Universidade Estadual de Maringá Departamento de Estatística





Programa de Pós-graduação em Bioestatística

SÉRGIO MARCUSSI GASPECHAK

INTERVALOS DE CONFIANÇA DE ORDEM SUPERIOR PARA A DISTRIBUIÇÃO EXPONENCIAL GENERALIZADA

Maringá - Paraná Setembro de 2015

SÉRGIO MARCUSSI GASPECHAK

INTERVALOS DE CONFIANÇA DE ORDEM SUPERIOR PARA A DISTRIBUIÇÃO EXPONENCIAL GENERALIZADA

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística da Universidade estadual de Maringá como requisito para obtenção do tíulo de mestre em Bioestatística.

Orientadora: Profa. Dra. Isolde Previdelli

Maringá - Paraná Setembro de 2015

SÉRGIO MARCUSSI GASPECHAK

INTERVALOS DE CONFIANÇA DE ORDEM SUPERIOR PARA A DISTRIBUIÇÃO EXPONENCIAL GENERALIZADA

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística da Universidade estadual de Maringá como requisito para obtenção do tíulo de mestre em Bioestatística.

Orientadora: Profa. Dra. Isolde Previdelli

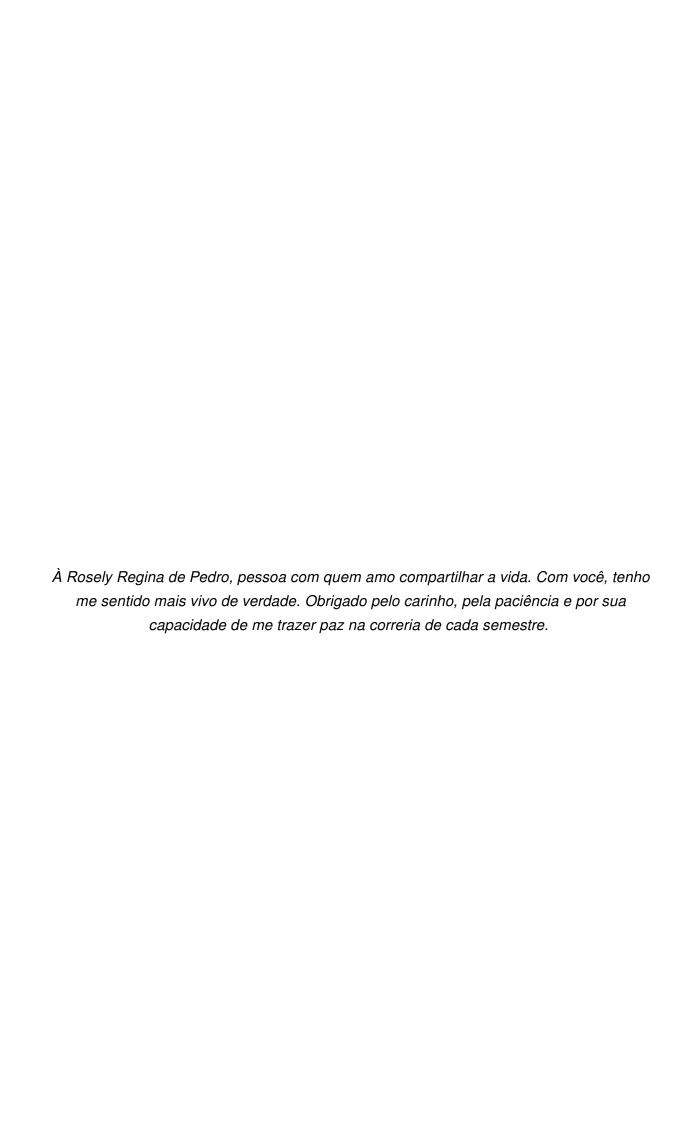
Trabalho aprovado. Maringá - Paraná, 11 de Setembro de 2015:

Prof^a. Dr^a. Isolde Previdelli-UEM
Orientadora

Prof. Dr. Artur Lemonte-UFPE

Prof^a. Dr^a. Rosangela G. Santana-UEM

Maringá - Paraná Setembro de 2015



Agradecimentos

A Deus, por ter me dado saúde e força para superar as dificuldades.

À minha orientadora Prof^a. Dr^a. Isolde Previdelli, pelo suporte no pouco tempo que lhe coube, pelas suas correções e incentivos.

Aos meus pais, pelo amor, incentivo e apoio incondicional.

À CAPES, pelo apoio financeiro.

E a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.



Resumo

Inferência estatística, geralmente é realizada com base em aproximações por distribuições assintóticas, uma vez que distribuições exatas nem sempre estão disponíveis. A inferência obtida por meio da estatística da Razão de verossimilhanças sinalizada, R, que possui distribuição assintótica normal padrão com erro de ordem $O(n^{-1/2})$, ditas, inferência de primeira ordem, e podem gerar resultados enviesados em pequenas amostras. Uma variante dessa estatística é a da Razão de verossimilhanças modificada, R^* , que possui distribuição assintótica normal padrão com erro de ordem $O(n^{-3/2})$, neste caso, de segunda ordem, pode ser uma alternativa em pequenas amostras. O objetivo desse trabalho é comparar as coberturas dos intervalos de confiança usando R e R^* , e a amplitude dos intervalos através de estudos de simulação sob vários cenários. Alem disso apresentamos uma aplicação a dados de psiquiatria para ilustrar que esta metodologia é uma alternativa nos casos de amostras pequenas a moderadas, o que são muito comuns em estudos relacionados a área da saúde.

Palavras-chave: Inferência. Log-verossimilhança. Teoria assintótica. Razão de verossimilhanças Modificada. Intervalos de Confiança.

Abstract

Abstract We usually make statistical inferences using approaches based on asymptotic distributions, since exact distributions can be unavailable. Inferences obtained from the sign-posted likelihood ratio statistics R, which has asymptotic standard normal distribution with error $O(n^{-1/2})$, called first order inferences, can produce biased results in the case of small samples. A variant of this statistics is the adjusted likelihood ratio statistics R*, which has asymptotic standard normal distribution with error $O(n^{-3/2})$, called in this case third order, can be an option in cases of small samples. This work aims at comparing the covers of the confidence intervals obtained from R with the ones obtained from R* and with the lengths of the intervals obtained via simulation studies in several scenarios, performing an application of it with psychiatric data and divulging an alternative methodology in cases from small to moderate samples, which are very common in studies related to health.

Keywords: Inference; Log-Likelihood; Asymptotic theory; Modified Log-Likelihood; Confidence intervals.

Lista de ilustrações

Figura 1 -	Comportamento da distribuição EG para diversos valores dos parâmetros.	24
Figura 2 -	Função taxa de falhas para $\lambda=2.5$ e $\delta=0.3;\ 0.8;\ 1;\ 2;\ 3.$	25
Figura 3 -	Distribuição das estatísticas de Teste para as metodologias em estudo	
	para $n=5;15;30,\delta=0.8$ e nível significância de $5\%.$	30
Figura 4 -	Distribuição das estatísticas de Teste para as metodologias em estudo	
	para $n=5;15;30,\delta=1.5$ e nível significância de $5\%.$	31
Figura 5 -	Comportamento dos Gastos com internações hospitalares	35

Lista de tabelas

Tabela 1 -	Cobertura dos intervalos de confiança e amplitude média dos intervalos com seus	
	coeficientes de variação para $\lambda=1$ e $\delta=0.8, 1.5, 2$ com nível significância de 5% e	
	1%	32
Tabela 2 -	Distribuição dos gastos com internações hospitalares (em milhões de reais) no	
	ano de 2002 em hospitais psiquiátricos no estado do Paraná	34
Tabela 3 -	Intervalos de confiança para δ : Sexo Feminino	35
Tabela 4 -	Intervalos de confiança para δ : Sexo Masculino	36

Sumário

1 INTRODUÇÃO							
	1.1	Introdu	ução	11			
2	Meto	odologi	ia	14			
	2.1	Metod	ologia	14			
		2.1.1	Transformação Normalizadora	16			
		2.1.2	Razão de Verossimilhança Sinalizada	18			
		2.1.3	Razão de Verossimilhança Sinalizada Modificada	18			
			2.1.3.1 Casos Uniparamétricos	19			
			2.1.3.2 Casos Multiparamétricos	20			
	2.2	Aproxi	mações para o ajuste de Barndorff-Nielsen	22			
		2.2.1	Aproximação de DiCiccio-Martin	22			
		2.2.2	Aproximações de Skovgaard e Severini (1999)	23			
		2.2.3	Aproximação de Fraser-Reid-Wu	23			
3	Dist	ribuiçã	o Exponencial Generalizada	24			
3.1 Distribuição Exponencial Generalizada							
		3.1.1	Funções de Risco e Sobrevivência	25			
		3.1.2	Quantis, Mediana e Moda	25			
		3.1.3	Função Geradora de Momentos	26			
		3.1.4	Média e Variância	26			
		3.1.5	Estimadores de Máxima verossimilhança	27			
		3.1.6	Intervalo de Confiança	27			
		3.1.7	Razão de verossimilhanças Modificada para EG	28			
4	Res	ultados	s das Simulações	29			
	4.1	Interva	alos de confiança para o parâmetro δ	29			
		4.1.1	Discussões	33			
5	Apli	cação		34			
	5.1	Aplica	ção	34			
		5.1.1	Considerações finais e estudos futuros	36			
Re	forôr	nciae		37			

INTRODUÇÃO

1.1 Introdução

Geralmente, inferência estatística é realizada com base em aproximações assintóticas, uma vez que distribuições exatas nem sempre estão disponíveis, funcionando bem em grandes amostras. Para amostras pequenas e moderadas, a suposição de convergência assintótica pode gerar resultados enviesados e a inferência obtida pode conduzir a tomadas de decisões com erros muito acima do aceitável.

A estatística da razão de verossimilhanças, $w(\theta)=2[\ell(\widehat{\theta})-\ell(\theta)]$, é uma estatística muito utilizada, mesmo quando não é claro a convergência assintótica em distribuição, neste caso χ_q^2 , em que q é a dimensão do vetor de parâmetros. Uma outra quantidade relacionada utilizada é dada por $R(\theta)=sign(\widehat{\theta}-\theta)\{2[\ell(\widehat{\theta})-\ell(\theta)]\}^{1/2}$, conhecida por razão de verossimilhanças sinalizada e que tem distribuição assintoticamente normal padrão com erro de ordem $O(n^{-1/2})$, que no caso de amostras pequenas a moderadas essa convergência se torna muito pobre, gerando resultados enviesados.

Vários autores trabalharam para fazer correções nessas estatísticas de teste. As correções foram originalmente propostas por Bartlett (1937), que fez uma transformação escalar aplicada à estatística da razão de verossimilhanças em modelos paramétricos. Mas a contribuição mais importante de Bartlett (1953) para a teoria estatística, ao introduzir uma fórmula para o viés de $O(n^{-1})$ em estimadores de máxima verossimilhança no caso iid (observações independentes e identicamente distribuídas) para os modelos uniparamétricos. Esse resultado foi estendido posteriormente por Cox e Snell (1968) para modelos multiparamétricos, cujos dados não são necessariamente identicamente distribuídos.

Nas últimas décadas tem sido desenvolvidos diversos métodos que visam a aproximações mais precisas com base nas aproximações ponto de sela. Dentre as principais referências temos os trabalhos de Daniels (1954), Barndorff-Nielsen e Cox (1979), Barndorff-

Nielsen e Cox (1984), Barndorff-Nielsen e Cox (1989), Cox e Barndorff-Nielsen (1994), Daniels (1987), Barndorff-Nielsen (1990), Barndorff-Nielsen (1991) e Reid (1988), Jensen (1992), Kolassa (1994), entre outros.

Em termos estatísticos essas aproximações mais precisas foram possíveis devido a um desenvolvimento das ideias de Fisher (1925), que faz uso de estatística ancilares que tem por objetivo complementar as informações sobre o parâmetro quando a estimativa de máxima verossimilhança não é suficiente, abordagem que teve reinicio com os trabalhos de Efron e Hinkley (1978) e, em particular, Barndorff-Nielsen (1980), Barndorff-Nielsen (1983), Barndorff-Nielsen (1986), Barndorff-Nielsen (1991) com uma equação para um p^* e uma estatística $R^*(\theta) = R(\theta) + R(\theta)^{-1} \log{(U(\theta)/R(\theta))}$, que depende de $R(\theta)$ e de uma quantidade $R^*(\theta)$ 0 obtida através de derivadas em relação ao espaço amostral que envolvem uma estatística ancilar $\mathbf{\tau}$ 1, sendo que R^* 2, possui distribuição assintótica normal padrão com erro de ordem $R^*(\theta)$ 2, conhecida como razão de verossimilhança modificada, que tem por objetivo melhorar a aproximação assintótica pela normal padrão.

Dentre muitos trabalhos que foram realizados para verificar as vantagens da razão de verossimilhanças modificada com relação a outras metodologias utilizadas para realizar inferências sobre parâmetros, pode-se citar Reid (1996), Brazzale, Davison e Reid (2007), que aborda as metodologias de alta ordem para obtenção de inferência no caso de pequenas amostras, e Brazzale e Davison (2008), que apresenta em varias situações a aplicação desta metodologia. Podemos citar também o artigo de Cortese e Ventura (2009), que aplica a teoria assintótica de ordem superior em modelo estresse-força, muito utilizado em engenharia e áreas médicas devido a sua ligação com a curva Roc, utilizada para testes de diagnósticos e também evidenciar a superioridade da metodologia em questão para pequenas amostras. Os pesquisadores Lozada-Can e Davison (2010), ilustram os cálculos utilizando um ensaio de diluição, uma regressão de modelo Poisson-inflado e uma série temporal, em que fica clara a superioridade dos resultados obtidos através das estatísticas modificadas. Lemonte (2010) em sua tese de doutorado apresenta vários resultados envolvendo as estatísticas modificadas assim, como diversas aproximações para determinação da quantidade U sem que haja a necessidade da especificação de uma estatística ancilar, que em algumas distribuições é de difícil obtenção ou até mesmo impossível, o que impossibilitaria o uso de R^* , sendo a principal referência deste trabalho.

As inferências podem ser pontuais ou intervalares, geralmente, intervalos de confiança são muito utilizados, pois é possível obter um grau de confiança das inferências obtidas, o que não ocorre nas estimativas pontuais. Existem várias metodologias que possibilitam encontrar intervalos de confiança (IC) para θ , como descrito em Brazzale, Davison e Reid (2007).

O procedimento geral para construção de intervalos de confiança para um parâmetro θ consiste nos seguintes passos:

- i) Obter uma estatística que também depende de θ , digamos $G=G(\boldsymbol{X},\theta)$, mas cuja distribuição não depende de θ ;
- *ii*) Através da distribuição de G encontrar a e b tais que $P(a < G < b) \ge 1 \alpha$;
- *iii*) Define $\{\theta : a < G(x, \theta) < b\}$ como intervalo (ou região) de confiança $100(1 \alpha)\%$.

A variável aleatória G, comumente denominada quantidade pivotal ou pivot, é fundamental para o funcionamento do método, e deve depender da amostra por meio de estatísticas suficientes minimais e ter distribuição conhecida.

Geralmente, quando se adota uma metodologia para construção dos intervalos de confiança, espera-se que ela forneça intervalos com uma cobertura próxima do valor nominal pré definido, porém metodologias que fornecem uma cobertura maior nem sempre são melhores, pois uma forma fácil de se obter uma cobertura maior é aumentando a amplitude do intervalo, mas essa amplitude é usada para indicar a confiabilidade de uma estimativa, ou seja, um IC pode ser usado para descrever o quanto os resultados de uma pesquisa são confiáveis. Portanto, metodologias que resultem num IC menor, é mais confiável do que em uma que resulte num IC maior. Assim é desejável que tenhamos metodologias que sejam capazes de fornecerem uma maior cobertura dos intervalos, mas com amplitudes menores.

Diante disso, esta pesquisa tem como principal objetivo obter a estatística modificada R^* para a distribuição exponencial generalizada, bem como, construir intervalos de confiança para comparar com os obtidos a partir das estatísticas de Wald e R.

No caso das metodologias em estudo, R e R^* , ambas convergem assintoticamente a uma normal padrão, mas R com erro de ordem $O(n^{-1/2})$ e R^* com erro de ordem $O(n^{-3/2})$, assim, espera-se que, para inferências obtidas de R^* , tenha-se resultados melhores; isto é, maior cobertura e também uma redução na amplitude dos intervalos, pois isto levará a ganhos nas inferências estatísticas, principalmente na área da saúde, uma vez que é comum pesquisas envolvendo pequenas amostras.

Este estudo está dividido como segue: No capítulo 2 apresentamos a metodologia contendo alguns aspectos teóricos importantes para desenvolver a pesquisa. No capítulo 3, apresentamos distribuição exponencial generalizada com alguns resultados teóricos. No capítulo 4, apresentamos os resultados das simulações de Monte Carlo para as estatísticas de Wald, R e R^* , em que foram comparadas as taxas de cobertura e amplitude dos intervalos de confiança para a distribuição exponencial generalizada e por fim no capítulo 5, tem-se uma aplicação com os gastos de internações em hospitais psiquiátricos do estado do Paraná no ano de 2002/ DATASUS.

Metodologia

2.1 Metodologia

Em uma situação mais simples, considere uma amostra independente e igualmente distribuída $(y_1,\ y_2,\dots,y_n)$, com função densidade de probabilidade $f(y;\pmb{\theta})$ em que $\pmb{\theta}$ é um parâmetro escalar desconhecido, a função de log-verossimilhança é dada por $\ell(\pmb{\theta}) = \sum\limits_{i=1}^n log(f(y_i;\pmb{\theta})), \ \hat{\pmb{\theta}}$ é o estimador de máxima verossimilhança de $\pmb{\theta}$ e informação observada expressa por $j(\pmb{\theta}) = -\frac{\partial^2 \ell(\pmb{\theta})}{\partial \pmb{\theta}^2}$.

As inferências para θ são, usualmente obtidas, através das estatísticas de (BRAZZALE; DAVISON; REID, 2007):

Razão de verossimilhanças:
$$R(\boldsymbol{\theta}) = sign(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \left\{ 2 \left(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}) \right) \right\}^{1/2} \sim N(0,1)$$
 (2.1)

Escore de Rao
$$s(\theta)=j(\widehat{\theta})^{-1/2}\partial\ell/\partial\theta\sim N(0,1);$$
 (2.2)

Estatística de Wald:
$$t(\theta) = j(\widehat{\theta})^{1/2} \left(\widehat{\theta} - \theta\right) \sim N(0, 1);$$
 (2.3)

Uma importante variação da razão de verossimilhanças é conhecida por razão de verossimilhanças modificada, que é definida da forma

$$R^*(\boldsymbol{\theta}) = R(\boldsymbol{\theta}) + \frac{1}{R(\boldsymbol{\theta})} \log \left(\frac{U(\boldsymbol{\theta})}{R(\boldsymbol{\theta})} \right) \sim N(0, 1).$$
 (2.4)

Se θ for um vetor de parâmetros, pode-se reescrevê-lo como $\theta=(\psi, \lambda)$, tal que ψ seja o parâmetro de interesse, e λ é um vetor formado pelos demais parâmetros, neste caso, tratados como parâmetros de perturbação ou incômodo, o que torna mais simples os cálculos. Caso haja mais de um parâmetro de interesse, geralmente, estes são estudados

um de cada vez. Deste modo tem-se $\hat{\boldsymbol{\theta}}=(\hat{\psi}, \hat{\boldsymbol{\lambda}})$. Fixando ψ , o estimador de máxima verossimilhança restrita será $\hat{\boldsymbol{\theta}}_{\psi}=(\psi, \hat{\boldsymbol{\lambda}}_{\psi})$, e a log-verossimilhança a ser maximizada para inferir sobre ψ se torna $\ell(\psi, \hat{\boldsymbol{\lambda}}_{\psi})=\ell_p(\psi)$. Consequentemente, tem-se

$$R(\psi) = sign(\widehat{\psi} - \psi) \left\{ 2 \left(\ell_p(\widehat{\psi}) - \ell_p(\psi) \right) \right\}^{1/2}, \tag{2.5}$$

e, portanto,

$$R^*(\psi) = \frac{1}{R(\psi)} + \log\left(\frac{U(\psi)}{R(\psi)}\right),\tag{2.6}$$

veja, Brazzale, Davison e Reid (2007).

A quantidade $R^*(\psi)$ fornece intervalos de confiança mais precisos em comparação com os fornecidos por $R(\psi)$. Para respostas contínuas, os intervalos de confiança obtidos com R^* tem erro de cobertura é de ordem $O(n^{-3/2})$, enquanto que para R o erro de cobertura de ordem $O(n^{-1/2})$ (LOZADA-CAN; DAVISON, 2010). A estatística $U(\psi)$, quando ψ for um escalar, pode ser escrita como

$$U(\psi) = \frac{\left| \varphi(\widehat{\boldsymbol{\theta}}) - \varphi(\widehat{\boldsymbol{\theta}}_{\psi}) - \varphi_{\lambda}(\widehat{\boldsymbol{\theta}}_{\psi}) \right|}{\left| \varphi_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) \right|} \frac{\left| j(\widehat{\boldsymbol{\theta}}) \right|^{1/2}}{\left| j_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_{\psi}) \right|^{1/2}}, \tag{2.7}$$

Em que φ_{θ} é a matriz da derivadas parciais $\partial \varphi / \partial \theta^{T}$. A expressão (2.7) é invariante para transformação φ , e pode simplificar muito os cálculos analíticos. O numerador do primeiro fator de U é o determinante da matriz $d \times d$ cuja primeira coluna é $(\varphi(\widehat{\boldsymbol{\theta}}) - \varphi(\widehat{\boldsymbol{\theta}}_{\psi}))^{T}$, o restante das colunas são $\varphi_{\lambda}(\widehat{\boldsymbol{\theta}}_{\psi})$.

Para amostras de observações independentes, φ é definido como

$$\varphi(\boldsymbol{\theta})^T = \sum_{i=1}^n \frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial y_i} \bigg|_{y=v^0} V_i,$$
(2.8)

em que y^0 representa os dados observados, e para respostas contínuas, tem-se

$$V_k = \left. \frac{\partial y_k}{\partial \boldsymbol{\theta}} \right|_{(y^0, \widehat{\boldsymbol{\theta}}^0)} = \left. - \left(\frac{\partial z_k}{\partial y_k} \right)^{-1} \left(\frac{\partial z_k}{\partial \theta} \right) \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^0 \ \boldsymbol{y} = \boldsymbol{y}^0}, \tag{2.9}$$

com $z=~\{z_1(y_1,m{ heta}),~z_2(y_2,m{ heta}),~\dots,z_n(y_n,m{ heta})\}$ em casos contínuos.

Fórmulas mais simples podem ser encontradas em casos específicos, como nos modelos discretos da família exponencial (DAVISON; FRASER DONALD AS E REID, 2006) modelos em que tem-se

$$V_k = \left. \frac{\partial E(y_k; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\widehat{\boldsymbol{\theta}}^0}.$$

Se θ é um escalar, então θ = ψ e λ não aparece na expressão. Assim, (2.7) torna-se

$$U(\psi) = \frac{\varphi(\widehat{\theta}) - \varphi(\widehat{\theta}_{\psi})}{\left|\varphi_{\theta}(\widehat{\theta})\right|} j(\theta)^{1/2}.$$
 (2.10)

Outra expressão apresentada por Brazzale, Davison e Reid (2007) para a quantidade U é

$$U(\psi) = \left(\ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) - \ell_{\boldsymbol{\lambda};\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) \cdot \ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\lambda}}}^{-1}(\widehat{\boldsymbol{\theta}}_{\psi}) \cdot \left\{\ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi})\right\}\right) \frac{\left|\ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi})\right|}{\sqrt{\left|j(\widehat{\boldsymbol{\theta}})\right| \cdot \left|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi})\right|}}, \quad (2.11)$$

em que

$$\ell_{;\widehat{\psi}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \widehat{\psi}}, \quad \ell_{;\widehat{\boldsymbol{\lambda}}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \widehat{\boldsymbol{\lambda}}}, \quad \ell_{\boldsymbol{\lambda};\widehat{\psi}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \widehat{\psi}}, \quad \ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\lambda}}} = \frac{\partial \ell^2(\boldsymbol{\theta})}{(\partial \boldsymbol{\lambda} \partial \widehat{\boldsymbol{\lambda}}^T)},$$

Alem disso,

$$j(\boldsymbol{\theta}) = -\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = -\frac{\partial \ell^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

e $j_{\psi\psi}(\psi, \lambda)$, $j_{\psi\lambda}(\psi, \lambda)$, $j_{\lambda\psi}(\psi, \lambda)$, e $j_{\lambda\lambda}(\psi, \lambda)$ são submatrizes da matriz de informação. No caso de distribuições da família exponencial, pode-se considerar $U(\theta) = j(\theta)^{1/2}(\hat{\theta} - \theta)$.

A seguir apresenta-se as ideias básicas para a construção de R^* a partir de uma transformação que visa melhorar a normalidade assintótica e uma fórmula p^* introduzida por Barndorff-Nielsen (1983). Mais detalhes destes resultados podem ser encontrados em Lemonte (2010) e Severini (2000).

2.1.1 Transformação Normalizadora

Dada uma amostra com distribuição aproximadamente normal, uma transformação dos dados pode ser aplicada com o objetivo de melhorar a ordem de convergência para a distribuição normal (LEMONTE, 2010). Considere uma variável aleatória X contínua de ordem $\mathcal{O}_p(1)$ com densidade da forma

$$X \sim \overline{c} \ h\left(\frac{x}{\sqrt{n}}\right) \phi(x),$$
 (2.12)

em que h tem ordem O(1), e h(0)=1, $\phi(.)$ denota a densidade de probabilidade da distribuição normal padrão, e \overline{c} é uma constante tal que $\overline{c}=1+O(n^{-1})$.

Tem-se que $h(x/\sqrt{n})=1+O(n^{-1/2})$ e, assim, X tem uma distribuição normal com erro de ordem $O(n^{-1/2})$. Uma aproximação para a função de distribuição de X é

$$\overline{c} \int_{-\infty}^{t} h(x/\sqrt{n})\phi(x)dx = 1 - \overline{c} \int_{t/\sqrt{n}}^{\infty} h(s)\sqrt{n}\phi(s\sqrt{n})ds,$$

que pode ser aproximada, segundo Severini (2000), por

$$F_n(t) = \Phi(t) - \phi(t) \frac{h(t/\sqrt{n}) - 1}{t},$$

com erro de ordem $O(n^{-3/2})$. Logo, a variável aleatória transformada

$$F_n(X) = \Phi(X) - \phi(X) \frac{h(X/\sqrt{n}) - 1}{X}$$

é uniformemente distribuída com erro de ordem $O(n^{-3/2})$.

Para obter uma variável aleatória que é aproximadamente distribuída de acordo com uma normal padrão, é necessário considerar a transformação $\Phi^{-1}(F_n(X))$. Expandindo a função Φ^{-1} em torno do ponto t e avaliando em $t+c/\sqrt{n}$, tem-se

$$\Phi^{-1}(t+c/\sqrt{n}) = \Phi^{-1}(t) + \frac{c}{\phi(\Phi^{-1}(t))\sqrt{n}} + \frac{1}{2} \frac{\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))^2} \frac{c^2}{n} + O(n^{-3/2}).$$
 (2.13)

Considerando-se a expansão em série de Taylor do $\log(1+w)$ em torno do ponto w=0, com $w=-c\,\Phi^{-1}(t)/\left\{\phi\left(\Phi^{-1}(t)\right)\sqrt{n}\right\}$, pode-se escrever

$$\log\left(-c\,\Phi^{-1}(t)/\left\{\phi\left(\Phi^{-1}(t)\right)\sqrt{n}\right\}\right) = \frac{c^{-1}\Phi^{-1}(t)}{\phi\left(\Phi^{-1}(t)\right)\sqrt{n}} - \frac{\left(c\,\Phi^{-1}(t)\right)^2}{2\left(\phi\left(\Phi^{-1}(t)\right)\sqrt{n}\right)^2} + O(n^{-3/2})$$

assim tem-se

$$\Phi^{-1}(t+c/\sqrt{n}) = \Phi^{-1}(t) - \frac{1}{\Phi^{-1}(t)} \log \left[1 - \frac{\Phi^{-1}(t)c}{\phi(\Phi^{-1}(t))\sqrt{n}} \right] + O(n^{-3/2}).$$
 (2.14)

Substituindo $t=\Phi^{-1}(X)$ e $c=-\sqrt{n}\phi(X)\frac{h(x/\sqrt{n})-1}{X}$ pode-se escrever (2.14) como

$$\Phi^{-1}\left(\Phi(t) + \frac{c}{\sqrt{n}}\right) = \Phi^{-1}\left(\Phi(X)\right) - \frac{1}{\Phi^{-1}\left(\Phi(X)\right)}\log\left[1 - \frac{\Phi^{-1}\left(\Phi(X)\right)}{\phi\left(\Phi^{-1}\left(\Phi(X)\right)\right)} \frac{-\sqrt{n}\phi(X)\left[h\left(\frac{X}{\sqrt{n}}\right) - 1\right]}{X\sqrt{n}}\right]$$

$$X^{*} = X - \frac{1}{X} \log \left[1 - \frac{X}{\phi(X)} \frac{-\sqrt{n}\phi(X) \left[h\left(\frac{X}{\sqrt{n}}\right) - 1 \right]}{X\sqrt{n}} \right]$$

$$X^{*} = X - \frac{1}{X} \log \left[h\left(\frac{X}{\sqrt{n}}\right) \right], \tag{2.15}$$

que é uma transformação normalizadora de X e, em geral, tem distribuição normal padrão com erro da ordem de $O(n^{-3/2})$.

2.1.2 Razão de Verossimilhança Sinalizada

Expandindo a log-verossimilhança em série de Taylor em torno de $\widehat{\theta}$, tem-se

$$\ell(\boldsymbol{\theta}) = \ell(\widehat{\boldsymbol{\theta}}) + \ell'(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + \frac{\ell''(\widehat{\boldsymbol{\theta}})}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^2 + \dots$$
 (2.16)

truncando no terceiro termo e considerando que $\ell'(\widehat{\boldsymbol{\theta}}) = 0$, têm-se

$$\left(2\left(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\right)\right)^{1/2} \cong j(\widehat{\boldsymbol{\theta}})^{1/2} \left|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right| \tag{2.17}$$

$$sinal(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \left(2 \left(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}) \right) \right)^{1/2} \cong j(\widehat{\boldsymbol{\theta}})^{1/2} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$
 (2.18)

em que $j(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T$.

A expressão do lado esquerdo é definida por razão de verossimilhança Sinalizada e do lado direito temos a estatística de Wald, que segue distribuição N(0,1) com erro de ordem $O(n^{-1/2})$. Em grandes amostras essas estatísticas se tornam equivalentes.

2.1.3 Razão de Verossimilhança Sinalizada Modificada

Uma variação importante da razão de verossimilhanças é a razão de verossimilhanças sinalizada modificada ou, simplesmente, razão de verossimilhanças modificada dada por $R^*(\theta) = R(\theta) + R(\theta)^{-1} \log(U(\theta)/R(\theta))$ conforme Barndorff-Nielsen (1983), obtida a partir da teoria de ponto de sela e uma quantidade p^* condicionada a uma estatística ancilar τ , que por definição sua distribuição não depende do parâmetro, como por exemplo: se X_1, X_2, \ldots, X_n é uma variável aleatória idd de uma família de localização, então a estatística $S(\underline{X}) = X_n - X_1$ é uma estatística ancilar para o parâmetro de localização. Supondo que a variável aleatória pertença a uma família de escala uma estatística ancilar será $X_1/X_n, X_2/X_n, \ldots, X_{n-1}/X_n$.

De acordo com Barndorff-Nielsen (1980) e Barndorff-Nielsen (1983), uma aproximação para a densidade do estimador de máxima verossimilhança condicionada a uma estatística ancilar τ é dada por

$$p^*(\widehat{\boldsymbol{\theta}}|,\boldsymbol{\theta}) = c |\widehat{j}|^{1/2} e^{\left(\ell(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}},\boldsymbol{\tau}) - \ell(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}},\boldsymbol{\tau})\right)}, \tag{2.19}$$

em que $|\hat{j}|^{1/2}$ denota o determinante da matriz de informação observada estimada e $c=c(\pmb{\theta},\pmb{ au})$ é uma constante normalizadora escolhida de tal forma que a integral de p^* seja igual a um para $x\in(-\infty,\infty)$, para um valor $\pmb{ au}$ fixo, que geralmente é determinado numericamente.

Além de ser uma boa aproximação, a expressão (2.19), possui também algumas propriedades interessantes, conforme apresenta Lemonte (2010);

- i) é invariante sob transformações um-a-um dos dados;
- ii) é invariante sob reparametrizações;

iii) a constante c é invariante sob reparametrizações.

A aproximação de $p(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\tau})$ por $p^*(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\tau})$ possui, em geral, uma precisão de ordem $O(n^{-3/2})$, veja Skovgaard et al. (1990). Assim, $p(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\tau}) = p^*(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\tau})\{1 + O(n^{-3/2})\}$.

Pace e Salvan (1997) demonstram que $c(\theta, \boldsymbol{\tau}) = (2\pi)^{-k/2}\{1 - O(n^{-1}\})$, sendo k o número de parâmetros da distribuição, podendo se obter uma aproximação para $p^*(\widehat{\theta}|\theta, \boldsymbol{\tau})$ dada por

$$p^{\dagger}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\tau};\boldsymbol{\theta}) = (2\pi)^{-k/2}|j|^{-1/2}e^{\left(\ell(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}},\boldsymbol{\tau}) - \ell(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}},\boldsymbol{\tau})\right)},\tag{2.20}$$

com erro relativo em relação a p^* de ordem $O(n^{-1})$.

2.1.3.1 Casos Uniparamétricos

Nos casos uniparamétricos vamos considerar uma variável aleatória tendo distribuição de probabilidade $f(x;\theta)$ com θ sendo um parâmetro desconhecido. A log-verossimilhança para θ dada por $\ell(\theta) = \sum_i^n \log f(x_i;\theta)$, a estatística R pode se escrita como

$$R(\theta) = sinal(\widehat{\theta} - \theta) \left\{ 2 \left(\ell(\widehat{\theta}) - \ell(\theta) \right) \right\}^{1/2}. \tag{2.21}$$

Primeiro obtêm-se uma aproximação para a densidade de R com base na fórmula de p^* , tomando $c=(2\pi)^{-1/2}\overline{c}$ com $\overline{c}=1+O(n^{-1})$. Assumiu-se, que a estatística da razão de verossimilhanças sinalizada é uma função de $\widehat{\theta}$ invertível, e r um valor da estatística R para algum valor arbitrário de θ . Observe que

$$(1/2)r^2 = \ell(\widehat{\theta}) - \ell(\theta); \tag{2.22}$$

substituindo em (2.20) mudando a variável $\widehat{\theta}$ para r, e considerando \widehat{j} uma função de r, sendo $\partial \widehat{\theta}/\partial r$ o jacobiano da transformação de $\widehat{\theta}$ para r.

De (2.22), com $\ell(\theta) = \ell(\theta; \widehat{\theta}, \boldsymbol{\tau})$ temos $r(\partial r/\partial \widehat{\theta}) = \ell_{\widehat{\theta}}(\widehat{\theta}) - \ell_{\widehat{\theta}}(\theta)$ e $|\partial \widehat{\theta}/\partial r| = |r|/|\ell_{\widehat{\theta}}(\widehat{\theta}) - \ell_{\widehat{\theta}}(\theta)|$, observando-se que $\sqrt{2\pi} \ \phi(r) = \exp(-r^2/2)$. Uma aproximação para a densidade condicional de r é dada por

$$p^{\dagger}(r|\boldsymbol{\tau};\theta) = \bar{c} \frac{\hat{j}^{1/2}|r|}{|\ell_{\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}) - \ell_{\widehat{\boldsymbol{\theta}}}(\theta)|} \phi(r), \tag{2.23}$$

sendo $\hat{\theta}$ interpretado como uma função de r. Essa expressão é da forma (2.12) com $h(r/\sqrt{n}) = \hat{j}^{1/2} \, |r|/|\ell_{\widehat{\theta}}(\hat{\theta}) - \ell_{\widehat{\theta}}(\theta)|$.

Assim, a normalização de R é dada por

$$R^* = R + \frac{1}{R}\log\left(h\left(\frac{R}{\sqrt{n}}\right)\right) = R + \frac{1}{R}\log\left(\frac{U}{R}\right),\tag{2.24}$$

em que $U(\theta)=j(\widehat{\theta})^{-1/2}\{\ell_{\widehat{\theta}}(\widehat{\theta})-\ell_{\widehat{\theta}}(\theta)\}$, sendo o sinal de U determinado de tal forma que U/R nunca seja negativo, conforme Barndorff-Nielsen (1991). Outras aproximações para U são apresentadas em Brazzale, Davison e Reid (2007) e Lemonte (2010).

2.1.3.2 Casos Multiparamétricos

Nos casos multiparamétricos vamos considerar uma variável aleatória tendo distribuição de probabilidade $f(x; \boldsymbol{\theta})$ contínua com parâmetros desconhecidos $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_d\}$, denotando a log-verossimilhança $\ell(\boldsymbol{\theta}) = \sum_i^n \log f(x_i; \boldsymbol{\theta})$. Particionando o vetor dos parâmetros $\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda})$, tal que $\psi = \theta_1$ seja um parâmetro escalar de interesse e $\boldsymbol{\lambda} = \{\theta_2, \theta_3, \dots, \theta_d\}$ será chamado de vetor de parâmetros de pertubação com dimensão (d-1). Seja $\widehat{\boldsymbol{\theta}} = (\widehat{\psi}, \widehat{\boldsymbol{\lambda}})$ o estimador de máxima verossimilhança do modelo completo, e $\widehat{\boldsymbol{\theta}}_{\psi} = (\psi, \widehat{\boldsymbol{\lambda}}_{\psi})$ o estimador de máxima verossimilhança restrito de $\boldsymbol{\theta}$ para um valor fixo de ψ .

A estatística da razão de verossimilhanças sinalizada para fazer inferências sobre o parâmetro ψ é da forma $R(\psi) = sinal(\hat{\psi} - \psi) \sqrt{2\left(\ell(\widehat{\pmb{\theta}}) - \ell(\widehat{\pmb{\theta}}_{\psi})\right)}$.

A estatística da razão de verossimilhanças modificada para o caso multiparamétrico pode ser obtida de forma semelhante ao caso uniparamétrico, mas agora considerando o vetor de parâmetros de pertubação λ .

A densidade condicional de
$$R$$
 dado $\boldsymbol{\tau}$, é dada por $p(r|\boldsymbol{\tau};\boldsymbol{\theta}) = \frac{p(r,\widehat{\boldsymbol{\lambda}}_{\psi}|\boldsymbol{\tau};\boldsymbol{\theta})}{p(\widehat{\boldsymbol{\lambda}}_{\psi}|r,\boldsymbol{\tau};\boldsymbol{\theta})}$.

Através de $p^*(\widehat{\theta}|\tau,\theta)$ é possível encontrar uma aproximação para a densidade de $p(r|\tau,\theta)$.

A fórmula p^* da distribuição condicional de $\widehat{\boldsymbol{\theta}}$ dada uma estatística ancilar $\boldsymbol{\tau}$ é dada por $p^*(\widehat{\boldsymbol{\theta}}|\boldsymbol{\tau},\theta) = \frac{\bar{c}}{(2\pi)^{k/2}}|\widehat{\boldsymbol{j}}|^{1/2} \exp\left(\ell(\boldsymbol{\theta}) - \ell(\widehat{\boldsymbol{\theta}})\right)$. Esta densidade pode ser transformada para aproximar a densidade de $(R; \widehat{\boldsymbol{\lambda}}_{\psi})$, logo,

$$p^*(r, \widehat{\boldsymbol{\lambda}}_{\psi} | \boldsymbol{\tau}; \boldsymbol{\theta}) = \frac{\overline{c}}{(2\pi)^{k/2}} |\widehat{\boldsymbol{j}}|^{1/2} \exp(\ell(\boldsymbol{\theta}) - \ell(\widehat{\boldsymbol{\lambda}}_{\psi})) \exp(-[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\psi})])) | \boldsymbol{J} |. \quad (2.25)$$

Usando o fato de que $-r^2/2 = \ell(\widehat{m{ heta}}) - \ell(\widehat{m{ heta}}_\psi)$ tem-se

$$p^*(r, \widehat{\boldsymbol{\lambda}}_{\psi} | \boldsymbol{\tau}; \boldsymbol{\theta}) = \frac{\overline{c}}{(2\pi)^{k/2}} |\widehat{\boldsymbol{j}}|^{1/2} \exp(\ell(\boldsymbol{\theta}) - \ell(\widehat{\boldsymbol{\theta}}_{\psi})) \exp\left(\frac{-r^2}{2}\right) | \boldsymbol{J}|,$$

em que $\mid \boldsymbol{J} \mid$ denota o determinante do jacobiano da transformação de $\widehat{\boldsymbol{\theta}}$ para r.

Nesse caso, torna-se conveniente encontrar $|{m J}|^{-1}=\partial(r,\widehat{{m \lambda}}_\psi)^T/\partial(\widehat{{m \theta}})$, ao invés de calcular diretamente $|{m J}|$. Como considera-se r uma função de $\widehat{{m \theta}}$ e diferenciando implicitamente $-r^2/2=\ell(\widehat{{m \theta}})-\ell(\widehat{{m \theta}}_\psi)$, tem-se

$$\frac{\partial r}{\partial \widehat{\psi}} = \frac{1}{r} \left(\ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) \right) \quad \text{e} \quad \frac{\partial r}{\partial \widehat{\boldsymbol{\lambda}}} = \frac{1}{r} \left(\ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \right), \tag{2.26}$$

lembrando que $|\boldsymbol{J}^{-1}| = |\boldsymbol{J}|^{-1}$, em que

$$\boldsymbol{J}^{-1} = \left[\begin{array}{ccc} \frac{\partial r}{\partial \widehat{\psi}} & & \frac{\partial r}{\partial \widehat{\boldsymbol{\lambda}}} \\ \\ \frac{\partial \widehat{\boldsymbol{\lambda}}_{\psi}^T}{\partial \widehat{\psi}} & & \frac{\partial \widehat{\boldsymbol{\lambda}}_{\psi}^T}{\partial \widehat{\boldsymbol{\lambda}}} \end{array} \right].$$

Os demais termos de J^{-1} , ou seja, $\partial \widehat{\lambda}_{\psi}^T/\partial \widehat{\psi}$ e $\partial \widehat{\lambda}_{\psi}^T/\partial \widehat{\lambda}$ podem ser obtidos através da equação de verossimilhança utilizada para determinar $\widehat{\lambda}_{\psi}$. Assim, $\ell_{\lambda}(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}}) = \{\mathbf{0}\}_{k \times 1}$ em que $\{\mathbf{0}\}_{k \times 1}$ denota uma matriz nula com k linhas e 1 coluna. Diferenciando ℓ em relação a $\widehat{\psi}$ e $\widehat{\lambda}$, e usando a regra da cadeia, é possível mostrar que

$$\frac{\partial \widehat{\boldsymbol{\lambda}}_{\psi}^T}{\partial \widehat{\boldsymbol{\psi}}} = \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi})^{-1} \ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\psi}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \quad \text{e} \quad \frac{\partial \widehat{\boldsymbol{\lambda}}_{\psi}^T}{\partial \widehat{\boldsymbol{\psi}}} = \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi})^{-1} \ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi})$$

em que $j_{\lambda\lambda}=-\ell_{\lambda\lambda}$ representa a sub-matriz da informação observada de λ (BARNDORFF-NIELSEN, 1991; COX; BARNDORFF-NIELSEN, 1994).

Substituindo em J^{-1} , tem-se

$$\boldsymbol{J}^{-1} = \left[\begin{array}{cc} \frac{1}{r} \left(\ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) \right) & \quad \frac{1}{r} \left(\ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \right) \\ \\ \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi})^{-1} \ell_{\boldsymbol{\lambda}:\widehat{\boldsymbol{\psi}}}(\widehat{\boldsymbol{\theta}}_{\psi}) & \quad \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi})^{-1} \ell_{\boldsymbol{\lambda}:\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \end{array} \right],$$

que pode ser reescrita como

$$\boldsymbol{J}^{-1} = \begin{bmatrix} r & \mathbf{0}_k \\ \mathbf{0}_k & \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi}) \end{bmatrix}^{-1} \begin{bmatrix} \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) & \ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \\ \ell_{\boldsymbol{\lambda};\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) & \ell_{\boldsymbol{\lambda};\widehat{\boldsymbol{\lambda}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \end{bmatrix},$$

em que $\mathbf{0}_k$ denota uma matriz nula $(k \times k)$. Assim

$$| \boldsymbol{J} | = r | \boldsymbol{j}_{\lambda\lambda}(\widehat{\boldsymbol{\lambda}}_{\psi}) | / \begin{vmatrix} \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) & \ell_{;\widehat{\lambda}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi}) \\ \ell_{\lambda;\widehat{\psi}}(\widehat{\boldsymbol{\theta}}_{\psi}) & \ell_{\lambda;\widehat{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi}) \end{vmatrix}.$$

Considerando-se

$$|u| = \left| egin{aligned} \ell_{:\widehat{oldsymbol{ heta}}}(\widehat{oldsymbol{ heta}}) - \ell_{:\widehat{oldsymbol{ heta}}}(\widehat{oldsymbol{\lambda}}_{\psi}) \ \ell_{oldsymbol{\lambda}:\widehat{oldsymbol{ heta}}}(\widehat{oldsymbol{\lambda}}_{\psi}) \end{aligned}
ight| \left| egin{aligned} oldsymbol{j}_{oldsymbol{\lambda}\lambda}(\widehat{oldsymbol{\lambda}}_{\psi})^{1/2} \middle| \widehat{oldsymbol{j}} \widehat{oldsymbol{j}} \end{aligned}
ight|^{1/2},$$

logo

$$p^*(r, \widehat{\boldsymbol{\lambda}}_{\psi} | \boldsymbol{\tau}; \boldsymbol{\theta}) = \frac{\overline{c}}{(2\pi)^{k/2}} \frac{r |\boldsymbol{j}_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_{\psi})|^{1/2}}{u} \exp\{\boldsymbol{\ell}(\boldsymbol{\theta}) - \boldsymbol{\ell}(\widehat{\boldsymbol{\theta}}_{\psi})\} \exp\left[\frac{-r^2}{2}\right].$$
 (2.27)

Para o modelo em que ψ é fixo, pode-se usar $p^*(\widehat{\lambda}_{\psi}|r, \boldsymbol{\tau}; \theta)$ para obter uma aproximação para a densidade condicional de $\widehat{\lambda}_{\psi}$ dado r e $\boldsymbol{\tau}$. Segundo Cox e Barndorff-Nielsen (1994) essa aproximação é expressa como

$$p^*(\widehat{\boldsymbol{\lambda}}_{\psi}|\boldsymbol{\tau},\theta) = \frac{\overline{c}_{\psi}}{(2\pi)^(k-1)/2} |\boldsymbol{j}_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_{\psi})|^{1/2} \exp\{\boldsymbol{\ell}(\boldsymbol{\theta}) - \boldsymbol{\ell}(\widehat{\boldsymbol{\theta}}_{\psi})\}, \tag{2.28}$$

em que $\overline{c}_{\psi} = 1 + O(n^{-1})$. Assim, as aproximações $p^*(\widehat{\lambda}_{\psi}|\tau,\theta)$ e $p^*(r,\widehat{\lambda}_{\psi}|\tau;\theta)$ conduzem a uma aproximação da densidade condicional de R dado τ da forma

$$p^*(r|\boldsymbol{\tau};\boldsymbol{\theta}) = \frac{p^*(r,\widehat{\boldsymbol{\lambda}}_{\psi}|\boldsymbol{\tau};\boldsymbol{\theta})}{p^*(\widehat{\boldsymbol{\lambda}}_{\psi}|\boldsymbol{\tau};\boldsymbol{\theta})} = \frac{\bar{c}}{\bar{c}_{\psi}} \frac{1}{(2\pi)^{1/2}} \frac{r}{u} = \frac{\bar{c}}{\bar{c}_{\psi}} \frac{r}{u} \phi(r)$$

que é da forma (2.15), com $h(r/\sqrt{n})=r/u$, em que u=u(r). Logo, a estatística da razão de verossimilhanças modificada é

$$R^*(\boldsymbol{\theta}) = R(\boldsymbol{\theta}) + \frac{1}{R(\boldsymbol{\theta})} \log \left(\frac{U(\boldsymbol{\theta})}{R(\boldsymbol{\theta})} \right),$$
 (2.29)

em que

$$U = \begin{vmatrix} \ell_{;\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}) - \ell_{;\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \\ \ell_{\boldsymbol{\lambda}\cdot\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}_{\psi}) \end{vmatrix} / \left(\left| j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_{\psi}) \right|^{1/2} \left| \widehat{\boldsymbol{j}} \right|^{1/2} \right).$$
 (2.30)

Intervalos de confiança para o parâmetro de interesse utilizando as quantidades R e R^* são obtidos resolvendo equações do tipo $R(\psi) = \Phi(1-\alpha)$ ou $R^*(\psi) = \Phi(1-\alpha)$, em que α é o nível de significância, o que não é de fácil solução e geralmente esses intervalos são obtidos por meio de uma interpolação. Pode-se observar na equação (2.29), quando $\theta = \widehat{\theta}$, que $R(\theta) = 0$ Em geral, a curva de R^* possui uma assíntota vertical, o que gera alguns problemas para obtenção dos intervalos pela interpolação. Para contornar isso Severini (2000) sugere que a curva de R^* seja interpolada, limitando |R| > 0.1 e utilizando-se aproximadamente 30 pontos.

2.2 Aproximações para o ajuste de Barndorff-Nielsen

Mesmo R^* , proposto por Barndorff-Nielsen (1983), sendo uma melhoria para R, existe em muitos casos, que não pertencentes à família exponencial ou a modelos de transformação (localização e escala) uma grande dificuldade para determinar uma estatística ancilar au para obtenção das derivadas com relação ao espaço amostral de modo ainda que $(\widehat{\theta}, au)$ seja suficiente, isto é, estatísticas que guardam toda a informação que a amostra contém sobre o parâmetro que possibilitam o resumo dos dados sem perda de informação sobre θ . Para amenizar esta dificuldade, muitos autores como DiCiccio e Martin (1993), Skovgaard (1996), Severini (1999) e Fraser, Reid e Wu (1999), tem sugerido algumas aproximações para a quantidade $U(\psi)$ que não requerem a determinação de uma estatística ancilar, mas em alguns casos geram um prejuízo quanto a ordem de convergência, para mais detalhes veja Severini (2000) e Lemonte (2010).

2.2.1 Aproximação de DiCiccio-Martin

No caso de parâmetros ortogonais, isto é, $E\left(\partial^2\ell(\boldsymbol{\theta})/\partial\psi\partial\lambda^T\right)=\mathbf{0}_{1\times(d-1)}$, o que resulta em uma matriz de informação de Fisher bloco diagonal, a aproximação sugerida

por DiCiccio e Martin (1993), denotada por R_0^* , é obtida aproximando U por

$$U_0(\psi) = \ell_{\psi}(\widehat{\boldsymbol{\theta}}_{\psi}) \frac{\left| \boldsymbol{J}_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_{\psi}) \right|^{1/2} \boldsymbol{I}_{\psi\psi}(\widehat{\boldsymbol{\theta}})^{1/2}}{\left| \boldsymbol{J}(\widehat{\boldsymbol{\theta}}) \right|^{1/2} \boldsymbol{I}_{\psi\psi}(\widehat{\boldsymbol{\theta}}_{\psi})^{1/2}},$$

em que ℓ_{ψ} é a derivada com relação ao parâmetro de interesse, J e I representam a informação de Fisher observada e esperada, respectivamente.

Quando o parâmetro de interesse é um escalar é sempre possível obter uma parametrização ortogonal, segundo Cox e Reid (1987) e Cox e Reid (1993), mas as resoluções das equações diferenciais necessárias nem sempre são triviais, de acordo com Lemonte (2010).

2.2.2 Aproximações de Skovgaard e Severini (1999)

Uma aproximação para R^* é \hat{R}^* , apresentada por Severini (1999), em que as derivadas relação ao espaço amostral em U, aproximado por \hat{U} são dadas por

$$\begin{split} \widehat{\ell}_{;\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) &= \widehat{Q}(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}) \, \widehat{I}(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}})^{-1} J(\widehat{\boldsymbol{\theta}}), \\ \widehat{\ell}_{;\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}) &= \widehat{Q}(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}}) \, \widehat{I}(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}})^{-1} J(\widehat{\boldsymbol{\theta}}), \\ \widehat{\ell}_{\boldsymbol{\theta}:\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}) &= \, \widehat{I}(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}})^{-1} \widehat{I}(\widehat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}})^{-1} J(\widehat{\boldsymbol{\theta}}), \end{split}$$

com $\widehat{Q}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) = \sum\limits_{j=1}^n \ell^{(j)}(\boldsymbol{\theta}) \ell^{(j)}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})^T$ e $\widehat{I}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) = \sum\limits_{j=1}^n \ell^{(j)}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \ell^{(j)}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})^T$, em que $\ell^{(j)}(\boldsymbol{\theta})$ é a logverossimilhança calculada no j-ésimo elemento da amostra.

2.2.3 Aproximação de Fraser-Reid-Wu

Neste caso, uma aproximação \tilde{R}^* para R^* é realizada substituindo U por \tilde{U} , em que as derivadas com relação ao espaço amostral são aproximadas por

$$\widetilde{\ell}_{:\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = \ell_{:y}(\boldsymbol{\theta})\widehat{V}\big(\ell(\widehat{\boldsymbol{\theta}})\widehat{V}\big)^{-1}J(\widehat{\boldsymbol{\theta}}) \quad \text{e} \quad \widetilde{\ell}_{\boldsymbol{\theta}:\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = \ell_{\boldsymbol{\theta}:y}(\boldsymbol{\theta})\widehat{V}\big(\ell(\widehat{\boldsymbol{\theta}})\widehat{V}\big)^{-1}J(\widehat{\boldsymbol{\theta}}),$$

em que $\widehat{V} = \left[\frac{-\partial F(y_1; \pmb{\theta})/\partial \pmb{\theta}}{f(y_1; \pmb{\theta})}, \ldots, \frac{-\partial F(y_n; \pmb{\theta})/\partial \pmb{\theta}}{f(y_n; \pmb{\theta})}\right]_{\pmb{\theta} = \widehat{\pmb{\theta}}}^T$; F(.), f(.) são as funções de densidade acumulada e de probabilidade, respectivamente; \pmb{J} é a matriz de informação observada e $\ell_y(\cdot) = \partial \ell(\cdot)/\partial y$ é a derivada em relação aos elementos da amostra.

De acordo com Fraser, Reid e Wu (1999), R^* é melhor aproximada por \tilde{R}^* , pois $\tilde{U}^*=U^*+O_p(n^{-3/2})$, logo $\tilde{R}^*=R^*+O_p(n^{-3/2})$, enquanto $U_0=U+O_p(n^{-1})$ conforme DiCiccio e Martin (1993) e $\hat{U}=U+O_p(n^{-1})$ segundo Severini (1999).

Mais detalhes sobre estas aproximações podem ser encontradas em Severini (2000) e Lemonte (2010).

Distribuição Exponencial Generalizada

3.1 Distribuição Exponencial Generalizada

A distribuição exponencial generalizada de dois parâmetros denotada por EG, foi introduzida por Gupta e Kundu (1999) com função de distribuição $F(x;\delta,\lambda)=(1-e^{-\lambda x})^{\delta},$ em que $\delta,\lambda,x>0$, sendo λ o parâmetro de escala e δ o parâmetro de forma.

A função densidade de probabilidade da EG é dada por

$$f(x; \delta, \lambda) = \delta \lambda (1 - e^{-\lambda x})^{\delta - 1} e^{-\lambda x}.$$
 (3.1)

A seguir, é apresentado o comportamento da distribuição EG para alguns valores dos parâmetros de escala e forma. Note que $\mathsf{EG}(1,\lambda)$ representa a distribuição exponencial com parâmetro de escala λ .

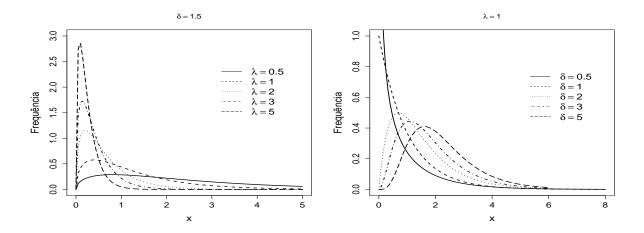


Figura 1 – Comportamento da distribuição EG para diversos valores dos parâmetros.

A distribuição $\mathrm{EG}(\delta,\lambda)$ pode ser usada de forma satisfatória em análise de dados de sobrevivência, particularmente, no lugar das distribuições Gamma e Weibull biparamétricas, por possuir uma maior flexibilidade na acomodação dos dados cuja característica aumentar ou diminuir a taxa de falhas, o que não ocorre com a distribuição exponencial que possui taxa de falha constante.

3.1.1 Funções de Risco e Sobrevivência

As funções de sobrevivência e de risco, associadas a EG são

$$S(x; \delta, \lambda) = 1 - (1 - e^{-\lambda x})^{\delta}$$
,

$$h(x; \delta, \lambda) = \frac{f(x; \delta, \lambda)}{1 - F(x; \delta, \lambda)} = \frac{\delta \lambda (1 - e^{-\lambda x})^{\delta - 1} e^{-\lambda x}}{1 - (1 - e^{-\lambda x})^{\delta}},$$

sendo que a forma da função taxa de falhas não depende do parâmetro λ .

Para um valor fixo de λ , a função taxa de falhas é crescente para $\delta>1$; de 0 à λ , e decrescente para $\delta<1$, de ∞ à λ ; para $\delta=1$, a função de risco, assim como no caso da exponencial, é constante e igual a λ , conforme a Figura 2.

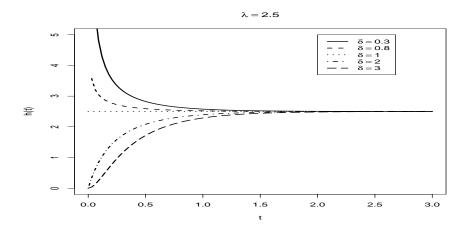


Figura 2 – Função taxa de falhas para $\lambda = 2.5$ e $\delta = 0.3$; 0.8; 1; 2; 3.

3.1.2 Quantis, Mediana e Moda

Pode-se definir uma função para os quantis, por meio da função acumulada isolando X em função de $q=F(x,\delta,\lambda)$ e, assim, tem-se $X_q=-(1/\lambda)\log\left(1-q^{1/\delta}\right)$. Para gerar amostras aleatórias, basta considerar uma amostra aleatória da distribuição uniforme, $q\sim U(0,1)$, e encontrar os respectivos valores de X_q . O tempo mediano de vida pode ser calculado substituindo q=0.5, assim $X_{med}=-(1/\lambda)\log\left(1-0.5^{1/\delta}\right)$. Se

 $\delta>1$, a moda pode ser obtida resolvendo a equação $\partial f(x;\delta,\lambda)/\partial x=0$, em que se tem $X_{mod}=\log(\delta)/\lambda$.

3.1.3 Função Geradora de Momentos

Por definição, $M(t) = E(e^{(tX)})$, assim, tem-se

$$M(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x; \delta, \lambda) dx = \int_0^\infty \delta \lambda \left(1 - e^{-\lambda x}\right)^{(\delta - 1)} e^{tx} e^{-\lambda x} dx, \tag{3.2}$$

considerando $y=e^{(-\lambda x)}\Rightarrow dy=-\lambda y\ dx$ e $x=-\log(y)/\lambda$, observa-se também que se $x=0\Rightarrow y=1$ e $x\to\infty\Rightarrow y\to0$ substituindo em (3.2), após algumas simplificações, tem-se que

$$M(t) = \delta \int_0^1 (1-y)^{\delta-1} y^{-\frac{t}{\lambda}} dy.$$
 (3.3)

Pela definição da função Beta

$$B(r,s) = \int_0^1 (1-t)^{s-1} t^{r-1} = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$
(3.4)

pode-se observar que

$$M(t) = \delta B(1 - t/\lambda, \delta)$$

$$= \delta \frac{\Gamma(1 - \frac{t}{\lambda})\Gamma(\delta)}{\Gamma(1 - \frac{t}{\lambda} + \delta)} = \frac{\Gamma(1 - \frac{t}{\lambda})\Gamma(\delta + 1)}{\Gamma(1 - \frac{t}{\lambda} + \delta)}; \quad t < \lambda,$$
(3.5)

em que $\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}dx$, fazendo uso da propriedade $\delta\Gamma(\delta)=\Gamma(\delta+1).$

3.1.4 Média e Variância

Calculando a primeira e segunda derivada de log(M(t)) em t=0, obtemos o primeiro e o segundo cumulante, que são a média e a variância, respectivamente. Assim

$$E(X) = \frac{1}{\lambda} \left(\psi(\delta + 1) - \psi(1) \right), \tag{3.6}$$

em que $\psi(x)=rac{\partial log(\Gamma(x))}{\partial x}=rac{\Gamma'(x)}{\Gamma(x)}$ é a função digama.

A variância pode ser obtida do cálculo do segundo cumulante, assim

$$Var(X) = \frac{1}{\lambda^2} \left(\psi'(1) - \psi'(\delta + 1) \right), \tag{3.7}$$

em que $\psi'(.)$ é a função trígama.

Se δ for conhecido, pode-se obter uma estimativa para λ através de (3.6), uma vez que $\widehat{\lambda}=(1/\overline{X})\,(\psi(\delta+1)-\psi(1))$, em que $\overline{X}=\left(\frac{1}{n}\right)\sum_{i=1}^n X_i$,

O coeficiente de variação não depende do parâmetro de escala λ , pois, por definição,

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{(\psi'(1) - \psi'(\delta + 1))}}{(\psi(\delta + 1) - \psi(1))}.$$
(3.8)

3.1.5 Estimadores de Máxima verossimilhança

Considera-se a estimação de δ e λ , ambos desconhecidos. Se $\{x_1, x_2, \ldots x_n\}$ é uma amostra aleatória a partir de EG (δ, λ) , então a verossimilhança é dada por

$$L(\delta, \lambda) = \prod_{i=1}^{n} f(x_i, \delta, \lambda) = \prod_{i=1}^{n} \delta \lambda \left(1 - e^{-\lambda x_i} \right)^{\delta - 1} e^{-\lambda x_i},$$

e, assim, a função de log-verossimilhança é

$$\ell(\delta, \lambda) = n \log(\delta) + n \log(\lambda) + (\delta - 1) \sum_{i=1}^{n} \log(1 - e^{-\lambda x_i}) - \lambda \sum_{i=1}^{n} x_i.$$
 (3.9)

Diferenciando (3.9) em relação aos parâmetros, as equações normais se tornam

$$\frac{\partial \ell(\delta, \lambda)}{\partial \delta} = 0 \Rightarrow \frac{n}{\delta} + \sum_{i=1}^{n} \log(1 - e^{-\lambda x_i}) = 0, \tag{3.10}$$

е

$$\frac{\partial \ell(\delta, \lambda)}{\partial \lambda} = 0 \Rightarrow \frac{n}{\lambda} + (1 - \delta) \sum_{i=1}^{n} \frac{x e^{-\lambda x_i}}{1 - e^{-\lambda x_i}} - \sum_{i=1}^{n} x_i = 0.$$
 (3.11)

A partir de (3.10), o estimador de δ como uma função de λ , diga-se $\widehat{\delta}(\lambda)$, é

$$\widehat{\delta}(\lambda) = \frac{-n}{\sum\limits_{i=1}^{n} \log(1 - e^{-\lambda x_i})}.$$
(3.12)

Para obter $\hat{\lambda}$, basta substituir (3.12) em (3.10) e resolver a equação resultante através de algum método numérico.

3.1.6 Intervalo de Confiança

Supondo que δ seja o parâmetro de interesse e λ o de perturbação, e ainda, decompondo $\boldsymbol{\theta}=(\lambda,\delta)$ com a notação $\widehat{\boldsymbol{\theta}}=(\widehat{\lambda},\widehat{\delta})$ e $\widehat{\boldsymbol{\theta}}_{\delta}=(\widehat{\lambda}_{\delta},\widehat{\delta})$, a razão de verossimilhanças Sinalizada pode ser escrita como $R(\delta)=sinal(\widehat{\delta}-\delta) \Big\{ 2 \left(\ell(\widehat{\lambda},\widehat{\delta})-\ell(\widehat{\lambda}_{\delta},\delta)\right) \Big\}^{1/2}$, em que $\widehat{\lambda}_{\delta}$ é o valor estimado de λ para um valor δ fixo.

O intervalo de confiança para o parâmetro δ pode ser obtido de $R(\delta)$, considerando que essa estatística possui distribuição assintótica normal padrão com erro $O(n^{-1/2})$.

Como alternativa, pode-se fazer uso da razão de verossimilhanças modificada, que depende da razão de verossimilhanças sinalizada e de uma quantidade $U(\delta)$ como segue

$$R^*(\delta) = R(\delta) + \frac{1}{R(\delta)} \log \left(\frac{U(\delta)}{R(\delta)} \right). \tag{3.13}$$

com

$$U(\delta) = \begin{vmatrix} \ell_{\widehat{\theta}}(\widehat{\boldsymbol{\theta}}) - \ell_{\widehat{\theta}}(\widehat{\boldsymbol{\theta}}_{\delta}) \\ \ell_{\lambda\widehat{\theta}}(\widehat{\boldsymbol{\theta}}_{\delta}) \end{vmatrix} / |j_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_{\delta})|^{1/2} |j(\widehat{\boldsymbol{\theta}})|^{1/2}, \tag{3.14}$$

em que $\ell_{;\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta})=\partial\ell(\boldsymbol{\theta})/\partial\widehat{\boldsymbol{\theta}},\quad \ell_{\lambda;\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta})=\partial^2\ell(\boldsymbol{\theta})/\partial\widehat{\boldsymbol{\theta}}\partial\lambda,\quad j_{\lambda\lambda}(\widehat{\boldsymbol{\theta}}_\delta)$ é o bloco da matriz de informação em relação a λ , sendo $\ell(\boldsymbol{\theta})=\ell(\delta,\lambda,\tau)$ e τ é uma estatística ancilar tal que $(\widehat{\delta},\widehat{\lambda},\tau)$ seja suficiente. Como a estatística ancilar não se obtêm facilmente, pode-se usar aproximações para obter as derivadas com relação ao espaço amostral. Neste caso, para a distribuição EG, adotou-se a aproximação sugerida por Fraser, Reid e Wu (1999), que fornece uma aproximação \tilde{R}^* para R^* , com erro de ordem $O(n^{-3/2})$.

3.1.7 Razão de verossimilhanças Modificada para EG

As aproximações necessárias para obtenção das derivadas com relação ao espaço amostral utilizadas na construção de \tilde{R}^* para a distribuição EG conforme visto na seção 2.2.3 são

$$\widehat{V} = \begin{bmatrix}
-\frac{(1 - e^{-\widehat{\lambda}y_1})\log(1 - e^{-\widehat{\lambda}y_1})e^{\widehat{\lambda}y_1}}{\widehat{\delta}\widehat{\lambda}} & -\frac{y_1}{\widehat{\lambda}} \\
\vdots & \vdots \\
-\frac{(1 - e^{-\widehat{\lambda}y_n})\log(1 - e^{-\widehat{\lambda}y_n})e^{\widehat{\lambda}y_n}}{\widehat{\delta}\widehat{\lambda}} & -\frac{y_n}{\widehat{\lambda}}
\end{bmatrix},$$
(3.15)

$$\ell_{;\mathbf{y}}(\boldsymbol{\theta}) = \left(\frac{(\delta-1)\lambda e^{(-\lambda y_1)}}{1 - e^{(-\lambda y_1)}} - \lambda \quad \frac{(\delta-1)\lambda e^{(-\lambda y_2)}}{1 - e^{(-\lambda y_2)}} - \lambda \quad \dots \quad \frac{(\delta-1)\lambda e^{(-\lambda y_n)}}{1 - e^{(-\lambda y_n)}} - \lambda\right) (3.16)$$

$$\ell_{\boldsymbol{\theta};y}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \delta \partial y_{1}} & \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \delta \partial y_{2}} & \cdots & \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \delta \partial y_{n}} \\ & & & \\ \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \lambda \partial y_{1}} & \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \lambda \partial y_{2}} & \cdots & \frac{\partial^{2}\ell(\boldsymbol{\theta})}{\partial \lambda \partial y_{n}} \end{bmatrix}, \tag{3.17}$$

$$\begin{array}{ll} \operatorname{com} & \frac{\partial \ell(\boldsymbol{\theta})}{\partial \delta \, \partial y_i} = \frac{\lambda e^{-\lambda y_i}}{1 - e^{-\lambda y_i}}, & \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda \partial y_i} = -\frac{(-\delta - 1 + y_i \lambda \delta - x \lambda) e^{-\lambda y_i} + \delta e^{-2\lambda x} + 1}{(-1 + e^{-\lambda y_i})}, \\ \operatorname{e} & i = 1, \ldots, n. \end{array}$$

A matriz de informação observada J é

$$J(\boldsymbol{\theta}) = \begin{bmatrix} \frac{n}{\delta^2} & \sum_{i=1}^{n} \frac{y_i e^{-\lambda y_i}}{1 - e^{-\lambda y_i}} \\ \sum_{i=1}^{n} \frac{y_i e^{-\lambda y_i}}{1 - e^{-\lambda y_i}} & \frac{n}{\lambda^2} + (\delta - 1) \sum_{i=1}^{n} \frac{y_i^2 e^{-\lambda y_i}}{(1 - e^{-\lambda y_i})^2} \end{bmatrix}.$$

Resultados das Simulações

4.1 Intervalos de confiança para o parâmetro δ

Para comparar as estatísticas de Wald, R e R^* , foram geradas 10000 amostras aleatórias por meio da transformação inversa para a distribuição EG, em que considerou-se o parâmetro de escala como conhecido e fixo, $\lambda=1$, e variando o parâmetro $\delta=0.8,\,1.5$ e 2, e com tamanhos amostrais de $n=5,\,10,\,20,\,30,\,50,\,100$, para os níveis 95% e 99% de confiança.

Foram realizados testes de hipóteses sob $H_0:\delta=0.8;\,1.5$ com objetivo de comparar a normalidade das estatísticas de teste. Foram construídos gráficos "Quantil \times Quantil" para uma visualização da convergência assintótica para distribuição normal padrão. E, ainda, obteve-se os intervalos de confiança usando três estatísticas e verificou-se as taxas de cobertura, amplitudes médias destes intervalos e as taxas de não coberturas a direita e à esquerda, bem como, os coeficientes de variação para amplitude média dos intervalos.

Para isso foram utilizadas rotinas do *Software* R Core Team (2015) com a *procedure* Fraser.Reid (Davison, 2009), disponível no site http://statwww.epfl.ch/AA, que gerou resultados das aproximações de Fraser, Reid e Wu (1999) para R^* da distribuição EG.

A seguir são apresentados gráficos e tabelas dos resultados da distribuição das estatísticas de teste, coberturas dos intervalos de confiança e as amplitudes médias dos intervalos com os respectivos coeficientes de variação para a amplitude das estatísticas Wald, R e R^* .

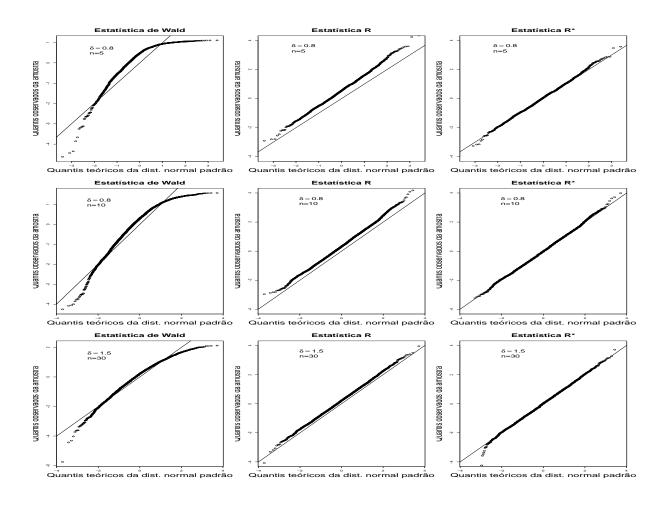


Figura 3 – Distribuição das estatísticas de Teste para as metodologias em estudo para $n=5;15;30,\,\delta=0.8$ e nível significância de 5%.

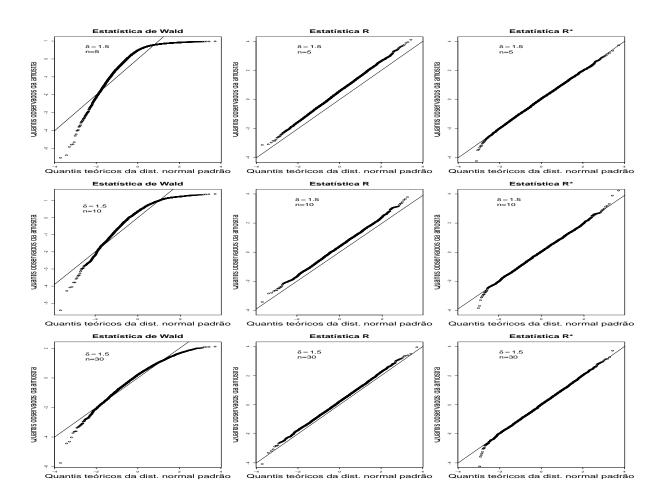


Figura 4 – Distribuição das estatísticas de Teste para as metodologias em estudo para $n=5;15;30,\,\delta=1.5$ e nível significância de 5%.

Pode-se verificar através dos gráficos, para $\delta=0.8$ e $\delta=1.5$, que para pequenas amostas a estatística de Wald não possui normalidade, R possui uma distribuição "normal" mas deslocada para direita, ou seja, com uma média diferente de zero, enquanto R^* tem a melhor convergência para a normal padrão. À medida que aumenta-se o tamanho da amostra, as estatísticas em estudo tendem a normal padrão, mas com vantagem clara para R^* , evidenciando assim ser uma estatística mais precisa, ou seja, menos sensível ao tamanho da amostral.

Tabela 1 — Cobertura dos intervalos de confiança e amplitude média dos intervalos com seus coeficientes de variação para $\lambda=1$ e $\delta=0.8,1.5,2$ com nível significância de 5% e 1% .

	I	I		δ -	- N Q			
n	Méto-	$\delta =$ 0.8 $_{ m IC~95\%}$ $_{ m IC~99\%}$						
••	dologia	Cobertura	Amplitude	CV	Cobertura	Amplitude	CV	
	W	0.9754; (0.0000; 0.0246)	61.1138	0.00654	0.9888; (0.0000; 0.0112)	35.3802	0.00606	
5	R	0.9020; (0.0906; 0.0074)	4.8747	0.00022	0.9738; (0.0252; 0.0010)	6.4335	0.00019	
	R*	0.9454; (0.0304; 0.0244)	3.9933	0.00024	0.9864; (0.0058; 0.0078)	5.3597	0.00021	
	W	0.9708; (0.0000; 0.0292)	1.9463	0.00027	0.9870; (0.0000; 0.0130)	2.5754	0.00025	
10	R	0.9330; (0.0558; 0.0112)	1.9974	0.00016	0.9810; (0.0162; 0.0028)	2.7373	0.00015	
	R*	0.9538; (0.0232; 0.0232)	1.7923	0.00016	0.9872; (0.0072; 0.0056)	2.4683	0.00015	
	W	0.9630; (0.0028; 0.0342)	1.0506	0.00009	0.9928; (0.0000; 0.0072)	1.3985	0.00008	
20	R	0.9412; (0.0408; 0.0180)	1.0805	0.00009	0.9866; (0.0108; 0.0026)	1.4679	0.00009	
	R*	0.9494; (0.0240; 0.0266)	1.029	0.00009	0.9924; (0.0044; 0.0032)	1.3985	0.00008	
	W	0.9604; (0.0090; 0.0306)	0.811	0.00006	0.9894; (0.0000; 0.0106)	1.0603	0.00006	
30	R	0.9462; (0.0378; 0.0160)	0.8254	0.00006	0.9884; (0.0080; 0.0036)	1.0925	0.00006	
	R*	0.9534; (0.0236; 0.0230)	0.7996	0.00006	0.9888; (0.0046; 0.0066)	1.0589	0.00006	
	W	0.9552; (0.0124; 0.0324)	0.5951	0.00004	0.9914; (0.0004; 0.0082)	0.7872	0.00004	
50	R	0.9476; (0.0320; 0.0204)	0.6009	0.00004	0.9898; (0.0080; 0.0022)	0.8008	0.00005	
	R*	0.9510; (0.0226; 0.0264)	0.5899	0.00004	0.9904; (0.0052; 0.0044)	0.7863	0.00005	
	W	0.9512; (0.0202; 0.0286)	0.4092	0.00003	0.9908; (0.0012; 0.0080)	0.537	0.00003	
100	R	0.9462; (0.0332; 0.0206)	0.4111	0.00003	0.9896; (0.0056; 0.0048)	0.541	0.00003	
	R*	0.9482; (0.0270; 0.0248)	0.40737	0.00003	0.9906; (0.0040; 0.0054)	0.536	0.00003	
				$\delta = 1.5$	7 7 7			
	W	0.9734; (0.0000; 0.0266)	341.0793	0.00367	0.9872; (0.0000; 0.0128)	460.5332	0.00310	
5	R	0.8906; (0.1006; 0.0088)	9.1626	0.00016	0.9687; (0.0300; 0.0014)	11.7634	0.00015	
-	R*	0.9478; (0.0264; 0.0258)	7.6383	0.00018	0.9810; (0.0080; 0.0110)	9.9263	0.00017	
	W	0.9698; (0.0000; 0.0302)	4.8842	0.00036	0.9894; (0.0000; 0.0106)	7.0666	0.00061	
10	R	0.9310; (0.0590; 0.0100)	4.4342	0.00013	0.9780; (0.0206; 0.0014)	5.9798	0.00013	
. •	R*	0.9496; (0.0258; 0.0246)	3.8858	0.00014	0.9842; (0.0072; 0.0086)	5.2692	0.00014	
	W	0.9698; (0.0000; 0.0302)	2.3482	0.00010	0.9870; (0.0000; 0.0130)	3.1132	0.00011	
20	R	0.9454; (0.0412; 0.0134)	2.4426	0.00009	0.9866; (0.0110; 0.0024)	3.3133	0.00010	
	R*	0.9554; (0.0216; 0.0230)	2.2906	0.00009	0.9906; (0.0038; 0.0056)	3.1098	0.00009	
	W	0.9608; (0.0052; 0.0340)	1.7606	0.00008	0.9886; (0.0000; 0.0114)	2.3062	0.00008	
30	R	0.9436; (0.0384; 0.0180)	1.8105	0.00008	0.9898; (0.0082; 0.0020)	2.42	0.00008	
	R*	0.9482; (0.0248; 0.0270)	1.7361	0.00008	0.9914; (0.0044; 0.0042)	2.3232	0.00008	
	W	0.9632; (0.0102; 0.0266)	1.2807	0.00005	0.9902; (0.0002; 0.0096)	1.6755	0.00005	
50	R	0.9474; (0.0366; 0.0160)	1.3017	0.00005	0.9902; (0.0072; 0.0026)	1.7233	0.00005	
	R*	0.9540; (0.0240; 0.0220)	1.2705	0.00005	0.9916; (0.0036; 0.0048)	1.6825	0.00005	
	W	0.9568; (0.0116; 0.0316)	0.8614	0.00004	0.9878; (0.0008; 0.0114)	1.1294	0.00004	
100	R	0.9486; (0.0300; 0.0214)	0.868	0.00004	0.9900; (0.0070; 0.0030)	1.1445	0.00004	
	R*	0.9514; (0.0220; 0.0266)	0.8576	0.00004	0.9882; (0.0056; 0.0062)	1.13104	0.00004	
			1	$\delta=2$	1			
-	W	0.9703; (0.0000 ; 0.0297)	315.0081	0.00302	0.9874; (0.0000; 0.0126)	393.5137	0.00379	
5	R	0.8964; (0.0960; 0.0076)	11.4357	0.00013	0.9756; (0.0240; 0.0004)	14.61056	0.00012	
	R*	0.9413; (0.0281; 0.0305)	9.6328	0.00016	0.9780; (0.0038; 0 0.018)	12.26198	0.00015	
	W	0.9663; (0.0000; 0.0337)	7.9932	0.00051	0.9816; (0.0000; 0.0184)	10.15386	0.00054	
10	R	0.9228; (0.0650; 0.0122)	6.1548	0.00012	0.9822; (0.0140; 0.0038)	7.920989	0.00011	
	R*	0.9417; (0.0274; 0.0309)	5.4007	0.00013	0.9828; (0.0046; 0.0126)	6.97835	0.00012	
	W	0.9667; (0.0000; 0.0333)	3.4166	0.00011	0.9852; (0.0000; 0.0148)	4.48133	0.00013	
20	R	0.9425; (0.0447; 0.0128)	3.4879	0.00009	0.9856; (0.0120; 0.0024)	4.628728	0.00009	
	R*	0.9513; (0.0246; 0.0243)	3.2405	0.00009	0.9872; (0.0064; 0.0064)	4.321421	0.00009	
	W	0.9621; (0.0006; 0.0373)	2.5226	0.00008	0.9864; (0.0000; 0.0136)	3.35107	0.00009	
30	R	0.9393; (0.0431; 0.0176)	2.5956	0.00008	0.9872; (0.0102; 0.0026)	3.506255	0.00008	
	R*	0.9493; (0.0248; 0.0260)	2.4766	0.00008	0.9894; (0.0054; 0.0052)	3.347255	0.00008	
	W	0.9623; (0.0074; 0.0303)	1.821	0.00006	0.9886; (0.0000; 0.0114)	2.405658	0.00006	
50	R	0.9489; (0.0345; 0.0166)	1.8579	0.00006	0.9868; (0.0092; 0.0040)	2.489153	0.00006	
	R*	0.9537; (0.0238; 0.0226)	1.8075	0.00006	0.9890; (0.0054; 0.0056)	2.422448	0.00006	
	W	0.9558; (0.0130; 0.0312)	1.2249	0.00004	0.9898; (0.0010; 0.0092)	1.606308	0.00004	
100	R	0.9484; (0.0316; 0.0200)	1.2367	0.00004	0.9892; (0.0070; 0.0038)	1.633325	0.00004	
	R*	0.9496; (0.0244; 0.0260)	1.2203	0.00004	0.9912; (0.0040; 0.0044)	1.611687	0.00004	

¹ Os valores entre parênteses são as taxas de não cobertura nas respectivas laterais dos intervalos.

Pode-se observar pela Tabela 1 que a estatística de Wald para pequenas amostras apresentou coberturas acima do valor nominal, mas com resultados ruins quanto a amplitude dos intervalos. Observe que para $n=5,\,\alpha=0.05$ e $\delta=0.8,$ a cobertura chega 98% (com uma amplitude de 68) contra de 90% para R (com uma amplitude 4.9) e 94.5% para R^* (com amplitude média de 3.99) com praticamente o mesmo coeficiente de variação.

Pode-se observar também que a medida que o tamanho da amostra aumenta, os resultados tornam-se cada vez mais semelhantes, como era esperado. Por exemplo, para n=100, $\alpha=0.01$ e $\delta=2$, os resultados são praticamente os mesmos, mas com uma pequena vantagem para R^* . Outro fato a ser observado são as taxas de não cobertura laterais que mesmo para pequenas amostras, R^* apresentam-se melhor distribuídas em torno do intervalo, evidenciando assim um melhoramento no viés da estatística R ocasionado pela assimetria comum para pequenas amostras.

4.1.1 Discussões

Os resultados obtidos para a distribuição da estatística de teste para os testes de hipóteses simulados deixam evidentes o melhor desempenho da estatística da razão de verossimilhança modificada mesmo quando o tamanho da amostra é pequeno. A estatística de Wald nestes casos tem um desempenho muito ruim quanto a convergência assintótica para a normal padrão, como era esperado, mostrando-se muito sensível ao tamanho da amostra, enquanto R apresentou distribuições com deslocamento para a direita, o que tem como consequência tomadas de decisões incorretas, uma vez que a região de aceitação encontra-se deslocada em relação a distribuição normal padrão, ao mesmo tempo em que R^* tem uma maior aderência à distribuição normal padrão em todos os casos analisados evidenciando ser uma estatística muito útil em situações onde há amostras pequenas e moderadas.

Quanto aos intervalos de confiança pode-se verificar que para amostras pequenas a cobertura dos intervalos de confiança para estatística de Wald foi maior, mas com uma amplitude muito superior as demais estatísticas, mostrando ser uma estatística imprecisa para pequenas amostras, no entanto, a medida que o tamanho da amostra aumenta, essa amplitude diminui, mas podemos observar que R^{*} em todos os cenários teve uma taxa de cobertura superior a de R com intervalos de menores amplitudes, evidenciando a superioridade de R^{*} em relação as demais estatísticas, resultados estes que corroboram com outros estudos divulgados e que serviram de referências para o desenvolvimento desta pesquisa.

Capítulo 5

Aplicação

5.1 Aplicação

Apresentaremos os intervalos de confiança obtidos através de R, R^* e da estatística de Wald utilizando um banco de dados disponibilizado pelo sistema de informática do sistema único de saude (DATASUS). Os dados representam gastos anuais de internações hospitalares de pacientes com problemas psiquiátricos no ano de 2002, dividido em gastos com pacientes do sexo feminino e sexo masculino, em diversos hospitais do estado do Paraná. Os dados estão apresentados na Tabela 2.

Tabela 2 – Distribuição dos gastos com internações hospitalares (em milhões de reais) no ano de 2002 em hospitais psiquiátricos no estado do Paraná.

Hospital	Masculino	Feminino
1	1.240.709,18	209.096,46
2	1.323.144,38	781.252,74
3	940.217,89	526.587,18
4	1.862.117,30	569.384,94
5	451.618,58	228.523,68
6	134.622,90	50.358,60
7	497.773,44	1.084.148,16
8	2.718.314,50	1.604.164,40
9	1.265.812,38	1.139.890,46
10	565.670,70	174.921,90
11	353.705,56	323.441,65
12	150.421,62	47.022,44
13	100.105,11	63.879,51
14	233.582,70	143.925,00
15	2.344.835,75	845.245,40
16	1.017.696,96	501.916,34
17	1.615.720,00	600.135,74

Fonte: Departamento de informática do sistema único de saúde (DATASUS).

Utilizamos a distribuição EG para ajustar os dados da Tabela 2 e pode-se observar por meio da Figura 5 que é razoável a suposição de que os dados seguem distribuição EG, apesar de termos um ponto discrepante.

Capítulo 5. Aplicação 35

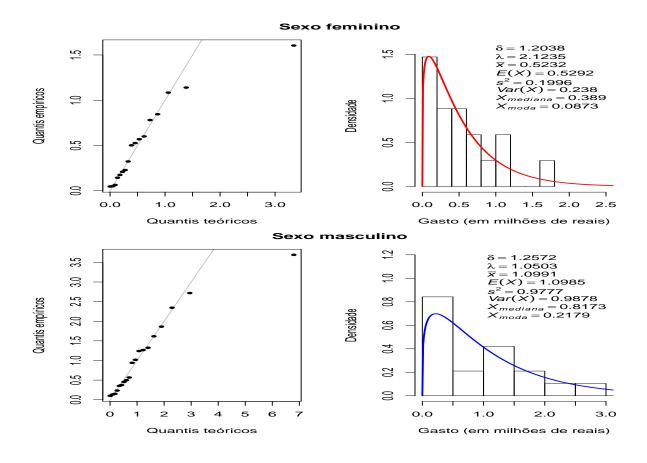


Figura 5 – Comportamento dos Gastos com internações hospitalares.

Para a obtenção de R^* , as derivadas com relação ao espaço amostral foram aproximadas da mesma forma que no estudo de a simulação anterior. Também, neste caso, utilizamos a função Frazer.reid do $software\ R$, utilizada dentro das simulações. Os resultados foram obtidos em 2 cenários, para amostras sorteadas aleatoriamente do banco de dados com n=5,10 e o banco de dados completo n=17, com o objetivo de verificar o comportamento das metodologias em questão.

Os resultados estão apresentados nas Tabelas 3 e 4.

		Parâmetro		IC(δ, 95%)			
n	Método	δ	se	Li	Ls	Amplitude	CV
	R	0.738688	0.414968	0.216036	2.013592	1.797556	0.561763
5	R^*	0.561911	0.343529	0.147499	1.646827	1.499327	0.611358
10	R	1.200378	0.507447	0.490659	2.600493	2.109834	0.422739
10	R^*	1.037252	0.457436	0.407220	2.316704	1.909484	0.441008
17	R	1.23672	0.404898	0.62724	2.27297	1.64573	0.32740
1 /	R^*	1.13496	0.38077	0.56578	2.11490	1.54912	0.33549

Tabela 3 — Intervalos de confiança para δ : Sexo Feminino

Capítulo 5. Aplicação 36

		Parâmetro		IC(δ, 95%)			
n	Método	δ	se	Li	Ls	Amplitude	$CV(\delta)$
5	R	1.516883	0.94929	0.390069	4.65506	4.264986	0.625818
	R^*	1.101238	0.746838	0.251869	3.684710	3.432841	0.67817
10	R	1.122411	0.475924	0.459245	2.447101	1.98786	0.424019
10	R^*	0.968248	0.424968	0.382035	2.174913	1.792878	0.438904
17	R	1.37892	0.45998	0.69114	2.567127	1.875987	0.33358
1 /	R^*	1.26084	0.43093	0.621780	2.38334	1.76156	0.34178

Tabela 4 — Intervalos de confiança para δ : Sexo Masculino

Como se pode observar nas Tabelas 3 e 4, em todas amostras as amplitudes dos intervalos obtidos por meio de R^* são menores do que aqueles obtidos por meio de R e os valores das amplitudes para as amostras de tamanho 5 e 10 estão mais próximos do valor obtidos quando usamos o banco de dados completo, indicando uma maior precisão dos resultados. Como média é diretamente proporcional ao parâmetro δ conforme 3.6 isto nos permite a obter planejamentos mais assertivos dos gastos com os internamentos.

5.1.1 Considerações finais e estudos futuros

Pode-se verificar que a normalidade assintótica é mais rapidamente atingida por R^* , como previsto em teoria, e os intervalos de confiança possuem uma cobertura muito próxima do valor nominal, assim como uma amplitude inferior às demais metodologias estudadas, o que é desejável, uma vez que, indica uma maior precisão dos resultados, o que torna R^* muito eficiente para estudos quando o tamanho da amostra é pequeno ou moderado, muito comum nas áreas de saúde e Biológicas.

Todos os resultados aqui obtidos foram realizados considerando apenas os parâmetros da distribuição EG. Como sugestão para estudos futuros pode-se verificar o desempenho de R^* sobre a razão entre médias e verificar o desempenho de R^* na presença de covariáveis, tanto no caso da distribuição EG, como para outras distribuições.

Referências

BARNDORFF-NIELSEN, O. Conditionality resolutions. *Biometrika*, Biometrika Trust, v. 67, n. 2, p. 293–310, 1980.

BARNDORFF-NIELSEN, O. On a formula for the distribution of the maximum likelihood estimator. biometrika 70 343–365. *Mathematical Reviews (MathSciNet): MR712023 Digital Object Identifier: doi*, v. 10, 1983.

BARNDORFF-NIELSEN, O.; COX, D. R. Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 279–312, 1979.

BARNDORFF-NIELSEN, O. E. Infereni on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, Biometrika Trust, v. 73, n. 2, p. 307–322, 1986.

BARNDORFF-NIELSEN, O. E. A note on the standardized signed log likelihood ratio. *Scandinavian Journal of Statistics*, JSTOR, p. 157–160, 1990.

BARNDORFF-NIELSEN, O. E. Modified signed log likelihood ratio. *Biometrika*, Biometrika Trust, v. 78, n. 3, p. 557–563, 1991.

BARNDORFF-NIELSEN, O. E.; COX, D. R. Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 483–495, 1984.

BARNDORFF-NIELSEN, O. E.; COX, D. R. *Asymptotic techniques for use in statistics*. [S.l.]: Chapman\ & Hall, 1989.

BARTLETT, M. S. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, JSTOR, p. 268–282, 1937.

BARTLETT, M. S. Approximate confidence intervals. ii. more than one unknown parameter. *Biometrika*, JSTOR, p. 306–317, 1953.

BOLFARINE, H.; SANDOVAL, M. C. Introdução à inferência estatística. [S.I.]: SBM, 2001. v. 2.

BRAZZALE, A. R.; DAVISON, A. C.; REID, N. *Applied asymptotics: case studies in small-sample statistics*. [S.I.]: Cambridge University Press, 2007. v. 23.

BRAZZALE, A. R.; DAVISON, A. C. e. o. Accurate parametric inference for small samples. *Statistical Science*, Institute of Mathematical Statistics, v. 23, n. 4, p. 465–484, 2008.

Referências 38

CORTESE, G.; VENTURA, L. Likelihood asymptotics for the stress-strength model p (x< y). In: MAGGIOLI EDITORE. S. Co. 2009. Sixth Conference. Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction. [S.I.], 2009. p. 133.

- COX, D.; BARNDORFF-NIELSEN, O. Inference and asymptotics. [S.I.]: CRC Press, 1994. v. 52.
- COX, D.; REID, N. A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 467–471, 1993.
- COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–39, 1987.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 248–275, 1968.
- DANIELS, H. E. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, JSTOR, p. 631–650, 1954.
- DANIELS, H. E. Tail probability approximations. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 37–48, 1987.
- DAVISON, A. C.; FRASER DONALD AS E REID, N. Improved likelihood inference for discrete data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 68, n. 3, p. 495–508, 2006.
- DICICCIO, T. J.; MARTIN, M. A. Simple modifications for signed roots of likelihood ratio statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 305–316, 1993.
- EFRON, B.; HINKLEY, D. V. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, Biometrika Trust, v. 65, n. 3, p. 457–483, 1978.
- FISHER, R. A. Theory of statistical estimation. In: CAMBRIDGE UNIV PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.I.], 1925. v. 22, n. 05, p. 700–725.
- FRASER, D. A. S.; REID, N.; WU, J. A simple general formula for tail probabilities for frequentist and bayesian inference. *Biometrika*, Biometrika Trust, v. 86, n. 2, p. 249–264, 1999.
- GUPTA, R. D.; KUNDU, D. Theory & methods: Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, Wiley Online Library, v. 41, n. 2, p. 173–188, 1999.
- JENSEN, J. L. The modified signed likelihood statistic and saddlepoint approximations. *Biometrika*, Biometrika Trust, v. 79, n. 4, p. 693–703, 1992.
- KOLASSA, J. Series approximation methods in statistics. *New York: Springer-Verl*, n. Lecture Notes in Statistics 88, 1994.
- LEMONTE, A. J. Estatistica gradiente e refinamento de métodos assintóticos no modelo de regressao Birnbaum-Saunders. Tese (Doutorado) Universidade de Sao Paulo, 2010.
- LOZADA-CAN, C.; DAVISON, A. Three examples of accurate likelihood inference. *The American Statistician*, v. 64, n. 2, 2010.
- PACE, L.; SALVAN, A. *Principles of statistical inference: from a Neo-Fisherian perspective.* [S.I.]: World Scientific, 1997. v. 4.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. Disponível em: http://www.R-project.org/>.

Referências 39

REID, N. Saddlepoint methods and statistical inference. *Statistical Science*, JSTOR, p. 213–227, 1988.

REID, N. Likelihood and higher-order approximations to tail areas: A review and annotated bibliography. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, JSTOR, p. 141–166, 1996.

SEVERINI, T. A. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, Biometrika Trust, v. 86, n. 2, p. 235–247, 1999.

SEVERINI, T. A. Likelihood methods in statistics. [S.I.]: Peterson's, 2000. v. 22.

SKOVGAARD, I. M. An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, JSTOR, p. 145–165, 1996.

SKOVGAARD, I. M. et al. On the density of minimum contrast estimators. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 18, n. 2, p. 779–789, 1990.