

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM BIOMETRIA

Tópicos Recentes em Estatística Paramétrica Aplicada à Biometria

Dissertação submetida à
Universidade Federal Rural de Pernambuco
como parte dos requisitos para a
obtenção do grau de Mestre em Biometria.

Mestrando: Lucas Gallindo Martins Soares
Orientador: Prof. Gauss Moutinho Cordeiro, PhD.

Recife, Agosto de 2007

Agradecimentos

À todos do DEINFO, pela paciência;

Ao professor Gauss M. Cordeiro pela orientação;

Ao professor Borko D. Stošić, pelas conversas na hora do cafezinho;

Ao professor Wilson Rosa pelas discussões sobre o que é matemática;

À professora Tatijana Stošić por insistir que os alunos sejam bons pesquisadores;

Aos colegas de turma Angela, Rosângela, Esdras, Adriano, Iran, Frank, Luiz José, Luiz Henrique, Janilson, Moacy, Domingos e Eduardo;

À meu pai e minha mãe, pois sem eles eu não existiria;

À Sir Ronald A. Fisher por ser indeciso quanto a ser matemático ou geneticista, e terminar criando o Planejamento de Experimentos e a Teoria de Verossimilhança.

Resumo

A palavra Biometria vem do grego, onde *Bios* significa **Vida** e *Metron* significa **Medida**, e pode ser traduzida como medida da vida. No contexto moderno, Biometria é um conjunto de métodos advindos das ciências exatas para o estudo de dados de origem biológica, sendo que a grande maioria destes métodos provém da Estatística que, de acordo com Rao (A technology for the millennium, *International Journal of Mathematics and Statistical Science*, Volume 8, Número 1, Junho de 1999, páginas 5-25),

”...é uma ciência que estuda e pesquisa sobre: o levantamento de dados com a máxima quantidade de informação possível para um dado custo; o processamento de dados para a quantificação da quantidade de incerteza existente na resposta para um determinado problema; a tomada de decisões sob condições de incerteza, sob o menor risco possível.”

No século XX, em especial após os trabalhos de Sir Ronald A. Fisher e contemporâneos sobre métodos ótimos de estimação e testes de hipóteses, uma classe de técnicas estatísticas tem dominado sobre as outras: os modelos paramétricos. Um modelo paramétrico é um relacionamento entre um fenômeno observado e uma família de funções de densidade de probabilidade, que permite fazer inferência sobre o fenômeno real através da inferência científica sobre os parâmetros da família de funções de densidade em estudo. O objetivo do presente trabalho é detalhar as propriedades dos mais recentes avanços em modelos paramétricos aplicados à Biometria, com ênfase na distribuição Conway-Maxwell-Poisson (COM-Poisson), nas distribuições estendidas por alternativas de Lehmann e similares, e de descobertas recentes sobre propriedades de estatísticas para testes de hipóteses baseadas na razão de verossimilhanças.

Abstract

The word Biometry comes from the Greek, where *Bios* means **Life** and *Metron* means **Measure**, and can be translated as measure of life. In the modern context, Biometry is a set of methods that came from exact sciences and are used to study data from biological studies, most of these methods coming from the Statistical Science which, according to Rao (A technology for the millennium, *International Journal of Mathematics and Statistical Science*, Volume 8, Number 1, June de 1999, pages 5-25),

”it is a science that studies and researches about: the gathering of data with the maximum possible amount of information for a given cost; the processing of data for the quantification of the quantity of uncertainty existing in the answer for a given problem; the making of decisions under uncertainty conditions, with the smallest possible risk.”

In the XX century, in special after the works of Sir Ronald A. Fisher and contemporaries about optimum estimation methods and hypothesis testing, a class of statistical techniques gained dominance over others: the parametric models. A parametric model is relationship between a family of probability density functions and a observed phenomenon, which allows inference about the real phenomenon through scientific inference about the parameters of the family of probability density function under study. The purpose of the present work is to work out details of the recent discoveries in parametric models applied to biometry, with emphasis on the Conway-Maxwell-Poisson distribution, on probability distributions extended by Lehmann alternatives and similar methods, and recent discoveries about properties of statistics for hypothesis testing based on likelihood ratios.

Sumário

Agradecimentos	p. ii
Lista de Figuras	p. vii
Lista de Tabelas	p. viii
1 Introdução	p. 1
1.1 Preliminares	p. 4
1.2 Testes de Hipóteses	p. 4
2 Distribuição de Conway-Maxwell-Poisson	p. 6
2.1 Propriedades Conhecidas e Função de Verossimilhança	p. 7
2.2 Viés de Primeira Ordem das Estimativas de Máxima Verossimilhança	p. 9
2.3 Estimação dos Parâmetros por Mínimos Quadrados	p. 10
3 Distribuições Estendidas por Alternativas de Lehmann	p. 14
3.1 Introdução	p. 14
3.2 Momentos - Cálculo Usando a Função Quantil	p. 15
3.3 Caracterização da Segunda Forma em Termos de Momentos Ponderados por Probabilidade	p. 16

3.4	Perda de Poder em Erros de Especificação de Distribuições Extendidas pela Primeira Alternativa	p. 18
3.4.1	Divergência Kullback-Leibler	p. 18
3.4.2	Especificação Errônea e Perda de Poder em Testes de Razão de Verossimilhanças	p. 19
4	Distribuição da Razão de Verossimilhanças	p. 22
4.1	Expansões Assintóticas	p. 23
4.1.1	Expansões Assintóticas Importantes: $(\hat{\theta} - \theta)_r$	p. 23
4.1.2	Expansões Assintóticas Importantes: $(\ell(\hat{\theta}) - \ell(\theta))$	p. 25
4.1.3	Expansões Assintóticas Importantes: r_k	p. 25
4.2	Família Exponencial	p. 26
4.3	Estatísticas s_k e r_k na Família Exponencial	p. 27
4.3.1	Família Exponencial Uniparamétrica	p. 27
4.4	Viés de Primeira Ordem dos Estimadores de Máxima Verossimilhança	p. 28
5	Conclusões	p. 29
	Referências	p. 30
	Apêndice A – Detalhes da Distribuição COM-Poisson	p. 33
A.1	Derivadas da Função Z	p. 33
A.2	Equações do Viés de Primeira Ordem da CMP	p. 34
A.3	Código em Linguagem R para gerar números aleatórios com distribuição CMP	p. 36

Lista de Figuras

1	Perda de Poder em Função de λ , para $\lambda > 1$	p.21
---	--	------

Lista de Tabelas

- 1 Resultados obtidos com amostras da distribuição de Poisson ($\lambda = 1$), Média da Amostra (estimador de máxima verossimilhança para λ utilizando a distribuição de Poisson, Variância da Amostra, e estimativas de λ e ν p. 12
- 2 Resultados obtidos com amostras da distribuição de Poisson ($\lambda = 10$), Média da Amostra (estimador de máxima verossimilhança para λ utilizando a distribuição de Poisson, Variância da Amostra, e estimativas de λ e ν p. 12
- 3 Resultados obtidos com amostras da distribuição Geométrica ($P(\text{sucesso}) = \lambda = 1/4$), Média da Amostra (o estimador de máxima verossimilhança para a distribuição geométrica é $1/\bar{X}$), e estimativas de λ e ν p. 13

1 Introdução

No formalismo de teoria da medida, define-se uma família de *funções de densidade* da seguinte forma: Seja D um conjunto contido em R^k , e seja $dm(x)$ a medida de contagem, se D é discreto, ou a medida de Lebesgue-Stieltjes, caso contrário.

Definição.: Uma função $f(x) : D \rightarrow R$ é uma *função de densidade de probabilidade* (ou fdp), se e somente se,

$$\int_D f dm(x) = 1 \quad \text{e} \quad f(x) \geq 0 \quad \forall x \in D$$

Diz-se que a variável X tem fdp f se a probabilidade de X estar no conjunto S é dada pela expressão $\int_S f(x) dm$.

Removendo os aspectos de teoria da medida e estudando o caso particular do R^1 , uma função de densidade de probabilidade é uma função f de R em R tal que a integral de f no intervalo menos infinito até infinito é igual a um, e em que $f(x)$ é maior ou igual a zero para todo valor de x . Uma fdp f está associada à variável X se a probabilidade de uma observação da variável X estar no intervalo (a, b) é igual à integral de f no intervalo (a, b) . Uma família de densidades é uma função f de $R \times \Theta$ (Θ , chamado de espaço paramétrico, é um conjunto aberto arbitrário contido em algum espaço vetorial euclidiano R^n) tal que para cada $\theta \in \Theta$, $f(x, \theta)$ é uma densidade. Usualmente, quando isto não implica em ambiguidades, refere-se à *famílias de funções de densidade de probabilidade* como *funções de densidade* ou simplesmente *densidades*.

O conjunto de pontos em que a fdp é maior que zero é chamado suporte de f . Se o suporte de uma densidade f for o mesmo para todo valor de $\theta \in \Theta$, esta família é

dita regular. O exemplo mais comum de fdp não-regular é a *densidade Uniforme* $(0, \theta)$, freqüentemente abreviado $U(0, \theta)$ cuja densidade é dada pela fórmula

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & x \in (0, \theta) \\ 0, & x \notin (0, \theta) \end{cases} \quad \forall \theta > 0 \quad (1.1)$$

O suporte depende de θ e, portanto, f não é regular.

O exemplo mais clássico de família de densidades é a função conhecida como curva normal, Gaussiana ou curva do sino, dada pela expressão

$$f(x, \sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x - \mu)^2}{2} \quad (1.2)$$

em que σ^2 é um valor qualquer maior que zero e μ pode assumir qualquer valor na reta real. Tipicamente, escreve-se $f(x|\sigma, \mu)$ no lugar de $f(x, \sigma, \mu)$, para diferenciar quem é a variável real e quem são as variáveis do espaço paramétrico, chamadas de parâmetros. A densidade normal (freqüentemente chamada de distribuição normal) tem uma propriedade interessante, dada por um resultado conhecido como Teorema Central do Limite, de que dada uma seqüência de variáveis $X_1, X_2, \dots, X_n, \dots$ cada uma com uma densidade bem-comportada (as funções de densidade não são necessariamente iguais), a seqüência de densidades associada à seqüência de variáveis definida como $S_n = \sum_1^n X_j$ converge para a distribuição normal quando $n \rightarrow \infty$.

Outra definição extremamente importante para o estudo de modelos paramétricos (ver, por exemplo, o Capítulo (3)) é a definição de função distribuição de probabilidade. Em termos formais, uma função F de D em R tal que

$$\lim_{x \rightarrow -\infty} F(x|\theta) = 0 \quad (1.3)$$

$$\lim_{x \rightarrow \infty} F(x|\theta) = 1 \quad (1.4)$$

$$F(x|\theta) \geq 0 \quad \forall x \in D \quad \forall \theta \in \Theta \quad (1.5)$$

é chamada de *Família de Funções de Distribuição de Probabilidade Acumulada* (fda). Em outras palavras, uma função de probabilidade acumulada é uma função que se aproxima

de 1 quando o argumento cresce indefinidamente (quando x tende à infinito), de zero quando o argumento decresce (quando x tende à menos infinito) e que nunca é negativa, sendo consequência disso o fato de toda fda ser não-decrescente. Esta função quantifica a probabilidade de uma variável aleatória X ter valor menor ou igual à x e, quando não causar ambiguidade, fdas serão chamadas de *distribuições acumuladas* ou *distribuições*.

Há um relacionamento entre fdps e fdas, pois para toda função de densidade de probabilidade f , a função F definida por

$$F(x|\theta) = \int_{-\infty}^x f dm(x) \quad (1.6)$$

é uma fda e, de maneira simétrica, para toda fda F diferenciável, a função f definida como

$$f(x, \theta) = \frac{dF(x, \theta)}{dx} \quad (1.7)$$

é uma função de densidade. Em outras palavras, pelo teorema fundamental do cálculo e pela primeira propriedade das distribuições, se F é uma distribuição, então a derivada de f é uma densidade. Cada densidade define unicamente uma distribuição e cada distribuição define unicamente uma densidade, e por isso na literatura os termos *distribuição* e *densidade* são muitas vezes trocados sem causar confusão.

Um modelo paramétrico é uma associação entre um fenômeno real, que gera uma variável Y , e uma família de densidades f , tal que para algum valor de $\theta \in \Theta$, X está associada à f . O objetivo principal da inferência paramétrica é fornecer evidências estatísticas que sobre este valor θ à partir de observações da variável Y proveniente de uma amostra ou de um experimento, sendo que, em geral, θ é uma função de outras variáveis, chamadas de $X_1, X_2, \dots, n, \dots$, sendo que tem-se controle sobre a seqüência de variáveis X .

Na prática geral da inferência paramétrica, sabem-se que valores que a variável Y em estudo pode tomar, mas tem-se pouca ou nenhuma evidência sobre que família de densidades está associada à essa variável. O procedimento mais comum nestes casos é

escolher dentre as famílias que tem como suporte o conjunto de valores aceitáveis para a variável Y a que tem forma mais simples ou menor dimensão no conjunto θ .

1.1 Preliminares

Seja Y um vetor de variáveis aleatórias com distribuição $f_Y(y, \theta)$, θ desconhecido, e seja $y_0^T = [y_1 \ y_2 \ \cdots \ y_n]$ uma observação de Y . Define-se a função de verossimilhança associada à observação y_0 como $L(\theta) = f_Y(y_0, \theta)$. Em situações mais comuns, tem-se que os componentes de Y são variáveis aleatórias independentes e identicamente distribuídas com função de densidade $f_y(y_i, \theta)$, e o y_0 é uma amostra das variáveis y_i . Tem-se então que função de verossimilhança associada com a amostra y_1, y_2, \dots, y_n é

$$L(\theta) = \prod_{i=1}^n f_y(y_i, \theta) \quad (1.8)$$

Sob esta ótica, a população de y tem uma distribuição da forma $f_y(y, \theta)$ com $\theta = \theta_0$, e o nosso problema é descobrir o valor de θ_0 . Uma forma de estimar seu valor proposta por [7] é escolher o valor de θ que maximiza o valor de $L(\theta)$. Como freqüentemente $L(\theta)$ envolve o cálculo de exponenciais complicadas (por exemplo, a equação (4.15), em geral trabalha-se com a função de log-verossimilhança, definida por $\ell(\theta) = \ln(L(\theta)) = \sum_{i=1}^n \ln(f_y(y_i, \theta))$, pois para qualquer função f estritamente monótona, maximizar $f(L(\theta))$ é equivalente à maximizar $L(\theta)$.

1.2 Testes de Hipóteses

Para testar uma hipótese sobre a densidade populacional, quando as hipótese tem a forma Em muitos problemas aplicados deseja-se testar uma hipótese sobre a densidade populacional que tem a forma

$$H_0 : \theta_{i_1} = b_1, \theta_{i_2} = b_2, \dots, \theta_{i_k} = b_k, \quad i_j \in \{1, 2, \dots, p\} \quad \forall j$$

em que os valores i_k são índices que especificam determinam o subconjunto de θ com o qual a hipótese lide. Em outros termos, temos um vetor de parâmetros θ e queremos testar uma hipótese sob um subconjunto do vetor de parâmetros, por exemplo, testar $H_0 : \theta_2 = 0, \theta_5 = 1$ em um problema em que θ tem dimensão maior ou igual à 5.

Particionando o vetor θ em duas partes $\theta = (\phi, \chi)$ de tal forma que todos os valores de θ a serem testados fiquem em χ , pode-se calcular o valor da verossimilhança no ponto de máximo, chamado de $\ell(\hat{\phi}, \hat{\chi})$ e o valor máximo da verossimilhança supondo H_0 verdadeira, chamado $\ell(\hat{\phi}, \chi_0)$. O cálculo da quantidade $\ell(\hat{\phi}, \chi_0)$ é simples, bastando maximizar ℓ sob a restrição H_0 . [7] propôs o uso da razão entre a verossimilhança máxima e a verossimilhança sob a hipótese H_0 , ou seja, a razão entre $\ell(\hat{\phi}, \hat{\chi})$ e $\ell(\hat{\phi}, \chi)$, como estatística para testar H_0 .

Neyman e Pearson definiram rigorosamente e estudaram o uso da razão de verossimilhanças para testes de hipóteses, fazendo uso do dobro do logaritmo da razão de verossimilhanças:

$$w(\chi_0) = 2 \ln \left\{ \frac{L(\hat{\phi}, \hat{\chi})}{L(\hat{\phi}, \chi_0)} \right\} = 2\{\ell(\hat{\phi}, \chi) - \ell(\hat{\phi}, \chi_0)\}$$

que tem distribuição assintótica χ^2 maximizada sob a hipótese em teste, sendo p a dimensão de ξ .

2 *Distribuição de Conway-Maxwell-Poisson*

Na análise de dados provenientes de processos de contagem, os modelos paramétricos binomial, multinomial, binomial negativa e de Poisson sempre dominaram os estudos, em grande parte devido à ênfase dada à estas distribuições nos cursos de probabilidade e a simplicidade dos cálculos envolvendo estas distribuições. Quando a contagem é ilimitada (por exemplo, o número de chamadas que chegam a estação telefônica de um hospital em um dia como função do número de acidentes de trânsito ocorridos naquele dia), dos dois modelos citados, apenas o modelo de Poisson tem o suporte infinito necessário para analisar os dados. Porém, isto não é tudo que é necessário para estudar este tipo de dados, pois o modelo de Poisson se restringe à casos em que a média da contagem é igual à variância da contagem, o que é bastante raro na prática da estatística.

Para problemas de contagem em que $E[X] \neq V[X]$, precisamos utilizar outras distribuições. Neste capítulo, estuda-se as propriedades da distribuição de Conway-Maxwell-Poisson [2], uma distribuição que permite formas bastante liberais de sobredispersão e subdispersão. A distribuição de Conway-Maxwell-Poisson, também chamada de *COM-Poisson* e *CMP*, é definida como:

$$P(X = x) = f(x) = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^x}{(x!)^\nu}, \text{ em que } Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \quad (2.1)$$

com $\lambda > 0, \nu \geq 0, x \in \mathbb{N} = \{0, 1, 2, 3, \dots, n, \dots\}$.

Para mostrar que a equação (2.1) realmente representa uma função de densidade de probabilidade, é necessário provar que ela é maior ou igual à zero em todo seu domínio

e que a integral dela sobre o domínio é igual à um. Condição de positividade é óbvia, pois a função Z definida é uma soma de termos positivos e f é a razão entre uma quantidade positiva e Z . A integral, como é usual em distribuições discretas, usa a medida de contagem, o que implica a relação

$$\int_0^{\infty} f(x)d\mu = \sum_{j=0}^{\infty} f(j) \quad (2.2)$$

r, então, a integral desejada é

$$\int_0^{\infty} f(x)d\mu = \sum_{x=0}^{\infty} \frac{1}{Z(\lambda, \nu)} \frac{\lambda^x}{(x!)^{\nu}} = \frac{1}{Z(\lambda, \nu)} \sum_{x=0}^{\infty} \frac{\lambda^x}{(x!)^{\nu}} = \frac{1}{Z(\lambda, \nu)} Z(\lambda, \nu) = 1 \quad (2.3)$$

2.1 Propriedades Conhecidas e Função de Verossimilhança

Os primeiros momentos da distribuição são derivadas de $\ln Z(\lambda, \nu)$

$$E[X] = \frac{1}{Z(\lambda, \nu)} \sum_{x=0}^{\infty} \frac{x\lambda^x}{(x!)^{\nu}} = \frac{\partial \ln Z}{\partial \lambda} \quad (2.4)$$

$$V[X] = E[X^2] - E^2[X] = \frac{\partial^2 \ln Z}{\partial \lambda^2} \quad (2.5)$$

Nenhum momento tem expressão fechada, mas a média pode ser aproximada de várias formas, incluindo usar a moda (de acordo com [17]), usando apenas os primeiros termos quando ν é grande, limitando $E[Y]$ quando ν é pequeno e usando uma expansão assintótica para Z . O paper [24] usa esta última abordagem, que será usada também nesta dissertação, obtendo a aproximação

$$E[Y] \approx \lambda^{\frac{1}{\nu}} + \frac{1}{2\nu} - \frac{1}{2} \quad (2.6)$$

que é particularmente precisa para $\lambda \geq 10^{\nu}$ ou $\nu < 1$.

De acordo com [24], o cálculo de características simples da distribuição, como sua função geradora de momentos ou sua função característica, é bastante difícil. De fato, basta verificar o artigo [23] dedicado à aproximação da função Z , dos momentos da distribuição e a geração de números aleatórios que seguem a distribuição COM-Poisson, mas

não chega a calcular a matriz de informação de Fisher associada a esta distribuição.

Uma função de verossimilhança de uma amostra de variáveis independente e identicamente distribuídas com a distribuição CMP de parâmetros (λ, ν) vai ser uma função proporcional à

$$L(\lambda, \nu) = \lambda^{\sum_{i=1}^n x_i} Z^{-n}(\lambda, \nu) \left(\prod_{i=1}^n x_i! \right)^{-\nu} \quad (2.7)$$

Maximizar L equivale a maximizar a função de log-verossimilhança, dada por

$$\ell(\lambda, \nu) = \ln L(\lambda, \nu) = \ln(\lambda) \sum_{i=1}^n x_i - \nu \sum_{i=1}^n \ln(x_i!) - n \ln Z(\lambda, \nu). \quad (2.8)$$

Denotando por Z_θ a derivada de Z com relação ao parâmetro θ , a função escore associada será

$$U(\lambda, \nu) = \begin{bmatrix} \frac{\partial \ell}{\partial \lambda} \\ \frac{\partial \ell}{\partial \nu} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^n x_i}{\lambda} - \frac{nZ_\lambda(\lambda, \nu)}{Z(\lambda, \nu)} \\ -\sum_{i=1}^n \ln(x_i!) - \frac{nZ_\nu(\lambda, \nu)}{Z(\lambda, \nu)} \end{bmatrix}. \quad (2.9)$$

A matriz Hessiana de ℓ , omitindo o elemento (λ, ν) da notação, é

$$\begin{aligned} J &= \begin{bmatrix} \frac{\partial^2 \ell}{\partial \lambda^2} & \frac{\partial^2 \ell}{\partial \lambda \partial \nu} \\ \frac{\partial^2 \ell}{\partial \nu \lambda} & \frac{\partial^2 \ell}{\partial \nu^2} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{\sum_{i=1}^n x_i}{\lambda^2} - \frac{nZ_{\lambda\lambda}}{Z} + \frac{nZ_\lambda Z_\lambda}{Z^2} & -\frac{nZ_{\lambda\nu}}{Z} + \frac{nZ_\lambda Z_\nu}{Z^2} \\ -\frac{nZ_{\nu\lambda}}{Z} + \frac{nZ_\nu Z_\lambda}{Z^2} & -\frac{nZ_{\nu\nu}}{Z} + \frac{nZ_\nu Z_\nu}{Z^2} \end{bmatrix} \end{aligned} \quad (2.10)$$

Disso, obtém-se a matriz de informação

$$I = -E[J] = - \begin{bmatrix} -\frac{nE[X]}{\lambda^2} - \frac{nZ_{\lambda\lambda}}{Z} + \frac{nZ_\lambda Z_\lambda}{Z^2} & 0 \\ 0 & -\frac{nZ_{\nu\nu}}{Z} + \frac{nZ_\nu Z_\nu}{Z^2} \end{bmatrix} \quad (2.11)$$

Expandindo as derivadas da função Z , os termos fora da diagonal principal se tornam nulos, o que faz com que os parâmetros sejam ortogonais.

2.2 Viés de Primeira Ordem das Estimativas de Máxima Verossimilhança

Calculando os termos do viés de primeira ordem da estimativa de máxima verossimilhança, equação (4.20), para o caso da distribuição CMP, tem-se os itens

$$\begin{aligned}
v_{\lambda\lambda\lambda} &= E \left[\frac{\partial^3 \ell}{\partial \lambda^3} \right] = \frac{2nE[X]}{\lambda^3} - \frac{nZ_{\lambda\lambda\lambda}}{Z} + \frac{3nZ_{\lambda\lambda}}{Z^2} - \frac{2nZ_{\lambda}Z_{\lambda}Z_{\lambda}}{Z^3} \\
v_{\lambda\lambda\nu} &= E \left[\frac{\partial^3 \ell}{\partial \lambda^2 \partial \nu} \right] = -\frac{nZ_{\lambda\lambda\nu}}{Z} + \frac{nZ_{\lambda\lambda}Z_{\nu}}{Z^2} + \frac{2nZ_{\lambda}Z_{\lambda\nu}}{Z^2} - \frac{2nZ_{\lambda}Z_{\lambda}Z_{\nu}}{Z^3} \\
v_{\lambda\nu\lambda} &= E \left[\frac{\partial^3 \ell}{\partial \lambda \partial \nu \partial \lambda} \right] = -\frac{nZ_{\lambda\nu\lambda}}{Z} + \frac{2nZ_{\lambda\nu}Z_{\lambda}}{Z^2} + \frac{nZ_{\lambda\lambda}Z_{\nu}}{Z^2} - \frac{2nZ_{\lambda}Z_{\nu}Z_{\lambda}}{Z^3} \\
v_{\lambda\nu\nu} &= E \left[\frac{\partial^3 \ell}{\partial \lambda \partial \nu^2} \right] = -\frac{nZ_{\lambda\nu\nu}}{Z} + \frac{2nZ_{\lambda\nu}Z_{\nu}}{Z^2} + \frac{nZ_{\lambda}Z_{\nu\nu}}{Z^2} - \frac{2Z_{\lambda}Z_{\nu}Z_{\nu}}{Z^3} \\
v_{\nu\lambda\lambda} &= E \left[\frac{\partial^3 \ell}{\partial \nu \partial \lambda^2} \right] = -\frac{nZ_{\lambda\nu\lambda}}{Z} + \frac{2nZ_{\lambda\nu}Z_{\lambda}}{Z^2} + \frac{nZ_{\lambda\lambda}Z_{\nu}}{Z^2} - \frac{2nZ_{\lambda}Z_{\nu}Z_{\lambda}}{Z^3} \\
v_{\nu\lambda\nu} &= E \left[\frac{\partial^3 \ell}{\partial \nu \partial \lambda \partial \nu} \right] = -\frac{nZ_{\lambda\nu\nu}}{Z} + \frac{2nZ_{\lambda\nu}Z_{\nu}}{Z^2} + \frac{nZ_{\lambda}Z_{\nu\nu}}{Z^2} - \frac{2Z_{\lambda}Z_{\nu}Z_{\nu}}{Z^3} \\
v_{\nu\nu\lambda} &= E \left[\frac{\partial^3 \ell}{\partial \nu^2 \partial \lambda} \right] = -\frac{nZ_{\lambda\nu\nu}}{Z} + \frac{2nZ_{\lambda\nu}Z_{\nu}}{Z^2} + \frac{nZ_{\lambda}Z_{\nu\nu}}{Z^2} - \frac{2Z_{\lambda}Z_{\nu}Z_{\nu}}{Z^3} \\
v_{\nu\nu\nu} &= E \left[\frac{\partial^3 \ell}{\partial \nu^3} \right] = -\frac{nZ_{\nu\nu\nu}}{Z} + \frac{3nZ_{\nu\nu}Z_{\nu}}{Z^2} - \frac{2nZ_{\nu}Z_{\nu}Z_{\nu}}{Z^3}
\end{aligned}$$

e os cumulantes

$$\begin{aligned}
v_{\lambda\lambda,\lambda} &= E \left[\frac{\partial^2 \ell}{\partial \lambda^2} \frac{\partial \ell}{\partial \lambda} \right] = -\frac{n^2 E[\bar{X}^2]}{\lambda^3} + \frac{n^2 E[\bar{X}] Z_{\lambda}}{\lambda^2 Z} - \frac{n^2 E[\bar{X}] Z_{\lambda\lambda}}{\lambda Z} \\
&\quad + \frac{n^2 E[\bar{X}] Z_{\lambda} Z_{\lambda}}{\lambda Z^2} + \frac{n^2 Z_{\lambda} Z_{\lambda\lambda}}{Z^2} - \frac{n^2 Z_{\lambda} Z_{\lambda} Z_{\lambda}}{Z^3} \\
v_{\lambda\lambda,\nu} &= E \left[\frac{\partial^2 \ell}{\partial \lambda^2} \frac{\partial \ell}{\partial \nu} \right] = \frac{n^2 E[\bar{X} \ln \bar{X}!]}{\lambda^2} + \frac{n^2 E[\bar{X}] Z_{\nu}}{\lambda^2 Z} - \frac{n^2 E[\ln \bar{X}!] Z_{\lambda\lambda}}{Z} \\
&\quad - \frac{n^2 E[\ln \bar{X}!] Z_{\lambda} Z_{\lambda} - n^2 Z_{\nu} Z_{\lambda\lambda}}{Z^2} - \frac{n^2 Z_{\lambda} Z_{\lambda} Z_{\lambda}}{Z^3} \\
v_{\nu\nu,\lambda} &= E \left[\frac{\partial^2 \ell}{\partial \nu^2} \frac{\partial \ell}{\partial \lambda} \right] = -\frac{n E[\bar{X}] Z_{\nu\nu}}{\lambda Z} + \frac{n^2 E[\bar{X}] Z_{\nu} Z_{\nu}}{\lambda Z^2} + \frac{n Z_{\lambda} Z_{\nu\nu}}{Z^2} - \frac{n^2 Z_{\lambda} Z_{\nu} Z_{\nu}}{Z^3} \\
v_{\nu\nu,\nu} &= E \left[\frac{\partial^2 \ell}{\partial \nu^2} \frac{\partial^2 \ell}{\partial \nu} \right] = \frac{n^2 E[\ln \bar{X}!] Z_{\nu\nu}}{Z} - \frac{n^2 E[\ln \bar{X}!] Z_{\nu} Z_{\nu} - n^2 Z_{\nu} Z_{\nu\nu}}{Z^2} - \frac{n^2 Z_{\nu} Z_{\nu} Z_{\nu}}{Z^3}
\end{aligned}$$

sendo que $\ln \bar{x}! = \frac{\sum_{j=1}^{\bar{x}} \ln x_j!}{n}$. Os seguinte cumulantes, também usados, são todos iguais à

zero:

$$v_{\lambda\nu,\lambda} = E \left[\frac{\partial^2 \ell}{\partial \lambda \nu} \frac{\partial \ell}{\partial \lambda} \right] \quad v_{\lambda\nu,\nu} = E \left[\frac{\partial^2 \ell}{\partial \lambda \nu} \frac{\partial^2 \ell}{\partial \nu} \right]$$

$$v_{\nu\lambda,\lambda} = E \left[\frac{\partial^2 \ell}{\partial \nu \lambda} \frac{\partial \ell}{\partial \lambda} \right] \quad v_{\nu\lambda,\nu} = E \left[\frac{\partial^2 \ell}{\partial \nu \lambda} \frac{\partial^2 \ell}{\partial \nu} \right]$$

O passo seguinte é inserir estas quantidades nas equações 4.21 e 4.22. Na distribuição COM-Poisson, temos que $i \neq j \Rightarrow i^{ij} = 0$ e $i \neq j \Rightarrow v_{ij,k} = 0$, o que reduz as equações à:

$$2B(\theta_\lambda) = i^{\lambda\lambda} i^{\lambda\lambda} (v_{\lambda\lambda\lambda} + 2v_{\lambda\lambda,\lambda}) + i^{\lambda\lambda} i^{\nu\nu} v_{\lambda\nu\nu} + O\left(n^{-\frac{3}{2}}\right) \quad (2.12)$$

$$2B(\theta_\nu) = i^{\nu\nu} i^{\lambda\lambda} v_{\nu\lambda\lambda} + i^{\nu\nu} i^{\nu\nu} (v_{\nu\nu\nu} + 2v_{\nu\nu,\nu}) + O\left(n^{-\frac{3}{2}}\right) \quad (2.13)$$

Embora não seja possível obter uma forma fechada para este viés, devido à função Z e suas derivadas serem todas definidas em termo de uma série infinita, as equações A.1 e A.2 são suficientes para avaliar numericamente a ordem do viés.

2.3 Estimação dos Parâmetros por Mínimos Quadrados

Uma das propriedades interessantes desta distribuição surge quando calcula-se a razão entre $P(X = x + 1 | \lambda, \nu)$ e $P(X = x | \lambda, \nu)$:

$$\frac{f(x + 1 | \lambda, \nu)}{f(x | \lambda, \nu)} = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^{x+1}}{((x + 1)!)^\nu} Z(\lambda, \nu) \frac{(x!)^\nu}{\lambda^x} = \frac{\lambda}{(x + 1)^\nu} \quad (2.14)$$

Logaritmando essa razão, chega-se à

$$\ln \left(\frac{P(x + 1)}{P(x)} \right) = \ln \lambda - \nu \ln(x + 1) \quad (2.15)$$

Escrevendo $\beta_0 = \ln \lambda$ e $\beta_1 = -\nu$, chega-se à

$$\ln \left(\frac{P(x + 1)}{P(x)} \right) = \beta_0 + \beta_1 \ln(x + 1), \quad (2.16)$$

uma forma similar à um modelo de regressão linear simples. As duas probabilidades do lado esquerdo da equação (2.16) são facilmente estimadas a partir de uma amostra, e de posse destas estimativas pode-se estimar os β utilizando mínimos quadrados ordinários ou um método equivalente para modelos de regressão linear simples, e a partir destas

estimativas obter estimativas dos parâmetros originais. O artigo de *Shmueli et alii* ([24]) sugere que estas estimativas são inferiores às de máxima verossimilhança, mas que são bons pontos de partida para métodos iterativos de maximização da função de verossimilhança.

Para obter uma avaliação rápida da qualidade desta estimação por mínimos quadrados utilizando a linguagem de programação R ([22]), foram geradas (através do equipamento para simulação de Monte Carlo que acompanha o pacote *stats* do interpretador padrão do R) amostras independentes e identicamente distribuídas da distribuição de Poisson com parâmetro $\lambda = 1$, sendo que 100 amostras tem tamanho 10, 100 tem tamanho 100, 100 tem tamanho 1000 e assim por diante até o tamanho 100000. O código para gerar estas amostras está na seção [A.3](#).

Foi computada então a média das estimativas dos parâmetros para cada tamanho de amostra, que devem ser próximas à $(\lambda = 1, \nu = 1)$. O processo foi repetido com $\lambda = 10$ e com $(\lambda = 0.25, \nu = 0)$, e os dois primeiros momentos amostrais junto com a média das estimativas obtidas para cada parâmetro estão nas Tabelas [1](#), [2](#) e [3](#).

Na Tabela [1](#) pode-se verificar que as estimativas para amostras de tamanho inferior à 1000 tem um comportamento caótico, tornando este processo de estimação não aplicável. É perceptível que, de acordo com a variância e média amostrais, a estimação por máxima verossimilhança também não produziria resultados mais interessantes, pois os primeiros momentos amostrais não condizem com os momentos da distribuição de Poisson com parâmetro $\lambda = 1$. Isto é devido ao fato que para estes tamanhos pequenos de amostra as características peculiares às distribuições de Poisson e COM-Poisson não aparecem. Na Tabela [2](#) pode-se verificar que, novamente, as estimativas para amostras de tamanho inferior à 1000 são inadmissíveis. O problema com amostras pequenas persiste com um aumento de 10 vezes no parâmetro λ , o que contradiz [24], que afirma que se $\lambda \geq 10^\nu$ a estimação por mínimos quadrados é admissível mesmo com amostras pequenas.

Outro caso comentado por [24] é com $\nu < 1$, que é ilustrado através de simulação de variáveis com distribuição geométrica na Tabela [3](#). O valor de λ utilizado aqui foi de 0.25.

Se relevarmos os valores inadmissíveis (pois são negativos) nas estimativas de ν , o que não é de todo incompreensível tendo em vista que o valor absoluto destas estimativas é praticamente zero, as estimativas utilizando mínimos quadrados são melhores que as estimativas de máxima verossimilhança (obtidas calculando o inverso das médias mostradas na tabela) para tamanhos de amostra grandes.

Vale ressaltar que nesta estimação por mínimos quadrados, não há nenhuma garantia que as suposições usuais de modelos de regressão linear sejam válidas. Como

$$Cov \left(\ln \left(\frac{\hat{P}(x+1)}{\hat{P}(x)} \right), \ln \left(\frac{\hat{P}(x)}{\hat{P}(x-1)} \right) \right) \approx -\frac{1}{nP(x)} \quad (2.17)$$

temos algo similar a correlação serial e heterocedasticidade.

Tabela 1: Resultados obtidos com amostras da distribuição de Poisson ($\lambda = 1$), Média da Amostra (estimador de máxima verossimilhança para λ utilizando a distribuição de Poisson, Variância da Amostra, e estimativas de λ e ν

Tamanho da Amostra	Média Amostral	Variância Amostral	$\hat{\lambda}$	$\hat{\nu}$
10	1.50	1.38	0.00	-6.12
100	0.88	0.83	4.04	2.57
1000	1.01	0.99	0.66	0.58
10000	0.99	0.99	0.87	0.87
100000	0.99	1.00	0.90	0.90

Tabela 2: Resultados obtidos com amostras da distribuição de Poisson ($\lambda = 10$), Média da Amostra (estimador de máxima verossimilhança para λ utilizando a distribuição de Poisson, Variância da Amostra, e estimativas de λ e ν

Tamanho da Amostra	Média Amostral	Variância Amostral	$\hat{\lambda}$	$\hat{\nu}$
10	9.10	3.21	0.28	-0.67
100	10.16	11.32	7.55	0.87
1000	9.91	9.90	9.83	0.98
10000	10.03	9.93	9.65	1.00
100000	9.97	10.04	10.72	1.05

Tabela 3: Resultados obtidos com amostras da distribuição Geométrica ($P(\text{sucesso}) = \lambda = 1/4$), Média da Amostra (o estimador de máxima verossimilhança para a distribuição geométrica é $1/\bar{X}$), e estimativas de λ e ν

Tamanho da Amostra	Média Amostral	$1/\bar{X}$	$\hat{\lambda}$	$\hat{\nu}$
10	3.30	0.30	0.78	-1.18
100	3.40	0.29	0.20	0.01
1000	3.04	0.33	0.27	-0.03
10000	2.91	0.34	0.29	-0.02
100000	3.00	0.33	0.26	-0.01

3 *Distribuições Estendidas por Alternativas de Lehmann*

3.1 Introdução

No contexto de modelos paramétricos para o estudo de tempos de vida, [14] popularizaram o estudo de distribuições estendidas por alternativas de Lehmann, que são distribuições que tem uma das formas

$$G_1(x; \lambda) = [F(x)]^\lambda \quad (3.1)$$

ou

$$G_2(x; \lambda) = 1 - [1 - F(x)]^\lambda, \quad (3.2)$$

em que $F(x)$ é uma função de distribuição acumulada qualquer e $\lambda > 0$. É fácil verificar que o suporte da distribuição estendida a partir de F é o mesmo de F e que as funções de densidade associadas são:

$$g_1(x, \lambda) = \lambda f(x) F(x)^{\lambda-1} \quad (3.3)$$

e

$$g_2(x, \lambda) = \lambda f(x) [1 - F(x)]^{\lambda-1} \quad (3.4)$$

em que $f(x)$ é a função de densidade associada à F .

Várias distribuições comumente usadas podem ser vistas como extensões de distribuições simplórias. Uma função de distribuição comumente usada que pode ser vista como uma distribuição estendida desta forma é a distribuição exponencial, que é a extensão,

pela segunda forma, da distribuição $F(x) = 1 - e^{-x}$. Ao estender a distribuição uniforme no intervalo $[0, 1]$ pela primeira forma (equações (3.1) e 3.3)), obtém-se a distribuição beta com parâmetros $(\alpha = \lambda, \beta = 1)$, que tem a forma:

$$f(x|\lambda) = \lambda x^{\lambda-1} \quad (3.5)$$

em que $x \in (0, 1)$ e $\lambda > 0$. Se estendermos pela segunda forma (equações (3.2) e (3.4)), obtém-se novamente a distribuição beta, porém desta vez com parâmetros $(\alpha = 1, \beta = \lambda)$, obtendo a forma:

$$f(x|\lambda) = \lambda [1 - x]^{\lambda-1} \quad (3.6)$$

em que $x \in (0, 1)$, $\lambda > 0$. As equações (3.5) e (3.6) são casos especiais da distribuição

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} \quad (3.7)$$

em que $x \in (0, 1)$, $\alpha > 0$ e $\beta > 0$.

3.2 Momentos - Cálculo Usando a Função Quantil

Fazendo uma mudança de variáveis $u = F(x)$ na expressão do k -ésimo momento de g , a saber

$$E[X^k|\lambda] = \mu_k = \int_{-\infty}^{\infty} x^k \lambda f(x) [F(x)]^{\lambda-1} dx \quad (3.8)$$

chega-se à

$$\mu_k = \int_0^1 Q^k(u) \lambda u^{\lambda-1} du \quad (3.9)$$

em que $Q(u) = F^{-1}(u)$ é a inversa da função de distribuição acumulada F , chamada de função quantil.

A função dentro da integral da equação (3.9), $Q^k(u) \lambda u^{\lambda-1}$, pode ser dividida em duas partes: $Q^k(u)$ e $\lambda u^{\lambda-1}$. Esta segunda parte é a equação (3.5), que é um caso especial da distribuição beta mostrada na equação (3.7). Então, pode-se reduzir o k -ésimo momento

à uma esperança da distribuição beta:

$$\mu_k = \int_0^1 \lambda Q^k(u) u^{\lambda-1} du = E_{\text{Beta}(\lambda,1)} [Q(u)]. \quad (3.10)$$

Raciocínio similar mostra que os momentos da segunda expansão são dados pela equação:

$$\mu_k = \int_0^1 \lambda Q^k(u) u^{\lambda-1} du = E_{\text{Beta}(1,\lambda)} [Q(u)]. \quad (3.11)$$

Combinando as equações (3.10) e (3.11) com a definição de função geradora de momentos, obtém-se que as funções geradoras de momentos das funções estendidas são dadas por:

$$M_1(t) = E_{\text{Beta}(1,\lambda)} \{ \exp [tQ(u)] \} \quad \text{e} \quad M_2(t) = E_{\text{Beta}(\lambda,1)} \{ \exp [tQ(u)] \} \quad (3.12)$$

3.3 Caracterização da Segunda Forma em Termos de Momentos Ponderados por Probabilidade

Expandindo a segunda forma de extensão, dada pelas equações (3.2) e (3.4)), em série infinita, pode-se obter caracterizações adicionais baseadas em fatoriais descendentes e Momentos Ponderados por Probabilidade (comumente denominado *Probability Weighted Moments* ou PWM). Primeiramente, utilizando a série de Taylor ao redor do ponto zero, verifica-se que

$$(1 - z)^\alpha = \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + 1 - j)} \frac{z^j}{j!}. \quad (3.13)$$

Fazendo $z = F(x)$ e $\alpha = \lambda$, tem-se que

$$(1 - z)^\alpha = (1 - F(x))^\lambda = G(x|\lambda) = \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda + 1 - j)} \frac{F(x)^j}{j!}, \quad (3.14)$$

e substituindo na equação (3.4), verifica-se que:

$$\begin{aligned} g(x|\lambda) &= \lambda f(x) \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(\lambda)}{\Gamma(\lambda - j)} \frac{F(x)^j}{j!} \\ &= f(x) \Gamma(\lambda) \sum_{j=0}^{\infty} \frac{(-1)^j}{\Gamma(\lambda - j)} \frac{F(x)^j}{j!}. \end{aligned} \quad (3.15)$$

Seja $M_k = \max\{X_1, X_2, \dots, X_k\}$ com $X_i \sim F$. A densidade de M_k é $f_{M_k}(x) = kf(x)F^{k-1}(x)$.

Fazendo $j = k - 1$ na equação (3.15), obtém-se

$$\begin{aligned} g(x|\lambda) &= f(x)\Gamma(\lambda) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{\Gamma(\lambda - k + 1)} \frac{F(x)^{k-1}}{(k-1)!} \\ &= \Gamma(\lambda) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{\Gamma(\lambda - k + 1)} \frac{f(x)F(x)^{k-1}}{(k-1)!} \\ &= \Gamma(\lambda) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{\Gamma(\lambda - k + 1)} \frac{f_{M_k}(x)}{k!} \end{aligned} \quad (3.16)$$

Definindo a quantidade

$$w_k = \frac{\Gamma(\lambda+1)(-1)^{k-1}}{k!\Gamma(\lambda+1-k)} = \frac{(-1)^{k-1}\lambda^k}{k!} \quad (3.17)$$

a expressão (3.16) se reduz à:

$$g(x|\lambda) = \sum_{k=1}^{\infty} w_k f_{M_k}(x) \quad (3.18)$$

Destas equações tem-se, respectivamente, a fórmula dos momentos e a fórmula da geratriz de momentos da função g :

$$\mu_r = \sum_{k=1}^{\infty} w_k \mu_{M_k} \quad (3.19)$$

$$Mg(t) = \sum_{k=1}^{\infty} w_k Mg_{M_k} \quad (3.20)$$

em que Mg_{M_k} é a função geradora de momentos de M_k . Logo, segue que em conjunto com a equação (3.11) tem-se que o r -ésimo momento de g é dado por:

$$\mu_r = \sum_{k=1}^{\infty} kw_k E_f\{X^r F(X)^{k-1}\} \quad (3.21)$$

$$= \sum_{k=1}^{\infty} kw_k \beta_{r,k-1} \quad (3.22)$$

em que $\beta_{r,k-1}$ são os Probability Weighted Moments (PWM ou ainda Momentos Ponderados por Probabilidade) de f .

3.4 Perda de Poder em Erros de Especificação de Distribuições Extendidas pela Primeira Alternativa

3.4.1 Divergência Kullback-Leibler

Dadas duas pdfs, [16] chamam a quantidade

$$D_{KL}(f|g) = \int_{\mathbb{R}} f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx \quad (3.23)$$

de Divergência de Kullback-Leibler (DKL). Frequentemente, esta quantidade é usada como medida de distância entre pdfs, mesmo não sendo uma métrica. Esta medida de divergência é sempre maior ou igual à zero com zero ocorrendo se e somente se $f = g$, mas não é simétrica pois $D_{KL}(f|g) \neq D_{KL}(g|f)$, além de não obedecer a desigualdade triangular.

Reescrevendo a equação (3.23), nós temos

$$\int_{\mathbb{R}} f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx = \int_{\mathbb{R}} f(x) \ln(f(x)) - f(x) \ln(g(x)) dx \quad (3.24)$$

$$= E_f [\ln(f(X))] - E_f [\ln(g(X))] \quad (3.25)$$

onde $E_f [h(X)]$ é a esperança da variável aleatória $h(X)$ com respeito à pdf f . Como $D_{KL}(f|g)$ é maior que zero, temos que

$$E_f [\ln(f(X))] > E_f [\ln(g(X))] \quad (3.26)$$

Maximizar a verossimilhança é equivalente a minimizar $D_{KL}(f|e)$, onde e é a função de distribuição empírica. Calculando $D_{KL}(f|e)$ chega-se em

$$D_{KL}(f|e) = E_f [\ln(f(X))] - \frac{1}{n} \sum_{j=1}^n \ln(f(x_j, \theta)) \quad (3.27)$$

onde o termo mais a direita é a log-verossimilhança empírica multiplicada por uma constante. Então, maximizando este termo minimiza-se toda a divergência. Então o processo de maximizar a verossimilhança é equivalente à minimizar a divergência entre a densidade empírica e o modelo paramétrico. Este resultado é comum na literatura, e mostrado em

todos os detalhes em [5], que dá uma dedução acessível e compacta das propriedades de métodos baseados em funções de verossimilhança usando DKL. [5] também estabelecem um resultado que diz que a DKL é uma medida da perda de poder dos testes de hipóteses baseados na razão de verossimilhança quando há erro de especificação na distribuição da estatística de teste sobre a hipótese alternativa.

3.4.2 Especificação Errônea e Perda de Poder em Testes de Razão de Verossimilhanças

Supondo que temos dados vindos de uma fdp $H(x|\theta, \lambda)$, e queremos testar a hipótese de que $(\theta = \theta_0, \lambda = \lambda_0)$. A log-verossimilhança é dada pela equação

$$\Lambda(\lambda_0, \theta_0) = \frac{\ell(\hat{\lambda}, \hat{\theta})}{\ell(\lambda_0, \theta_0)} \quad (3.28)$$

em que a notação $\hat{\xi}$ é usada para a estimativa de máxima verossimilhança do parâmetro ξ sem restrições ao espaço paramétrico. Se não é possível calcular $\ell(\hat{\lambda}, \hat{\theta})$ porquê a estimativa do parâmetro λ é complicada e aproxima-se a estatística da razão de verossimilhanças usando $\ell(\lambda_1, \tilde{\theta})$ no lugar da verossimilhança sobre a hipótese alternativa, em que $\tilde{\theta}$ é o estimador de máxima verossimilhança de θ dado que $\lambda = \lambda_1$. Então obtém-se a relação

$$\Lambda(x) \approx \frac{\ell(\lambda_1, \tilde{\theta})}{\ell(\lambda_0, \theta_0)} \quad (3.29)$$

Um resultado de [5], seção 3, estabelece que a estatística de teste gerada desta forma é menos poderosa que a usual, e a perda de poder é igual à

$$\Delta_{\text{Power}} = D_{KL} \left(f(x|\hat{\lambda}, \hat{\theta}), f(x|\lambda_1, \tilde{\theta}) \right) \quad (3.30)$$

No caso em que os dados seguem uma distribuição estendida pela primeira alternativa de Lehmann com distribuição original tal que $F = F(x|\theta)$ para um parâmetro θ . Sendo a hipótese nula da forma

$$H_0 : \theta = \theta_0, \lambda = 1 \quad (3.31)$$

contra uma hipótese alternativa

$$H_A : \theta \neq \theta_0, \lambda \neq 1 \quad (3.32)$$

Se for considerado que os dados não seguem uma distribuição extendida da forma $G(x|\lambda, \theta)$, mas de uma população que segue a distribuição original $F(x|\theta)$, isso equivale a aproximar a log-verossimilhança sobre a hipótese alternativa como no parágrafo anterior. Neste caso, a log-verossimilhança será tomada sobre a hipótese

$$H_{A'} : \theta \neq \theta_0, \lambda = 1 \quad (3.33)$$

o que gera a seguinte expressão para a log-verossimilhança:

$$\Lambda(x) \approx \frac{\ell(1, \tilde{\theta})}{\ell(1, \theta_0)} \quad (3.34)$$

Então tem-se que o teste tem menos poder que um usando a distribuição G , e a diferença no poder do teste é dada por

$$\Delta_{\text{Power}} = D_{KL} \left(g(x|\hat{\lambda}, \hat{\theta}) | g(x|1, \tilde{\theta}) \right) \quad (3.35)$$

Então para testar hipóteses sobre o parâmetro ξ original, testes usando a versão extendida são sempre mais poderosos, com uma diferença considerável na taxa de erros tipo II.

Expandindo a equação (3.35) obtém-se que

$$\Delta_P = D_{KL} \left(g(x|\hat{\lambda}, \hat{\theta}) | g(x|1, \tilde{\theta}) \right) \quad (3.36)$$

$$= \int_{\mathbb{R}} g(x|\hat{\lambda}, \hat{\theta}) \ln \left(\frac{g(x|\hat{\lambda}, \hat{\theta})}{g(x|1, \tilde{\theta})} \right) dx \quad (3.37)$$

$$= \int_{\mathbb{R}} \lambda f(x|\hat{\lambda}, \hat{\theta}) F^{\lambda-1}(x|\hat{\lambda}, \hat{\theta}) \ln \left(\frac{\lambda f(x|\hat{\lambda}, \hat{\theta}) F^{\lambda-1}(x|\hat{\lambda}, \hat{\theta})}{f(x|1, \tilde{\theta})} \right) dx \quad (3.38)$$

$$= \int_{\mathbb{R}} \lambda f(x|\hat{\lambda}, \hat{\theta}) F^{\lambda-1}(x|\hat{\lambda}, \hat{\theta}) \ln \left(\lambda F^{\lambda-1}(x|\hat{\lambda}, \hat{\theta}) \right) dx \quad (3.39)$$

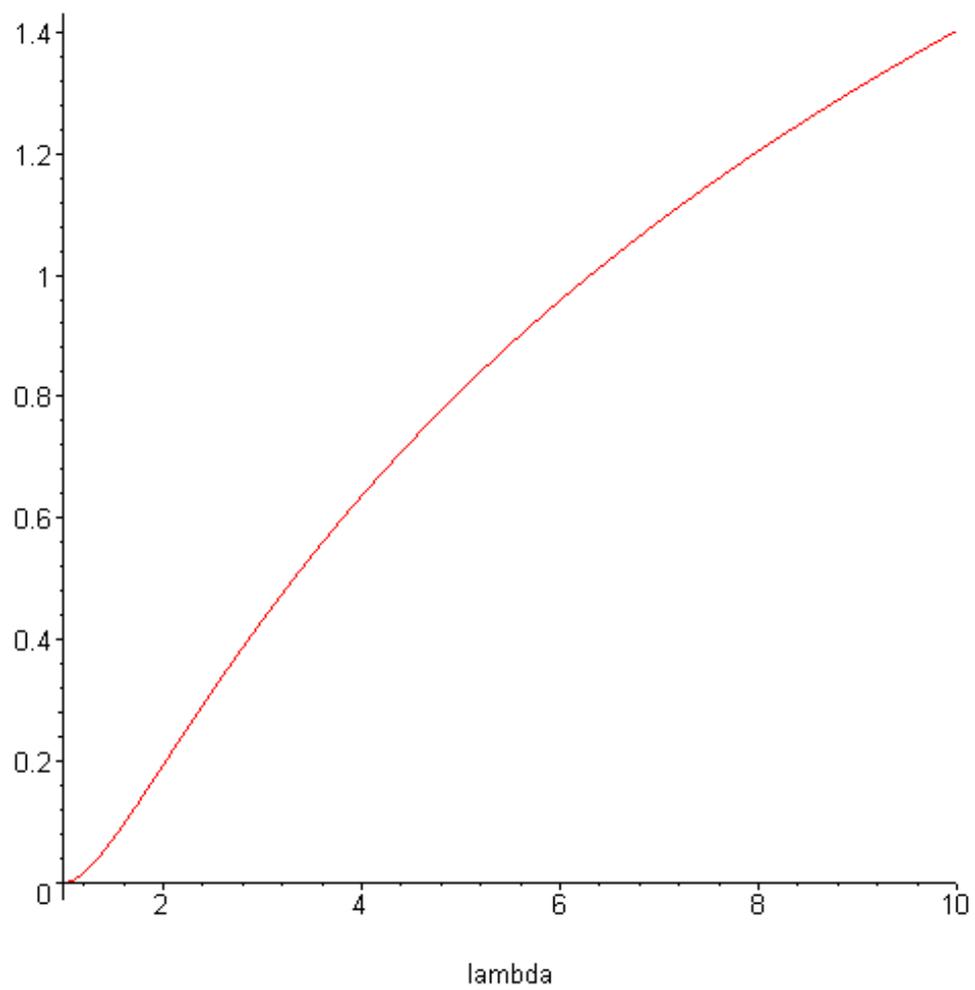
$$= \ln \lambda + \int_{\mathbb{R}} \lambda(\lambda - 1) f(x|\hat{\lambda}, \hat{\theta}) F^{\lambda-1}(x|\hat{\lambda}, \hat{\theta}) \ln \left(F(x|\hat{\lambda}, \hat{\theta}) \right) dx \quad (3.40)$$

Integrando por partes, obtém-se

$$\Delta_{\text{Power}} = \ln \lambda + \frac{1 - \lambda}{\lambda} \quad (3.41)$$

O gráfico desta função mostra a perda de poder que tem-se no teste quando a distribuição dos dados é uma extendida pela primeira alternativa de Lehmann e falha-se em perceber isto, sendo mostrado na Figura 1 para valores de λ maiores que um.

Figura 1: Perda de Poder em Função de λ , para $\lambda > 1$.



4 *Distribuição da Razão de Verossimilhanças*

Neste capítulo, o interesse reside em sequências de hipóteses da forma

$$H_0 : \theta_0 = t_0$$

$$H_0 : \theta_1 = t_1 | \theta_0 = t_0$$

$$H_0 : \theta_2 = t_2 | \theta_1 = t_1, \theta_0 = t_0$$

$$H_0 : \theta_3 = t_3 | \theta_2 = t_2, \theta_1 = t_1, \theta_0 = t_0$$

...

$$H_0 : \theta_k = t_k | \theta_{k-1} = t_{k-1}, \dots, \theta_0 = t_0$$

utilizando as estatísticas r_k , em que k é a k -ésima hipótese da lista acima (definidas em [1], página 101), que tem a forma:

$$r_k = \frac{\theta_k - b_k}{|\theta_k - b_k|} \sqrt{w(\chi = b_k)} \quad (4.1)$$

De acordo com Barndorff-Nielsen e Cox (1989), página 154, pode-se calcular as esperanças e covariâncias assintóticas destas estatísticas reescrevendo-as em forma matricial como

$$R = I^{-\frac{1}{2}} S, \quad R = \begin{pmatrix} r_1 \\ r_2 \\ \dots \\ r_p \end{pmatrix}$$

em que $I^{-\frac{1}{2}}$ é uma matriz tal que $(I^{-\frac{1}{2}})^2 = I^{-\frac{1}{2}}I^{-\frac{1}{2}} = I^{-1}$ é a inversa de $E[-J]$, sendo J a matrix Hessiana de $\ell(\theta)$, diferenciando com relação aos θ , e s é um vetor em que o r -ésimo elemento é dado por

$$\begin{aligned} s_r &= l_r + \\ &+ \frac{1}{6}(3H_{rs}l^s + v_{rst}l^s l^t) \\ &+ \frac{1}{72}(27H_{rs}H_{tu}i^{st}l^u + 30H_{rs}v_{tuv}i^{st}l^u l^v + 12H_{rst}l^s l^t + \\ &+ 3v_{rstu}l^s l^t l^u + 8v_{rst}v_{uvw}i^{tu}l^s l^v l^w) + \dots \end{aligned} \quad (4.2)$$

em que, para todo conjunto R de índices válidos para θ , $v_R = E[\ell_R]$, $H_R = \ell_R - v_R$ e, sendo $q = \#R$, $\ell^R = \frac{\partial^q \ell}{\partial \theta_{i_1} \partial \theta_{i_2} \dots \partial \theta_{i_q}}$. Aqui faz-se uso da convenção da soma de Einstein, comum em trabalhos com tensores, de forma que $a_{ij}x^j$ deve ser lido como $\sum_{j=j_0}^{j_1} a_{ij}x_j$ em que j_0 e j_1 ficarão implícitos pelo contexto. Nosso objetivo será deduzir as propriedades destas estatísticas para distribuições da família exponencial mostrada na equação (4.15).

4.1 Expansões Assintóticas

4.1.1 Expansões Assintóticas Importantes: $(\hat{\theta} - \theta)_r$

A expansão mostrada aqui é de fundamental importância no cálculo de viéses e esperança de estatísticas de razão e verossimilhanças. Consideremos a função $\ell_r(\theta)$, e expandimos ela em série de Taylor multivariada ao redor de um ponto arbitrário $\theta = \theta_0$, obtendo

$$\begin{aligned} \ell_r(\theta) &= \ell_r(\theta_0) + \frac{\ell_{rs}(\theta_0)(\theta - \theta_0)^s}{1!} + \frac{\ell_{rst}(\theta_0)(\theta - \theta_0)^s(\theta - \theta_0)^t}{2!} \\ &+ \frac{\ell_{rstu}(\theta_0)(\theta - \theta_0)^s(\theta - \theta_0)^t(\theta - \theta_0)^u}{3!} + \dots \end{aligned} \quad (4.3)$$

Se considerarmos apenas o primeiro termo, pode-se reescrever na forma matricial

$$\frac{d\ell}{d\theta}(\theta) = \frac{d\ell}{d\theta}(\theta_0) + \frac{d^2\ell}{d\theta^2}(\theta_0)(\theta - \theta_0) \quad (4.4)$$

em que $\frac{d\ell}{d\theta}$ é o vetor no qual a k -ésima linha é a derivada de ℓ em relação ao k -ésimo θ , $\frac{d^2\ell}{d\theta^2}$ é a matrix hessiana de ℓ ou, equivalentemente, a derivada de $\frac{d\ell}{d\theta}$. Estudando os pontos críticos de ℓ , tem-se

$$\ell_r(\theta) = 0 \quad (4.5)$$

tem-se que

$$\begin{aligned} (\theta - \theta_0) &= - \left[\frac{d^2\ell}{d\theta^2}(\theta_0) \right]^{-1} \frac{d\ell}{d\theta}(\theta_0) \\ \Rightarrow (\theta - \theta_0)_r &= - \left(\left[\frac{d^2\ell}{d\theta^2}(\theta_0) \right]^{-1} \right)^r \frac{d\ell}{d\theta}(\theta_0) \end{aligned} \quad (4.6)$$

$$\Rightarrow (\theta - \theta_0)_r = j^{rs} \ell_s(\theta_0) \quad (4.7)$$

em que os j^{rs} são os mesmos definidos nas preliminares. A fórmula (4.7) pode ser considerada como uma aproximação para a quantidade desejada, mas queremos refiná-la. Igualando (4.3) à zero, isolando o primeiro termo e resolvendo para $(\theta - \theta_0)_r$ obtém-se

$$\begin{aligned} (\theta - \theta_0)_r &= j^{rs} \ell_s(\theta_0) + \frac{\ell_{stu}(\theta_0)(\theta - \theta_0)^t(\theta - \theta_0)^u}{2!} \\ &+ \frac{\ell_{stuv}(\theta_0)(\theta - \theta_0)^t(\theta - \theta_0)^u(\theta - \theta_0)^v}{3!} + \dots \end{aligned} \quad (4.8)$$

Substituindo (4.7) no primeiro termo de (4.8) e ignorando os termos de ordem mais alta, obtém-se

$$(\theta - \theta_0)_r = j^{rs} \ell_s(\theta_0) + \frac{j^{rs} \ell_{stu}(\theta_0) j^{tv} \ell_v(\theta_0) j^{ux} \ell_x(\theta_0)}{2!} \quad (4.9)$$

que é uma expansão de segunda ordem para $(\hat{\theta} - \theta)_r$. Substituímos agora (4.9) no segundo termo de (4.8) e ignora-se os termos de ordem mais alta para obter a expansão de terceira ordem

$$\begin{aligned} (\theta - \theta_0)_r &= j^{rs} \ell_s(\theta_0) + \frac{j^{rs} \ell_{stu}(\theta_0) j^{tv} \ell_v(\theta_0) j^{ux} \ell_x(\theta_0)}{2!} + \\ &+ \frac{j^{rs} \ell_{stuv}(\theta_0)}{3!} \cdot \left(j^{tx} \ell_x(\theta_0) + \frac{j^{tx} \ell_{xyz}(\theta_0) j^{yb} \ell_b(\theta_0) j^{za} \ell_a(\theta_0)}{2!} \right) \\ &\cdot \left(j^{ux} \ell_x(\theta_0) + \frac{j^{ux} \ell_{xyz}(\theta_0) j^{yb} \ell_b(\theta_0) j^{za} \ell_a(\theta_0)}{2!} \right) \\ &\cdot \left(j^{vx} \ell_x(\theta_0) + \frac{j^{vx} \ell_{xyz}(\theta_0) j^{yb} \ell_b(\theta_0) j^{za} \ell_a(\theta_0)}{2!} \right) \end{aligned} \quad (4.10)$$

Este processo pode ser repetido *ad infinitum* gerando expansões de qualquer ordem desejada, e estaremos interessados no caso onde $\theta = \hat{\theta}$, o que justifica a equação (4.5).

4.1.2 Expansões Assintóticas Importantes: $(\ell(\hat{\theta}) - \ell(\theta))$

Define-se a função f como $f(\theta) = \ell(\theta) - \ell(\theta_0)$ e expandiremos esta função em série de Taylor ao redor do ponto θ_0 . Por praticidade, chamaremos $\ell_R(\theta)$ de ℓ_R e $(\theta - \theta_0)$ de $\Delta\theta$. O resultado é

$$f(\theta) = \ell_r \Delta\theta^r + \frac{\ell_{rs} \Delta\theta^r \Delta\theta^s}{2!} + \frac{\ell_{rst} \Delta\theta^r \Delta\theta^s \Delta\theta^t}{3!} + \dots \quad (4.11)$$

na qual, se o ponto de interesse é $\theta = \hat{\theta}$, os termos $\Delta\theta^k$ são dados pelas expansões da seção anterior.

4.1.3 Expansões Assintóticas Importantes: r_k

Por definição,

$$\begin{aligned} r_k &= \frac{\theta_k - b_k}{|\theta_k - b_k|} \sqrt{w(\chi = b_k)} \\ &= \frac{\theta_k - b_k}{|\theta_k - b_k|} \sqrt{2\{\ell(\phi, \chi) - \ell(\phi, \hat{\chi}_0)\}} \\ \Rightarrow r_k^2 &= 2\{\ell(\phi, \chi) - \ell(\phi, \hat{\chi}_0)\} \end{aligned} \quad (4.12)$$

Então, pode-se relacionar estas estatísticas com as expansões previamente estabelecidas, com as restrições

$$\begin{aligned} \Delta\theta^j &= 0 \quad \forall j < k \\ \ell_R &= 0 \quad \text{se } \exists j \in R \text{ talque } j < k \end{aligned}$$

Deixando a convenção de soma de lado neste trecho, a expansão (4.11) implica que

$$r_k^2 = \sum_{r=k+1}^p \ell_r \Delta\theta^r + \sum_{r=k+1}^p \sum_{s=k+1}^p \frac{\ell_{rs} \Delta\theta^r \Delta\theta^s}{2!} +$$

$$+ \sum_{r=k+1}^p \sum_{s=k+1}^p \sum_{t=k+1}^p \frac{\ell_{rst} \Delta\theta^r \Delta\theta^s \Delta\theta^t}{3!} + \dots \quad (4.13)$$

Combinando isto com a expansão de primeira ordem para $\Delta\theta^r$ dada por (4.7), obtém-se

$$\begin{aligned} r_k^2 &= \sum_{r=k+1}^p \ell_r \left(\sum_{s=k+1}^p j^{rs} \ell_s \right) \\ &+ \sum_{r=k+1}^p \sum_{s=k+1}^p \frac{\ell_{rs} \left(\sum_{t=k+1}^p j^{rt} \ell_t \right) \left(\sum_{u=k+1}^p j^{su} \ell_u \right)}{2!} \\ &+ \sum_{r=k+1}^p \sum_{s=k+1}^p \sum_{t=k+1}^p \frac{\ell_{rst} \left(\sum_{x=k+1}^p j^{rx} \ell_x \right) \left(\sum_{y=k+1}^p j^{sy} \ell_y \right) \left(\sum_{z=k+1}^p j^{rz} \ell_z \right)}{3!} \\ &+ \dots \end{aligned} \quad (4.14)$$

4.2 Família Exponencial

Diz-se que um vetor X de variáveis aleatórias X_i pertence a família exponencial se sua função de densidade pode ser escrita na forma

$$f_X(x|\theta) = \exp \left\{ \frac{1}{A(\phi)} (S(x)^T \theta - B(\theta)) + C(x, \phi) \right\} \quad (4.15)$$

com $A(\phi) > 0$, θ vetor de dimensão p e $S(x)$ é um vetor de estatísticas suficientes para θ . Neste caso, a função de verossimilhança dos x_i tem a forma da densidade de X , e a log-verossimilhança é dada por

$$\ell(\theta) = \frac{1}{A(\phi)} (S(x)\theta - B(\theta)) + C(x, \phi)$$

de onde é dedutível que as derivadas de $\ell(\theta)$ são dadas por

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_j} &= \frac{1}{A(\phi)} \left(S(x)_j - \frac{\partial B}{\partial \theta_j} \right) \\ \frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} &= -\frac{1}{A(\phi)} \frac{\partial^2 B}{\partial \theta_j \partial \theta_k} \\ \frac{\partial^3 \ell}{\partial \theta_j \partial \theta_k \partial \theta_l} &= -\frac{1}{A(\phi)} \frac{\partial^3 B}{\partial \theta_j \partial \theta_k \partial \theta_l} \\ &\dots \end{aligned}$$

em que $S(x)_j$ é a j -ésima estatística suficiente. Tem-se, então, que as derivadas parciais de ordem dois ou maior não dependem dos dados, fazendo com que $E[\ell_R] = v_R = \frac{1}{A(\phi)} \left(S(x)_j - \frac{\partial B}{\partial \theta_j} \right)$ caso $R = \{j\}$ e $E[\ell_R] = v_R = \ell_R$ caso contrário. Então, tem-se que $H_R = \ell_R - v_R = \frac{1}{A(\phi)} E[S(x)_r]$ caso $R = \{j\}$ e identicamente nulo caso contrário. Tem-se, também, que, a partir do resultado bem estabelecido $E[\ell_r] = 0$, que $E[\ell_r] = \frac{1}{A(\phi)} \left(E[S(Y)_r] - \frac{\partial B}{\partial \theta_r} \right)$ implica que $E[S(Y)_r] = \frac{\partial B}{\partial \theta_r}$. Mais um resultado desta família é que o estimador de máxima verossimilhança tem forma fechada se e somente se a função $\frac{\partial B}{\partial \theta_r}$ é invertível, pois diferenciando $\ell(\theta)$ e igualando à zero tem-se

$$\begin{aligned} \frac{1}{A(\phi)} \left(S(Y)_r - \frac{\partial B(\theta)}{\partial \theta_r} \right) &= 0 \\ S(Y)_r &= \frac{\partial B(\theta)}{\partial \theta_r} \end{aligned}$$

e a solução deste sistema tem forma fechada quando o segundo termo de (4.16) é invertível.

4.3 Estatísticas s_k e r_k na Família Exponencial

4.3.1 Família Exponencial Uniparamétrica

Na equação (4.15), se considerarmos $p = 1$ e escrevermos $L(\theta)$ na forma (1.8) tem-se θ escalar, apenas uma estatística suficiente $S(x)$ e

$$J = \left[\frac{\partial^2 \ell}{\partial \theta^2} \right] = \ell_2 = -n \frac{d(B(\theta))}{d\theta} \quad (4.16)$$

$$I^{-\frac{1}{2}} = \frac{1}{\sqrt{-n \frac{dB(\theta)}{d\theta}}} \quad (4.17)$$

Ainda devido à θ ser um escalar, tem-se que $r = 1$, $i^{rr} = ?$ inversa ou normal?, que

$$v_R = E \left[\frac{d^q \ell}{d\theta^q} \right] = \frac{d^q \ell}{d\theta^q} = \ell_R \quad \forall q \geq 2 \quad (4.18)$$

implicando que $H_R = 0$, o que reduz

4.4 Viés de Primeira Ordem dos Estimadores de Máxima Verossimilhança

Seja $v_{ijk} = E \left[\frac{\partial^3 \ell}{\partial \theta_i \partial \theta_j \partial \theta_k} \right]$ e $v_{ijk} = E \left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \frac{\partial \ell}{\partial \theta_k} \right]$ e i^{rs} o termo da r -ésima linha e s -ésima coluna da matriz de formação esperada. De acordo com [1], página 150, o viés do estimador de máxima verossimilhança, sob condições de regularidade, é dado por

$$B(\theta_r) = \frac{1}{2} i^{rs} i^{tu} (v_{stu} + 2v_{st,u}) + O \left(n^{-\frac{3}{2}} \right) \quad (4.19)$$

Reescrevendo sem notação tensorial, temos

$$B(\theta_r) = \frac{1}{2} \sum_{s=1}^p \sum_{t=1}^p \sum_{u=1}^p i^{rs} i^{tu} (v_{stu} + 2v_{st,u}) + O \left(n^{-\frac{3}{2}} \right) \quad (4.20)$$

No caso biparamétrico, quando temos apenas θ_1 e θ_2 , isso se reduz a duas equações:

$$\begin{aligned} 2B(\theta_1) &= i^{11} i^{11} (v_{111} + 2v_{11,1}) + i^{11} i^{12} (v_{112} + 2v_{11,2}) \\ &+ i^{11} i^{21} (v_{121} + 2v_{12,1}) + i^{11} i^{22} (v_{122} + 2v_{12,2}) \\ &+ i^{12} i^{11} (v_{211} + 2v_{21,1}) + i^{12} i^{12} (v_{212} + 2v_{21,2}) \\ &+ i^{12} i^{21} (v_{221} + 2v_{22,1}) + i^{12} i^{22} (v_{222} + 2v_{22,2}) \\ &+ O \left(n^{-\frac{3}{2}} \right) \end{aligned} \quad (4.21)$$

$$\begin{aligned} 2B(\theta_2) &= i^{21} i^{11} (v_{111} + 2v_{11,1}) + i^{21} i^{12} (v_{112} + 2v_{11,2}) \\ &+ i^{21} i^{21} (v_{121} + 2v_{12,1}) + i^{21} i^{22} (v_{122} + 2v_{12,2}) \\ &+ i^{22} i^{11} (v_{211} + 2v_{21,1}) + i^{22} i^{12} (v_{212} + 2v_{21,2}) \\ &+ i^{22} i^{21} (v_{221} + 2v_{22,1}) + i^{22} i^{22} (v_{222} + 2v_{22,2}) \\ &+ O \left(n^{-\frac{3}{2}} \right) \end{aligned} \quad (4.22)$$

Se os parâmetros forem globalmente ortogonais, temos que $i^{12} = i^{21} = 0$, reduzindo à

$$2B(\theta_1) = i^{11} i^{11} (v_{111} + 2v_{11,1}) + i^{11} i^{22} (v_{122} + 2v_{12,2}) + O \left(n^{-\frac{3}{2}} \right) \quad (4.23)$$

$$2B(\theta_2) = i^{22} i^{11} (v_{211} + 2v_{21,1}) + i^{22} i^{22} (v_{222} + 2v_{22,2}) + O \left(n^{-\frac{3}{2}} \right) \quad (4.24)$$

5 *Conclusões*

No presente trabalho duas contribuições foram feitas:

- Foi desenvolvida a fórmula para viés do estimador de máxima verossimilhança, incluindo expansões utilizadas no cálculo deste viés (Capítulo 4), dos parâmetros da distribuição COM-Poisson, que tem grande potencial para uso na Biometria, vide [15] (Seção 4).
- Foi computada a perda de poder ao errar na especificação de uma distribuição, quando uma distribuição desejada é uma estendida pela primeira alternativa de Lehmann é aproximada por uma distribuição clássica.

Referências

- [1] BARNDORFF-NIELSEN, O.E. E COX, D. (1994) Inference and Asymptotics. Chapman and Hall, London.
- [2] CONWAY, R. W. E MAXWELL, W. L. (1961). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, Volume **12**, páginas 132-136.
- [3] CORDEIRO, G. M. E STOŠIĆ, B. D. (2007). Comunicação pessoal.
- [4] EDWARDS, A.W.F. E FRACCARO, M. (1960). Distribution and sequences of sexes in a selected sample of Swedish families, *Annals of Human Genetics*, Volume **24**, páginas 245-252.
- [5] EGUCHI, S. E COPAS, J. (2006). Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, vol. **97**, Issue 9, pages 2034-2040.
- [6] FISHER, R. A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *Journal of Agricultural Science*, Volume **11**, páginas 107-135.
- [7] FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society London*, Series A, Volume **222**, páginas 309-368.
- [8] FISHER, R. A. E MACKENZIE, W.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *Journal of Agricultural Science*, Volume **13**, páginas 311-320.
- [9] FISHER, R.A. (1924). Studies in crop variation: III. The influence of rainfall on the yield of wheat at Rothamsted. *Philosophical Transactions of the Royal Society London*, Series B, Volume **213**, páginas 89-142.
- [10] FISHER, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, Volume **22**, páginas 700-725.
- [11] FISHER, R. A. E ÉDEN, T. (1927). Studies in crop variation. IV. The experimental determination of the value of top dressings with cereals. *Journal of Agricultural Science*, Volume **17**, páginas 548-562.
- [12] FISHER, R. A. E ÉDEN, T. (1929). Studies in crop variation. VI. Experiments on the response of the potato to potash and nitrogen. *Journal of Agricultural Science*, Volume **19**, páginas 201-213.

- [13] GUPTA, R. C., GUPTA, P. L. E GUPTA, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communication in Statistics: Theory and Methods*, vol. **27**, pages 887-904.
- [14] GUPTA, R. D. E KUNDU, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, Volume **41**, páginas 173 - 188.
- [15] KADANE, J.B., SHMUELI, G., MINKA, T.P., BORLE, S. E BOATWRIGHT, P. (2006). Conjugate Analysis of the Conway-Maxwell-Poisson Distribution. *Bayesian Analysis*, Volume **1**, Número 2, páginas 363-374.
- [16] KULLBACK, S. E LEIBLER, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, vol. **22**, Number 1, pages 79-86.
- [17] LORD, D., GUIKEMA, S. E GEEDIPALLY, S.R. (2008). Application of the Conway-Maxwell-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes, manuscrito não publicado.
- [18] MCCULLAGH, P. (1987). Tensor Methods in Statistics. Chapman and Hall.
- [19] NADARAJAH, S., KOTZ, S. (2006). The Exponentiated Type Distributions. *Acta Applicandae Mathematicae*, vol. **92**, pages 97-111.
- [20] NADARAJAH, S. 2006. The exponentiated Gumbel distribution with climate application. *Environmetrics*, vol. **17**, Number 1, pages 13-23.
- [21] NEYMAN, J. E PEARSON, E. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London Series A - Papers of a Mathematical or Physical Character*, Volume **231**, páginas 289-337.
- [22] R DEVELOPMENT CORE TEAM, 2004. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3. Disponível em: <http://www.r-project.org>
- [23] SHMUELI, G., MINKA, T.P., KADANE, J.B., BORLE, S. E BOATWRIGHT, P. (2005). Computing with the COM-Poisson distribution.
- [24] SHMUELI, G., MINKA, T.P., KADANE, J.B., BORLE, S. E BOATWRIGHT, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution, *Journal of the Royal Statistical Society Series C*, Volume **54**, Número 1 , páginas 127-144.
- [25] SKOVGAARD, I. M. (2001). Likelihood asymptotics, *Scandinavian Journal of Statistics*, Volume **28**, páginas 3-32.
- [26] SOARES, L. G. M. E DE ARAÚJO, L. H. G. D. (2007). "Estimação por Mínimos Quadrados dos Parâmetros da Distribuição de Conway-Maxwell-Poisson", *Anais da 10ª Escola de Modelos de Regressão*.
- [27] THIELE, T. N. (1889). Almindelig Iagttagelseslcere: Sandsynlighedsregning og mindste Kvadraters Methode. Reitzel, Kobenhavn.

- [28] VERHULST, P. F. (1838), "Notice sur la loi la population suit dans son accroissement", *Correspondence mathématique et physique*, publiée L. A. J. QUETELET, Volume **10**, páginas 113 - 121.

APÊNDICE A – Detalhes da Distribuição COM-Poisson

A.1 Derivadas da Função Z

Além da ortogonalidade dos parâmetros, as derivadas são necessárias para mostrar a ortogonalidade dos parâmetros. As derivadas mais simples tem a forma

$$\begin{aligned}
 Z_\lambda &= \sum_{j=0}^{\infty} \frac{j\lambda^{j-1}}{(j!)^\nu} \\
 Z_\nu &= -\nu \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu+1}} \\
 Z_{\lambda\lambda} &= \sum_{j=0}^{\infty} \frac{j(j-1)\lambda^{j-1}}{(j!)^\nu} \\
 Z_{\lambda\nu} &= -\nu \sum_{j=0}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu+1}} \\
 Z_{\nu\lambda} &= -\nu \sum_{j=0}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu+1}} \\
 Z_{\nu\nu} &= (\nu^2 + \nu) \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu+2}}
 \end{aligned}$$

Tomando $j(j-1)\dots(j-k) = \prod_{i=0}^{k-1} (j-i)$ e $(-\nu)(-\nu-1)\dots(-\nu-k) = \prod_{i=0}^{k-1} (-\nu-i)$, temos dois termos gerais

$$\begin{aligned}
 Z_{\lambda\dots(\text{derivando } k \text{ vezes em } \lambda)\dots\lambda} &= \sum_{j=k}^{\infty} \frac{\lambda^{j-k} \prod_{i=0}^{k-1} (j-i)}{(j!)^\nu} \\
 Z_{\nu\dots(\text{derivando } k \text{ vezes em } \nu)\dots\nu} &= \left(\prod_{i=0}^{k-1} [-\nu-i] \right) \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu+k}}
 \end{aligned}$$

$$= \left([-1]^k \prod_{i=0}^{k-1} [v+i] \right) \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu+k}}$$

Agora temos as derivadas necessárias para mostrar que a diagonal da matriz hessiana (equação 2.10) é nula. Temos

$$\begin{aligned} J[2, 1] &= -\frac{nZ_{\nu\lambda}}{Z} + \frac{nZ_{\nu}Z_{\lambda}}{Z^2} \\ &= n \left(-\frac{Z_{\nu\lambda}Z}{Z^2} + \frac{Z_{\nu}Z_{\lambda}}{Z^2} \right) \\ &= \frac{n\nu}{Z^2} \left(\left[-\sum_{j=0}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu+1}} \right] \left[\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu}} \right] + \left[\sum_{j=0}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu+1}} \right] \left[\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu}} \right] \right) \\ &= 0 \end{aligned}$$

A.2 Equações do Viés de Primeira Ordem da CMP

$$B(\theta_\lambda) = - \frac{\lambda Z(-2E[X] + \lambda^3 Z^2 Z_\lambda Z_{\lambda\lambda} - 3\lambda Z^3 Z_\lambda^3 - 2n\lambda^3 Z^3_\lambda - 2n\lambda^3 Z^6 E[X^2] - 2n\lambda^4 Z^5 E[X] Z_\lambda + 2n\lambda^5 Z^3 E[X])}{2n(\lambda^2 Z^2_\lambda - E[X] Z^2 - \lambda Z Z_{\lambda\lambda})^2} \quad (\text{A.1})$$

$$B(\theta_\nu) = \frac{Z(-Z_\nu Z^2 Z_{\nu\nu} + 3Z Z_\nu Z_{\nu\nu} - 2Z^3_\nu + nE[\ln X] Z^2 Z_{\nu\nu} - nE[\ln X] Z Z^2_\nu + nZ Z_\nu Z_{\nu\nu} - nZ^3_\nu)}{2n(-Z^2_\nu + Z Z_{\nu\nu})^2} \quad (\text{A.2})$$

A.3 Código em Linguagem R para gerar números aleatórios com distribuição CMP

```

# Antes de declarar o gerador de números aleatórios (a função rcmp)
# se faz necessário declarar a função Z e a função de distribuição
# COM-Poisson.
cmp.Z <- function (lambda, nu, prec=150) {
  soma<-1
  for (i in 1:prec) {
    tmp<-(lambda^i)/(factorial(i)^nu)
    if (!is.nan(tmp)) {
      soma<-soma+tmp
    } else {
      break
    }
  }
  return(soma)
}

# A função acima tem precisão variável, se a precisão não for informada
# ela usa 150 iterações. Ela vai parar quando chegar a 150 iterações ou
# quando o interpretador do R não conseguir mais calcular iterações.

# Esta função computa a função de distribuição f
cmp.f <- function (x, lambda, nu, prec=150) {
  tmp<-(lambda^x)/(factorial(x)^nu)
  return(tmp/cmp.Z(lambda,nu,prec))
}

# A mesma regra para precisão da função acima vale.

```

```

# A função a seguir segue o formato das funções geradoras
# de números aleatórios do pacote base;
# Para ver exemplos, usar o comando
?runif
# para ver a ajuda do R.

# O argumento n é número de números aleatórios a serem gerados.
# O código usa o bem conhecido método de aceitação-rejeição.
rcmp <- function (n, lambda=1, nu=1) {
  resp<-vector(mode="numeric",length=n)
  for (i in 1:n) {
    tmp<-cmp.f(0, lambda, nu)
    u<-runif(1)
    j<-0
    while (tmp<u) {
      j<-j+1
      tmp<-tmp+cmp.f(j,lambda,nu)
    }
    resp[i]<-j }
  return(resp)
}

# Como teste, gera uma amostra iid de tamanho 100
# da distribuição Poisson com média 10:
rcmp(100,lambda=10)

# Gera uma amostra iid de tamanho 100
# da distribuição CMP com ni=10 e lambda=10:
rcmp(100,lambda=10,nu=10)

# Fim do código.

```