] Stacy, E. W. e Mihram, G. A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics* 7, 349-358.
- [86] Sun, H. J. (1987). A general reduction method for n-variate normal orthant probability. *Communications in Statistics, Theory and Methods* 17, 3913-3921.
- [87] Sun, H. J. (1988). A fortran subroutine for computing normal orthant probabilities of dimensions up to nine. *Communications in Statistics Simulation and Computation* 17, 3, 1097-1111.
- [88] Tong, Y. L. (1990). *The Multivariate Normal Distribution*. New York: Springer.
- [89] Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* 82, 782-793.
- [90] Wolak, F. A. (1989). Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models. *Econometric Theory* 5, 1-35.
- [91] Wolak, F. A. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica* 59, 981-995.
- [92] Wong, M. Y. (1989). Likelihood estimation of a simple regression model when both variables have error. *Biometrika* 76, 141-148.
- [93] Wrigth, F. T. e Tran, T. (1985). Approximating the level probabilities in order restricted inference: The simple tree ordering. *Biometrika* 72, 429-439.