Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Modelagem Matemática e Computacional

APERFEIÇOAMENTO DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS BASEADO NA FUNÇÃO DE VEROSSIMILHANÇA PERFILADA

Emília Gonçalves de Lima Neta

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional, UFPB, da Universidade Federal da Paraíba, como parte dos requisitos necessários à obtenção do título de Mestre em Modelagem Matemática e Computacional.

Orientadores: Claudio Javier Tablada
Renilma Pereira da Silva

João Pessoa
Fevereiro de 2021

APERFEIÇOAMENTO DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS BASEADO NA FUNÇÃO DE VEROSSIMILHANÇA PERFILADA

Emília Gonçalves de Lima Neta

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL (PPGMMC) DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DA PARAÍBA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM MODELAGEM MATEMÁTICA E COMPUTACIONAL.

Examinada por:

Prof. Claudio Javier Tablada, Dr.

Profª. Renilma Pereira da Silva, Dra.

Prof. Pedro Rafael Diniz Marinho, Dr.

Profª. Mariana Correia de Araújo, Dra.

JOÃO PESSOA, PB – BRASIL
FEVEREIRO DE 2021

Catalogação na publicação
Seção de Catalogação e Classificação

N469a Lima Neta, Emília Gonçalves de.
Aperfeiçoamento do teste da razão de verossimilhanças baseado na função de verossimilhança perfilada / Emília Gonçalves de Lima Neta. - João Pessoa, 2021.
64 f. : il.

Orientação: Claudio Javier Tablada.
Coorientação: Renilma Pereira da Silva.
Dissertação (Mestrado) - UFPB/CI.

1. Verossimilhança perfilada. 2. Testes aperfeiçoados. 3. Distribuições de probabilidade. 4. Qui-quadrado inf. I. Tablada, Claudio Javier. II. Silva, Renilma Pereira da. III. Título.

UFPB/BC CDU 519.226(043)

Elaborado por ANNA REGINA DA SILVA RIBEIRO - CRB-15/024

Aos meus pais, Francisco (in memoriam) e Maria, por todo apoio ao longo da minha trajetória.

Agradecimentos

A Deus, por sempre me dar força e coragem ao longo da minha trajetória acadêmica.

Aos meus pais, Francisco Gonçalves (in memoriam) e Maria Soares pelos ensinamentos, apoio, confiança e amor que me dedicaram durante toda minha vida. Sem dúvidas, essa conquista é mais de vocês do que minha.

Aos meus irmãos, Raimundo e Cícero, por todo apoio.

Ao meu marido, Leandro Dantas, por todo apoio e amor durante minha vida acadêmica. Sem dúvidas, foi uma pessoa essencial para concretização desse sonho.

A todos os meus amigos pelas contribuições, especialmente a André Ricardo e Janete Alves.

Aos meus colegas de trajetória, em especial a Adriana Ribeiro, Katy Castro e Júnior Duarte, por todos os momentos vivenciados, pela garra, por toda dedicação e por terem tornado o caminho mais fácil com suas contribuições e amizade.

Aos meus orientadores, Dr. Claudio Tablada e Dra. Renilma Pereira, por todo incentivo, ensinamentos e paciência, durante o desenvolvimento deste trabalho.

Ao professor Dr. Pedro Rafael, por toda contribuição com as simulações na linguagem de programação R, sem dúvida, foi fundamental para concretização desta pesquisa.

A banca examinadora por todas as sugestões de melhoria para este trabalho.

A todos os professores do curso por todo o conhecimento compartilhado.

Enfim, agradeço a todos que de alguma forma contribuíram ao longo dessa caminhada.

Resumo da Dissertação apresentada ao PPGMMC/CI/UFPB como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

APERFEIÇOAMENTO DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS BASEADO NA FUNÇÃO DE VEROSSIMILHANÇA PERFILADA

Emília Gonçalves de Lima Neta

Fevereiro/2021

Orientadores: Claudio Javier Tablada
Renilma Pereira da Silva

Programa: Modelagem Matemática e Computacional

Na área de Estatística, uma das formas de realizar inferência sobre os parâmetros de um determinado modelo probabilístico é por meio de teste de hipóteses. Porém, muitas vezes, é conveniente realizar o estudo inferencial apenas para um subconjunto desses parâmetros, os quais são denominados de parâmetros de interesse e os demais, de perturbação. Inferências para parâmetros de interesse podem ser feitas utilizando a função de verossimilhança perfilada. Porém, o uso desta função pode conduzir a resultados imprecisos quando o número de parâmetros de perturbação é grande em relação ao tamanho da amostra. Além disso, a função de verossimilhança perfilada não é uma função de verossimilhança genuína. Assim, algumas propriedades básicas de uma função de verossimilhança podem não ser válidas. Com o intuito de atenuar esses problemas, Barndorff-Nielsen (1983) e Severini (1998) propuseram versões ajustadas da função de verossimilhança perfilada. É conhecido da literatura que a estatística da razão de verossimilhanças, sob hipótese nula, tem distribuição assintótica qui-quadrado. Portanto, para amostras de tamanho pequeno ou moderado, a aproximação da distribuição assintótica nula pela distribuição nula exata pode não ser satisfatória. Visando conferir inferências baseadas em amostras de tamanho pequeno ou moderado mais confiáveis, Sousa (2020) propôs um método de aperfeiçoamento do teste da razão de verossimilhanças que consiste em fazer uma correção na cauda da distribuição nula assintótica através da distribuição qui-quadrado inf. Neste trabalho, o principal objetivo é comparar o desempenho dos testes baseados na estatística da razão de verossimilhanças (considerando a função de verossimilhança pefilada e versões modificadas) com o método de aperfeiçoamento proposto por Sousa (2020) em amostras finitas. Também serão incluídas na comparação testes

corrigidos pela técnica de reamostragem bootstrap. Especificamente, essa comparação será feita aplicando as diferentes abordagens às distribuições Lindley Ponderada e Weibull Exponencializada. Para isso, serão realizadas simulações de Monte Carlo, considerando diferentes cenários. Por fim, realizamos exemplos numéricos com base em conjuntos de dados reais.

Abstract of Dissertation presented to PPGMMC/CI/UFPB as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

IMPROVEMENT OF THE LIKELIHOOD RATIO TEST BASED ON THE PROFILED LIKELIHOOD FUNCTION

Emília Gonçalves de Lima Neta

February/2021

Advisors: Claudio Javier Tablada
Renilma Pereira da Silva

Program: Computational Mathematical Modelling

In Statistics, hypothesis tests are used to make inferences of the parameters of a given probabilistic model. However, it is often convenient to perform inferential study only for a subset of parameters, which are called parameters of interest, and the others, nuisance parameters. Inferences of parameters of interest can be made based on profiled likelihood function. However, making inferences based on this function can lead to inaccurate results when the number of nuisance parameters is large comparing to the sample size. Also, the profiled likelihood function is not genuine. Thus, some basic properties of the likelihood function may not be valid. To mitigate these problems, Barndorff-Nielsen (1983) and Severini (1998) proposed adjusted versions of the profiled likelihood function. It is known from the literature that the likelihood ratio statistic, under the null hypothesis, has an asymptotic chi-square distribution. Therefore, for small or moderate samples, the asymptotic distribution is not a good approximation for the exact null distribution. To improve inferences, Sousa (2020) proposed a method for improving the likelihood ratio test, which consists of correcting the tail of the asymptotic null distribution through the chi-square inf. The main aim of this work is to compare the performance of tests based on the likelihood ratio statistic (considering the profile likelihood function and the modified versions) with the improvement method proposed by Sousa (2020) in finite samples. Tests corrected by the bootstrap resampling technique will also be included in the comparison. Specifically, this comparison will be made by applying the different approaches to the Weighted Lindley and Exponentialized Weibull

distributions. For that, Monte Carlo simulations will be performed, considering different scenarios. Finally, we performed four numerical examples based on real data sets.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

$(\Omega, \mathcal{F}, P)$	Espaço de probabilidade, p. 4
P	Probabilidade, p. 4
Ω	Espaço amostral, p. 4
$\mathcal{F}$	Álgebra de eventos, p. 4
$K(\cdot)$	Matriz de informação esperada, p. 9
$L(\cdot)$	Função de verossimilhança, p. 9
$\ell(\cdot)$	Função de log-verossimilhança, p. 9
$\mathbf{U}(\cdot)$	Função score, p. 9
H_0	Hipótese nula, p. 10
H_1	Hipótese alternativa, p. 10
$J(\cdot)$	Matriz de informação observada, p. 10
Θ	Espaço paramétrico, p. 10
α	Probabilidade de erro tipo I, p. 10
β	Probabilidade de erro tipo II, p. 10
$\pi(\cdot)$	Função poder, p. 11
$\mathbf{C}$	Região crítica referente ao quantil da verdadeira distribuição de $\lambda(x)$, p. 11
C^*	Região crítica referente ao quantil q, p. 12
W	Estatística da razão de verossimilhanças, p. 12
γ	Nível de significância, p. 12
$\hat{\theta}$	Estimador de máxima verossimilhança de θ, p. 12

$\tilde{\theta}$	Estimador de máxima verossimilhança restrito, p. 12
q	Quantil, p. 12
$L_p(\cdot)$	Função de verosimilhança perfilada, p. 14
$\ell_p(\cdot)$	Função de log-verosimilhança perfilada, p. 14
W_p	Estatística da razão de verossimilhança baseada na verossimilhança perfilada, p. 14
$\ell_{\mathrm{BNS}}(\cdot)$	Função de log-verossimilhança a aproximação de Severini, p. 15
$\ell_{\mathrm{BN}}(\cdot)$	Função de log-verossimilhança da versão modificada Barndorff-Nielsen, p. 15
W_{BNS^*}	Estatística para a aproximação de Severini baseada em covariâncias empíricas., p. 16
W_{BNS}	Estatística para a aproximação de Severini baseada em covariâncias populacionais., p. 16
$\lambda_p(x)$	Estatística da razão de verossimilhanças baseada na verossimilhança perfilada, p. 17
$\mathbf{C}_p$	Região crítica do teste baseada na verossimilhança perfilada, p. 17
W_p^*	Teste da razão de verossimilhanças aperfeiçoado, proposto por Sousa (2020), p. 18
C_p^*	Região crítica do teste aperfeiçoado, p. 18
W_b	Teste da razão de verossimilhanças corrigido por bootstrap., p. 19
W_{bb}	Estatística da razão de verossimilhanças corrigida Bootstrap-Bartlett, p. 19
$\ell_{\mathrm{BNS}^*}(\cdot)$	Função de log-verossimilhança a aproximação de Severini, p. 32

Lista de Abreviaturas

f.d.a	Função de distribuição acumulada, p. 5
v.a	Variável aleatória, p. 5
f.d.p	Função densidade, p. 8
i.i.d	Independentes e identicamente distribuídas, p. 8
EMV	Estimador de máxima verossimilhança, p. 9
TRV	Teste da razão de verossimilhanças, p. 11

Capítulo 1

Introdução

A Inferência Estatística trata de técnicas que objetivam inferir sobre as características de uma população ou fenômeno por meio de subconjuntos da população, denominados amostras. Para realizar tais inferências é necessário obter um modelo que melhor represente a população de interesse. Nestes modelos, são indexadas quantidades fixas e comumente desconhecidas denotadas de parâmetros.

Podemos realizar inferência sobre os parâmetros por meio da estimação pontual, da estimação intervalar e dos testes de hipóteses. Esse trabalho tem como foco o estudo do aperfeiçoamento do teste da razão de verossimilhanças (TRV) baseado na função de verossimilhança perfilada.

Em diversas situações, é comum realizar inferência em apenas alguns parâmetros de um modelo, estes são denominados parâmetros de interesse e os demais, de perturbação. Nestes casos, podemos fazer inferência utilizando a função de verossimilhança perfilada, que consiste em substituir na verossimilhança usual os parâmetros de perturbação por suas estimativas de máxima verossimilhança para valores fixos dos parâmetros de interesse.

Por se tratar de uma função que depende apenas dos parâmetros de interesse, a verossimilhança perfilada possui menor custo computacional quando comparada à função de verossimilhança usual. No entanto, se a quantidade de parâmetros de perturbação é grande em relação ao tamanho da amostra, a aproximação da distribuição da estatística da razão de verossimilhanças, sob a hipótese nula, pela distribuição qui-quadrado pode ser bastante pobre. Além disso, a verossimilhança perfilada não é uma função de verossimilhança genuína, pois algumas de suas propriedades não são válidas, tais como a função escore não ter média zero e a informação ser viesada (SEVERINI, 1998).

Visando atenuar essas limitações da verossimilhança perfilada, diversos ajustes foram propostos na literatura. Barndorff-Nielsen (1983) propôs uma versão modificada que é obtida como aproximação a uma verossimilhança marginal ou condicional. Porém, a versão do autor faz uso de uma estatística ancilar que normalmente é difícil

de ser obtida.

Devido a dificuldade em determinar a estatística ancilar, Severini (1998) propôs aproximações para a versão modificada de Barndorff-Nielsen (1983). Para casos em que o valor esperado da distribuição de probabilidade é fácil de ser obtido, é utilizada uma aproximação baseada em matriz de covariâncias de derivadas da função de log-verossimilhança. Em caso contrário, é usada uma aproximação por matriz de covariâncias empíricas. Para um estudo mais aprofundado consultar Severini (2000).

A estatística da razão de verossimilhanças (baseada na função de verossimilhança perfilada ou usual), sob a hipótese nula, tem distribuição assintótica qui-quadrado. No entanto, para amostras de tamanho pequeno ou moderado a distribuição assintótica pode não conferir uma boa aproximação para a distribuição exata. Com intuito de melhorar a inferência baseada em amostras de tamanho pequeno ou moderado, Sousa (2020) propôs um método de aperfeiçoamento do teste da razão de verossimilhanças que consiste em fazer uma correção na cauda da distribuição nula assintótica por meio da distribuição qui-quadrado inf.

O objetivo deste trabalho é comparar o desempenho dos testes baseados na função de verossimilhança perfilada e suas versões modificadas com o método do aperfeiçoamento proposto por Sousa (2020) em amostras finitas. Para efeito de comparação, também consideramos as correções de bootstrap paramétrico e bootstrap Bartlett. Iremos aplicar as abordagens descritas nas distribuições Lindley Ponderada e Weibull Exponencializada, simulando diferentes cenários de Monte Carlo por meio da linguagem R Core Team (2019).

Este trabalho está organizado em seis capítulos. No Capítulo 2 são apresentados alguns conceitos básicos de probabilidade e inferência, para situar o leitor acerca dos assuntos que serão utilizados no desenvolvimento do trabalho.

O Capítulo 3 foi destinado para os testes da razão de verossimilhanças baseados na função de verossimilhança perfilada e suas versões modificadas. Na Seção 3.1 abordamos o conceito de verossimilhança perfilada. Em seguida, na Seção 3.2 apresentamos a versão modificada de Barndorff-Nielsen (1983) e suas aproximações propostas por Severini (1998). Na Seção 3.3 dispomos do aperfeiçoamento do teste da razão de verossimilhanças proposto por Sousa (2020). Por fim, na Seção 3.4 abordamos o teste da razão de verossimilhanças corrigido via bootstrap paramétrico e Bartlett bootstrap.

No Capítulo 4 derivamos as funções de verossimilhança para parâmetro de interesse nas distribuições Lindley Ponderada e Weibull Exponencializada. Na Seção 4.1 realizamos um estudo sobre a distribuição Lindley Ponderada, no qual mostramos que esta pertence à família exponencial. Além disso, a partir das versões modificadas apresentadas no Capítulo 3, derivamos ajustes baseados em covariâncias populacionais para o parâmetro de interesse da Lindley Ponderada. Na Seção

4.2 apresentamos um estudo referente à distribuição Weibull Exponencializada, no qual derivamos os ajustes baseados em covariâncias empíricas para o parâmetro de interesse das versões modificadas (Capítulo 3).

No Capítulo 5 expomos os resultados numéricos das simulações e exemplos aplicados para a distribuição Lindley Ponderada e para a distribuição Weibull Exponencializada. Para cada uma das distribuições serão simulados dois cenários, a fim de comparar as abordagens de aperfeiçoamento de teste apresentadas no Capítulo 3. Finalmente, no Capítulo 6 discutimos as considerações finais deste trabalho.

Capítulo 2

Conceitos Preliminares

2.1 Conceitos Básicos de Probabilidade

A teoria da probabilidade consiste em estudar fenômenos do cotidiano por meio de experimentos aleatórios. Um experimento é dito aleatório quando for possível realizá-lo várias vezes sob as mesmas condições, porém os resultados obtidos podem ser diferentes. Além disso, deve ser possível descrever o conjunto de todas as possibilidades do experimento. Por fim, se o experimento é repetido um grande número de vezes, surgirá uma configuração definida ou uma regularidade.

Os resultados de um experimento aleatório são caracterizados por três componentes:

- Espaço amostral: conjunto de todos os possíveis resultados, denotado por Ω;
- Coleção de conjuntos de interesse: é formada por subconjuntos de Ω e intitulada de σ-álgebra de eventos [1], denotada por $\mathcal{F}$;
- Probabilidade: é um valor numérico entre 0 e 1, que representa a chance de ocorrência de cada um dos conjuntos de resultados de interesse, é denotada por P.

A terna $(\Omega, \mathcal{F}, P)$ é denominada espaço de probabilidade, em que a função $P : \mathcal{F} \to R$ é uma medida de probabilidade.

Na realização de experimentos aleatórios não sabemos seu resultado, porém podemos estabelecer um espaço de probabilidade adequado para analisar a probabilidade

[1] Definição: Uma σ-álgebra de eventos $\mathcal{F}$ é uma coleção de subconjuntos do espaço amostral Ω quando satisfaz as condições:

1. $\mathcal{F} \neq \emptyset$;
2. Se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$;
3. Se $\{A\}_{j \geq 1}$ é uma sequência de eventos em $\mathcal{F}$, então $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.

de qualquer evento de interesse da σ-álgebra. A partir disso, obtemos o conceito de variável aleatória.

Definição 2.1. (Variável aleatória) *Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. Denominamos de variável aleatória (v.a), qualquer função $X : \Omega \to \mathbb{R}$ tal que*

$$X^{-1}(I) = \{w \in \Omega : X(w) \in I\} \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$. Em palavras, X é tal que sua imagem inversa de intervalos $I \subset \mathbb{R}$ pertencem a σ-álgebra $\mathcal{F}$.

De modo geral, uma v.a pode ser entendida como uma variável quantitativa, no qual os resultados dependem de fatores aleatórios. Formalmente, é uma função real definida no espaço amostral Ω de um experimento aleatório.

Para caracterizar a estrutura probabilística de uma v.a, precisamos entender algumas definições, tais como, função de distribuição acumulada, função de probabilidade e função densidade.

Definição 2.2. (Função de distribuição acumulada) *Sendo X uma v.a em $(\Omega, \mathcal{F}, P)$, sua função de distribuição acumulada (f.d.a) é definida por*

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x),$$

com x percorrendo todos os reais.

Proposição 2.1. (Propriedades da f.d.a) *Uma f.d.a de uma v.a X em $(\Omega, \mathcal{F}, P)$ obedece as seguintes propriedades:*
(F1) $\lim_{x\to-\infty} F(x) = 0$ *e* $\lim_{x\to\infty} F(x) = 1$*;*
(F2) *F é contínua à direita;*
(F3) *F é não decrescente, isto é, $F(x) \leq F(y)$ sempre que $x \leq y$, $\forall x, y \in \mathbb{R}$.*

Segundo Magalhães (2006), conhecer a função de distribuição acumulada de uma v.a nos permite obter qualquer informação sobre a variável. Além disso, mesmo que a variável assuma valores apenas em um subconjunto dos reais, a f.d.a é definida em toda a reta.

Definição 2.3. (v.a discreta e função de probabibilidade)[2] *A função de probabilidade de uma v.a discreta é uma função que atribui probabilidade a cada um dos possíveis valores assumidos pela variável. Isto é, sendo X uma variável com valores $x_1, x_2, \ldots,$ temos para $i = 1, 2, \ldots,$*

$$p(x_i) = P(X = x_i) = P(\{w \in \Omega : X(w) = x_i\}).$$

[2]Uma v.a é considerada discreta quando assume somente um número enumerável de valores.

Proposição 2.2. (Propriedades da função de probabilidade) *A função de probabilidade X em $(\Omega, \mathcal{F}, P)$ satisfaz:*

(FP1) $0 \leq p(x_i) \leq 1, \forall i = 1, 2, \ldots$;

(FP2) $\sum_i p(x_i) = 1$ *com a soma percorrendo todos os possíveis valores.*

Definição 2.4. (v.a contínua e função densidade) *Uma v.a X em $(\Omega, \mathcal{F}, P)$, com f.d.a F, é contínua se existir uma função não negativa f tal que:*

$$F(x) = \int_{-\infty}^{x} f(w)dw, \forall x \in \mathbb{R}$$

A função f é denominada função densidade de probabilidade (f.d.p).

Proposição 2.3. (Propriedades da f.d.p) *A f.d.p de X em $(\Omega, \mathcal{F}, P)$ satisfaz:*

(FD1) $f(x) \geq 0, \forall x \in \mathbb{R}$;

(FD2) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Vimos que para variáveis aleatórias discretas, utilizamos a função de probabilidade. Já para a v.a contínua utilizamos a função densidade. É importante entender que os estudos deste trabalho serão focados em variáveis aleatórias contínuas.

Vale ressaltar que a função densidade caracteriza a v.a contínua. Pode-se notar uma relação entre a função de distribuição acumulada e a função densidade, pois dada a função densidade, obtém-se a função de distribuição acumulada por integração. Por outro lado, derivando a função de distribuição acumulada nos pontos em que ela é derivável, obtém-se a função densidade.

Para realizar simulações em diferentes conjuntos de valores de uma v.a que segue uma determinada distribuição, precisamos gerar números pseudo-aleatórios. Dessa forma, um resultado importante é o conceito de função quantílica, no qual apresentaremos a seguir.

Definição 2.5. *Seja F uma função de distribuição qualquer. A função quantílica da distribuição F, denotada por Q, é definida da seguinte forma:*

$$Q(y) = F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Note que se a inversa de F existe no sentido usual, isto é, F^{-1} possui forma fechada, ela coincide com a função quantílica.

Na realização de experimentos aleatórios, nem sempre se tem apenas uma variável de interesse, comumente envolvem-se várias variáveis. Dessa forma, faz-se necessário trabalhar com duas ou mais variáveis de maneira conjunta, pois facilita a obtenção de resultados. Para isso, veremos a definição de vetor aleatório.

Definição 2.6. (Vetor aleatório) *Um vetor* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ *cujas componentes são v.a's definidas no mesmo espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*, é chamado vetor aleatório ou v.a n-dimensional.*

A função de distribuição $F = F_{\mathbf{X}} = F_{X_1,X_2,\ldots,X_n}$ de um vetor aleatório $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, é definida por:

$$F(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$. F é também denotada por função de distribuição conjunta das variáveis aleatórias $X_1, \ldots, X_n$.

Proposição 2.4. (Propriedades da Função de distribuição conjunta) *Seja* $\boldsymbol{X}$ *um vetor aleatório em* $(\Omega, \mathcal{F}, P)$ *então, para qualquer* $x \in \mathbb{R}^n$, $F(\boldsymbol{x})$ *satisfaz as seguintes propriedades:*

(FC1) $F(\boldsymbol{x})$ *é não decrescente em cada uma de suas coordenadas;*

(FC2) $F(\boldsymbol{x})$ *é contínua à direita em cada uma de suas coordenadas;*

(FC3) *Se para algum* j, $x_j \to -\infty$, *então* $F(x) \to 0$ *e, ainda, se para todo* $j, x_j \to \infty$, *então* $F(\boldsymbol{x}) \to 1$*;*

(FC4) $F(\boldsymbol{x})$ *é tal que, para* $\forall a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $1 \leq i \leq m$, *temos* $P(a_1 < X_1 < b_1, a_2 < X_2 \leq b_2, \ldots, a_n < X_n \leq b_n) \geq 0$.

Definição 2.7. (Variáveis independentes) *As v.a's* $X_1, \ldots, X_n$ *são coletivamente independentes se, e somente se:*

$$F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

para todo $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

Proposição 2.5. *Sejam* $X_1, X_2, \ldots, X_n$ *variáveis aleatórias independentes e identicamente distribuídas (i.i.d), ou seja, além de independentes possuem a mesma função de distribuição* $F(x)$*. As expressões da função de distribuição de* $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ *e* $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ *são dadas, respectivamente, por:*

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - [1 - F(x)]^n$$

e

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = [F(x)]^n.$$

Definição 2.8. (Esperança Matemática) *Seja* X *uma v.a contínua com f.d.p f. De-*

finimos a esperança matemática ou valor esperado de X por

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x,$$

desde que a integral esteja bem definida.

Definição 2.9. (Variância) *Seja X uma v.a contínua, então definimos a variância da v.a X, por*

$$Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2.$$

2.2 Conceitos Básicos de Inferência Estatística

A Inferência Estatística consiste em técnicas que objetivam inferir sobre as características de uma população ou fenômeno, por meio de informações contidas na amostra, subconjunto da população. É importante ressaltar que o processo de inferência não é exato, pois são feitas conclusões a partir de dados amostrais. Dessa forma, sempre haverá um grau de incerteza sobre as conclusões.

Para realizar inferência, devemos considerar um modelo que represente o fenômeno ou população de interesse. Para Cordeiro e Cribari-Neto (2014), um modelo é uma representação simples de uma realidade mais abrangente. Além disso, um bom modelo deve reter características importantes do fenômeno representado.

Segundo Cordeiro e Cribari-Neto (2014), um modelo estatístico representa um fenômeno que ocorre de maneira incerta. No modelo são indexadas quantidades fixas e comumente desconhecidas, denominadas de parâmetros. Dessa forma, a inferência estatística é realizada nesses parâmetros, assim como no modelo e, consequentemente, no fenômeno de interesse.

Podemos realizar inferência sobre os parâmetros por meio da estimação pontual, da estimação intervalar e dos testes de hipóteses. A estimação consiste em utilizar estatísticas amostrais para estimar os parâmetros populacionais desconhecidos. Já uma hipótese estatística é uma suposição sobre algum parâmetro populacional desconhecido, assim, a suposição pode ser verdadeira ou falsa. Por fim, os testes de hipóteses se referem aos procedimentos estatísticos para aceitar ou rejeitar uma hipótese.

Existem diversos métodos para realizar estimação pontual de parâmetros desconhecidos, porém o mais utilizado é o método de máxima verossimilhança.

Definição 2.10. *Uma sequência $X_1, \ldots, X_n$ de v.a's independentes e identicamente distribuídas (i.i.d) com f.d.p $f(x|\boldsymbol{\theta})$ com $\boldsymbol{\theta} \in \Theta$ [3], é dita ser uma amostra aleatória*

[3] O conjunto Θ em que $\boldsymbol{\theta}$ toma valores é denominado espaço paramétrico.

de tamanho n da distribuição de X e neste caso, temos:

$$f(x_1, \ldots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) = f(x_1|\boldsymbol{\theta}) \ldots f(x_n|\boldsymbol{\theta}).$$

A função densidade conjunta obtida como função de $\boldsymbol{\theta}$ é denominada função de verossimilhança de $\boldsymbol{\theta}$, correspondente à amostra observada $\mathbf{x} = (x_1, \ldots, x_n)^\top$ e será denotada:

$$L(\boldsymbol{\theta}; x) = L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}). \tag{2.1}$$

O logaritmo natural da função de verossimilhança de $\boldsymbol{\theta}$ é denominado de log-verossimilhança e é denotado por $\ell(\boldsymbol{\theta}; x) = \ell(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$.

Definição 2.11. *O estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ é o valor de $\hat{\boldsymbol{\theta}} \in \Theta$ que maximiza $L(\boldsymbol{\theta})$(se $\hat{\boldsymbol{\theta}}$ existir). Seu valor observado é a estimativa de máxima verossimilhança.*

É importante entender que o EMV também maximiza a função log-verossimilhança, pois a função logaritmo é estritamente crescente, ou seja, o valor de $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta})$ é o mesmo que maximiza $\ell(\boldsymbol{\theta})$.

As estimativas de máxima verossimilhança gozam de diversas propriedades significativas, a saber, consistência, invariância e eficiência assintótica.

A primeira derivada da função log-verossimilhança é chamada função (ou vetor) escore

$$\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Definição 2.12. *A matriz de informação de Fisher, ou matriz de informação esperada, para $\boldsymbol{\theta} \in \mathbb{R}^p$ é a matriz $p \times p$ definida por*

$$K(\boldsymbol{\theta}) = E\{\mathbf{U}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})^\top\}.$$

Se certas condições de regularidade (CORDEIRO, 1999) são satisfeitas, temos

$$E[\mathbf{U}(\boldsymbol{\theta})] = 0 \quad \text{e} \quad K(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right].$$

Além disso, temos que a matriz de primeiras derivadas da função escore com sinal negativo é denominada *matriz de informação observada*, da forma

$$J(\boldsymbol{\theta}) = -\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right].$$

Em inferência, quando resumimos a informação que os dados contêm sobre $\boldsymbol{\theta}$, usamos uma estatística, sendo interessante que não haja perda de informações sobre $\boldsymbol{\theta}$. Dessa forma, a estatística precisa conter toda a informação possível sobre $\boldsymbol{\theta}$ presente na amostra. De modo geral, se tivermos uma estatística em que podemos extrair toda informação que a amostra possui sobre $\boldsymbol{\theta}$, então dizemos que essa estatística é suficiente para $\boldsymbol{\theta}$.

Definição 2.13. *Seja $X_1, \ldots, X_n$ uma amostra aleatória $\boldsymbol{X}$ com f.d.p $f(x|\boldsymbol{\theta})$. Dizemos que a estatística $T = T(X_1, \ldots, X_n)$ é suficiente para $\boldsymbol{\theta}$, quando a distribuição condicional de $X_1, \ldots, X_n$ dado T for independente de $\boldsymbol{\theta}$.*

2.2.1 Teste de Hipóteses

Na inferência, qualquer afirmação que se estabeleça sobre um parâmetro desconhecido, é denotada de hipótese estatística. O principal objetivo é utilizar informações sobre a amostra para definir uma regra de decisão em que rejeitamos ou não rejeitamos a hipótese proposta. Esta regra de decisão é chamada de teste. Em um teste formulam-se duas hipóteses complementares, em que são chamadas de hipótese nula e hipótese alternativa, denotadas por H_0 e H_1, respectivamente.

Segundo Bolfarine e Sandoval (2001), dizemos que a distribuição de uma v.a X está totalmente especificada quando conhecemos sua função densidade $f(x|\boldsymbol{\theta})$ e $\boldsymbol{\theta}$, e está parcialmente especificada quando conhecemos a função densidade $f(x|\boldsymbol{\theta})$, porém $\boldsymbol{\theta}$ é desconhecido.

De modo geral, se temos uma amostra aleatória $X_1, \ldots, X_n$ extraída de uma distribuição que envolve um parâmetro $\boldsymbol{\theta}$ desconhecido, no qual está estabelecido em um espaço paramétrico Θ, então as hipóteses podem ser definidas como $H_0 : \boldsymbol{\theta} \in \Theta_0$ e $H_1 : \boldsymbol{\theta} \in \Theta_1$, em que Θ_0 e Θ_1 formam uma partição de Θ. Portanto, um teste é determinado particionando o espaço amostral em dois subconjuntos. Um destes relacionam os valores observados de $\mathbf{X}$ para os quais H_0 será rejeitada, denominada região crítica do teste. Por fim, o outro envolve os valores observados de $\mathbf{X}$ para os quais H_0 não será rejeitada, denotada região de aceitação do teste.

Quando $\dim(\Theta_0) = 1$ dizemos que H_0 é simples. Caso contrário, ou seja, $\dim(\Theta_0) > 1$ dizemos que H_0 é composta. Da mesma forma acontece com H_1.

Na decisão de rejeitar ou não rejeitar H_0, há risco de cometermos erros. Denotaremos, γ e β como probabilidades dos erros tipo I e II, respectivamente. Além disso, γ é também conhecido como nível de significância ou nível nominal, e representa o quanto estamos dispostos a cometer o erro do tipo I.

O erro tipo I consiste em rejeitar H_0, sendo H_0 verdadeira. Já o erro tipo II acontece quando não rejeitamos H_0, sendo H_0 falsa. Na Tabela 2.1 é apresentada as possíveis consequências na realização de um teste de hipótese.

Tabela 2.1: Tipos de erros em teste de hipóteses.

Decisão	H_0 é verdadeira	H_0 é falsa
Não rejeitar H_0	Decisão correta	Erro do tipo II
Rejeitar H_0	Erro do tipo I	Decisão correta

Suponha que $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\nu}^\top)^\top$, onde $\boldsymbol{\xi}$ e $\boldsymbol{\nu}$ são vetores de parâmetros de interesse e de perturbação, respectivamente. Neste trabalho estamos interessados no estudo de hipóteses compostas bilaterais, do tipo $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi_0}$ *versus* $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi_0}$.

Definição 2.14. (Função poder) *A função $\pi(\boldsymbol{\theta})$ é chamada função de poder do teste. Denotando por $\boldsymbol{C}$ a região crítica, a função de poder é definida como:*

$$\pi(\boldsymbol{\theta}) = P(X \in \boldsymbol{C} \,|\, \boldsymbol{\theta}),\ \forall \boldsymbol{\theta} \in \Theta.$$

Segundo Bolfarine e Sandoval (2001), um procedimento que produz testes razoáveis e pode ser utilizado em muitos casos, sem dificuldade, é o Teste da Razão de Verossimilhanças (TRV).

Definição 2.15. *O* TRV *para as hipóteses $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_1$, pode ser definido como o teste com região crítica dada por*

$$\boldsymbol{C} = \left\{ \boldsymbol{x}; \lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_0} L(\boldsymbol{\theta}; \boldsymbol{x})}{\sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}; \boldsymbol{x})} \leq c \right\}.$$

Vale ressaltar que $0 \leq \lambda(\mathbf{x}) \leq 1$, pois o numerador é o supremo da função de verossimilhança quando $\boldsymbol{\theta} \in \Theta_0$, enquanto o denominador é o supremo sobre todo o espaço paramétrico Θ. Dessa forma, se H_0 é verdadeira, é esperado que $\lambda(\mathbf{x})$ esteja próximo de 1, caso contrário, esperamos que o denominador seja grande em relação ao numerador, assim, $\lambda(\mathbf{x})$ deve ser próximo de 0.

Portanto, ao nível de significância γ é razoável rejeitar H_0 se $\lambda(\mathbf{x}) \leq c$, no qual a constante c pode ser obtida da forma:

$$\gamma = \sup_{\boldsymbol{\theta}\in\Theta_0} P(\lambda(\mathbf{x}) \leq c).$$

No entanto, para isso, é necessário conhecer a verdadeira distribuição da estatística $\lambda(\mathbf{x})$, que normalmente não é obtida de forma trivial. A seguir, será apresentada a distribuição assintótica da estatística do TRV, resolvendo esse problema para o caso em que se tem grandes amostras.

Teorema 2.1. *Seja $X_1, \ldots, X_n$ uma amostra aleatória da v.a X com f.d.p $f(x|\boldsymbol{\theta})$. Sob as condições de regularidade, se $\boldsymbol{\theta} \in \Theta_0$, então a distribuição da estatística $W = -2\log(\lambda(\boldsymbol{x}))$ converge para a distribuição qui-quadrado quando o tamanho da*

amostra n tende ao infinito. O número de graus de liberdade da distribuição limite é a diferença entre o número de parâmetros não especificados em Θ e o número de parâmetros não especificados em Θ_0.

Para facilitar a compreensão, considere que $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são estimadores de máxima verossimilhança sob todo espaço paramétrico e a hipótese nula, respectivamente. Então, teremos:

$$W = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\}.$$

Dessa forma, por meio do TRV, a região crítica para testar as hipóteses $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_1$, pode ser definida como:

$$C^* = \{x; W > q\}.$$

Portanto, ao nível de significância γ é plausível rejeitar H_0 se $W > q$, no qual a constante q pode ser obtida da forma:

$$\gamma = P(W > q).$$

Assim, fixando uma probabilidade γ, o número q pode ser igualado ao quantil $1-\gamma$ obtido a partir da distribuição qui-quadrado ($\mathcal{X}^2_k$), em que k é o número de restrições impostas na hipótese nula (H_0).

Vale ressaltar que este resultado é útil para valores de n (tamanhos amostrais) elevados. Dessa forma, entende-se que usar a distribuição qui-quadrado como uma aproximação à verdadeira distribuição W pode acarretar em inferências imprecisas quando o tamanho da amostra for pequeno ou moderado.

Para um estudo mais aprofundado dos conceitos apresentados nas seções 2.1 e 2.2, veja as referências Bolfarine e Sandoval (2001), James (1996) e Magalhães (2006).

Capítulo 3

Testes de Hipóteses Baseados em Verossimilhança Perfilada e suas Versões Modificadas

3.1 Função de Verossimilhança Perfilada

Em várias situações, desejamos realizar inferência em apenas alguns parâmetros de um modelo, esses parâmetros são denominados de parâmetros de interesse e os demais, parâmetros de perturbação. Em modelos que envolvem parâmetros de perturbação, podemos realizar inferência por meio de uma pseudo-verossimilhança, que são funções da verossimilhança que dependem apenas dos parâmetros de interesse. Além disso, as pseudo-verossimilhanças têm menor custo computacional, pois possuem uma menor quantidade de parâmetros a serem estimados. Nesse trabalho iremos utilizar a pseudo-verossimilhança conhecida como verossimilhança perfilada.

O conceito da verossimilhança perfilada consiste em substituir os parâmetros de perturbação por suas estimativas de máxima verossimilhança na verossimilhança usual, isto para cada valor fixo do parâmetro de interesse.

É importante ressaltar que usar a verossimilhança perfilada na presença de um grande número de parâmetros de perturbação em relação ao tamanho amostral, pode prejudicar a qualidade das inferências que baseiam-se em resultados assintóticos, uma vez que a verossimilhança perfilada poderá ter curvatura diferente da verossimilhança genuína. Isso pode prejudicar a estimação da estrutura de covariâncias dos EMV's.

Seja $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\nu}^\top)^\top$ [1] um vetor de parâmetros que indexa uma distribuição de probabilidade e $L(\boldsymbol{\theta})$ (definida em (2.1)) a função de verossimilhança associada a essa distribuição, em que $\boldsymbol{\xi}$ e $\boldsymbol{\nu}$ são parâmetros de interesse e de perturbação,

[1] $\boldsymbol{\xi}$ e $\boldsymbol{\nu}$ podem ser vetores ou escalares.

respectivamente. Então, a função de verossimilhança perfilada é dada por

$$L_p(\boldsymbol{\xi}) = L(\boldsymbol{\xi}, \hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}),$$

em que $\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}$ [2] é o estimador de máxima verossimilhança de $\boldsymbol{\nu}$ para $\boldsymbol{\xi}$ fixo.

Sendo $\ell_p(\boldsymbol{\xi}) = \log[L_p(\boldsymbol{\xi})]$, então a estimativa de máxima verossimilhança perfilada de $\boldsymbol{\xi}$, $\hat{\boldsymbol{\xi}}_{\boldsymbol{p}}$, é obtido maximizando $\ell_p(\boldsymbol{\xi})$. Dessa forma, podemos definir a estatística da razão de verossimilhanças com base na função de verossimilhança perfilada para testar a hipótese $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi_0}$ *versus* $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi_0}$, por

$$W_p = 2\{\ell_p(\hat{\boldsymbol{\xi}}_{\boldsymbol{p}}) - \ell_p(\boldsymbol{\xi_0})\}. \tag{3.1}$$

Sob as condições de regularidade e sob a hipótese nula, a distribuição da estatística W_p converge para a distribuição qui-quadrado, com k graus de liberdade ($\mathcal{X}^2_k$), em que k é a dimensão do vetor $\boldsymbol{\xi_0}$, quando $n \longrightarrow \infty$.

Vale ressaltar que a função de verossimilhança perfilada $L_p(\boldsymbol{\xi})$ não é uma função de verossimilhança genuína, pois estamos tratando $\boldsymbol{\nu}$ como se fosse um parâmetro conhecido e igual a $\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}$. Em particular, segundo Cordeiro (1999) algumas das propriedades da função de verossimilhança não são válidas para a função de verossimilhança perfilada, tais como:

- A função escore não necessariamente tem valor esperado zero e a informação pode ser viesada;
- A aproximação da distribuição nula da estatística da razão de verossimilhanças, baseada na função de verossimilhança perfilada, pela distribuição qui-quadrado pode não ser satisfatória.

Em contrapartida, segundo Cordeiro (1999) a função de verossimilhança perfilada possui algumas propriedades interessantes, tais como:

- $\hat{\boldsymbol{\xi}}_{\boldsymbol{p}} = \hat{\boldsymbol{\xi}}$, em que $\hat{\boldsymbol{\xi}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\xi}$;
- Ao testar hipóteses sobre $\boldsymbol{\xi}$, a estatística da razão de verossimilhanças baseada em $\ell_p(\boldsymbol{\xi})$ é equivalente à baseada em $\ell(\boldsymbol{\xi}, \boldsymbol{\nu})$ (função log-verossimilhança usual).

[2] Para $\boldsymbol{\xi}$ fixo, $\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}$ é o valor que maximiza $L(\boldsymbol{\xi}, \boldsymbol{\nu})$ e pode ser obtido fazendo $\dfrac{\partial \ell(\boldsymbol{\xi}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} = 0$.

3.2 Versões Modificadas e o Teste da Razão de Verossimilhanças

Sabemos que a função de verossimilhança perfilada não pode ser considerada uma função de verossimilhança genuína, uma vez que, algumas de suas propriedades não são válidas. Tendo em vista essas limitações, na literatura foram propostas diversas modificações para a função de verossimilhança perfilada.

Cox e Reid (1987) propuseram um ajuste à função de verossimilhança perfilada que requer uma parametrização ortogonal. Há também uma aproximação para o ajuste de Cox e Reid para o caso em que o parâmetro de interesse é escalar e não necessariamente ortogonal aos parâmetros de perturbação, disponível em Cox e Reid (1993). Outras modificações também foram sugeridas por Barndorff-Nielsen (1995) e Severini (1998). Um estudo mais detalhado sobre as versões modificadas para a função de verossimilhança perfilada pode ser encontrado em Severini (2000).

3.2.1 Barndorff-Nielsen (1983)

Uma modificação para a função de verossimilhança perfilada foi proposta por Barndorff-Nielsen (1983), sendo obtida como uma aproximação a uma verossimilhança marginal ou condicional para $\boldsymbol{\xi}$, se houver. Neste caso, a versão modificada de ℓ_p é dada por

$$\ell_{\mathrm{BN}}(\boldsymbol{\xi}) = \ell_p(\boldsymbol{\xi}) + \frac{1}{2}\log|j_{\boldsymbol{\nu}\boldsymbol{\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}, \boldsymbol{\xi})| - \log|\ell_{\boldsymbol{\nu},\hat{\boldsymbol{\nu}}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}, \boldsymbol{\xi}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\xi}}, a)|,$$

em que $j_{\boldsymbol{\nu}\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi})$ é a informação observada em relação a $\boldsymbol{\nu}$, e $\hat{\boldsymbol{\nu}}$ e $\hat{\boldsymbol{\xi}}$ as estimativas de máxima verossimilhança para $\boldsymbol{\nu}$ e $\boldsymbol{\xi}$, respectivamente. Além disso,

$$\ell_{\boldsymbol{\nu},\hat{\boldsymbol{\nu}}}(\boldsymbol{\nu}, \boldsymbol{\xi}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\xi}}, a) = \frac{\partial}{\partial \hat{\boldsymbol{\nu}}}\left[\frac{\partial \ell(\boldsymbol{\nu}, \boldsymbol{\xi}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\xi}}, a)}{\partial \boldsymbol{\nu}}\right]$$

é uma derivada do espaço amostral da função de log-verossimilhança e a é uma estatística ancilar [3] (SEVERINI, 2000).

No entanto, essa modificação pode se tornar difícil de ser calculada, devido a especificação da estatística ancilar, que em muitos casos, pode ser inviável de ser obtida.

3.2.2 Aproximação Baseada em Covariâncias Populacionais

Devido a dificuldade em determinar a estatística ancilar, iremos considerar uma aproximação para a modificação de Barndorff-Nielsen (1983) que foi proposta por

[3] Uma estatística $S(\cdot)$ diz-se ser ancilar se a sua distribuição não depender do parâmetro.

Severini (1998), na qual é definida como:

$$\ell_{\text{BNS}}(\boldsymbol{\xi}) = \ell_p(\boldsymbol{\xi}) + \frac{1}{2}\log|j_{\boldsymbol{\nu}\boldsymbol{\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}, \boldsymbol{\xi})| - \log|I_{\boldsymbol{\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\xi}}, \boldsymbol{\xi}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\xi}})|, \tag{3.2}$$

em que

$$I_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi}; \boldsymbol{\nu_0}, \boldsymbol{\xi_0}) = E_{(\boldsymbol{\nu_0}, \boldsymbol{\xi_0})}[\ell_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi})\ell_{\boldsymbol{\nu}}(\boldsymbol{\nu_0}, \boldsymbol{\xi_0})^\top] \tag{3.3}$$

e

$$\ell_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi}) = \frac{\partial \ell(\boldsymbol{\nu}, \boldsymbol{\xi})}{\partial \boldsymbol{\nu}}.$$

Observe que $\ell_{\text{BNS}}(\boldsymbol{\xi})$ tem seu máximo em $\hat{\boldsymbol{\xi}}_{\text{BNS}}$. Dessa forma, a estatística da razão de verossimilhanças baseado nessa modificação para testar $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi_0}$ *versus* $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi_0}$ é dada por

$$W_{\text{BNS}} = 2[\ell_{\text{BNS}}(\hat{\boldsymbol{\xi}}_{\text{BNS}}) - \ell_{\text{BNS}}(\boldsymbol{\xi_0})].$$

3.2.3 Aproximação Baseada em Covariâncias Empíricas

Severini (1998) propôs outra alternativa de aproximação para a função de log-verossimilhança modificada de Barndorff-Nielsen (1983), em que o termo $I_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi}; \boldsymbol{\nu_0}, \boldsymbol{\xi_0})$ em (3.3) é substituído por

$$\breve{I}_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{\xi}; \boldsymbol{\nu_0}, \boldsymbol{\xi_0}) = \sum_{j=1}^{n} \ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu}, \boldsymbol{\xi})\ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu_0}, \boldsymbol{\xi_0})^\top, \tag{3.4}$$

em que $\ell_{\boldsymbol{\theta}}^{(j)}(\boldsymbol{\theta}) = (\ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\theta}), \ell_{\boldsymbol{\xi}}^{(j)}(\boldsymbol{\theta}))$ é a função escore para a j-ésima observação.

Denotaremos por ℓ_{BNS^*} a função de log-verossimilhança modificada que faz uso de (3.4). A ℓ_{BNS^*} será útil em casos que há dificuldade em calcular (3.3).

Veja que ℓ_{BNS^*} possui seu máximo em $\hat{\boldsymbol{\xi}}_{\text{BNS}^*}$. Portanto, a estatística da razão de verossimilhanças baseada nessa modificação para testar $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi_0}$ *versus* $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi_0}$ é dada por

$$W_{\text{BNS}^*} = 2[\ell_{\text{BNS}^*}(\hat{\boldsymbol{\xi}}_{\text{BNS}^*}) - \ell_{\text{BNS}^*}(\boldsymbol{\xi_0})].$$

3.3 Aperfeiçoamento do Teste da Razão de Verossimilhanças (SOUSA, 2020)

Seja $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\nu}^\top)^\top$, em que $\boldsymbol{\xi}$ e $\boldsymbol{\nu}$ são os vetores de parâmetros de interesse e perturbação, respectivamente. Para testar hipóteses compostas do tipo $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi_0}$ *versus* $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi_0}$, podemos utilizar o TRV baseado na verossimilhança perfilada. Conforme foi comentado, sob H_0, a estatística W_p, definida em (3.1), tem distribuição assintótica $\mathcal{X}^2_k$, em que k é a dimensão do vetor $\boldsymbol{\xi_0}$. Assim, a região crítica do TRV assintótico é dada por:

$$\mathbf{C}_p = \left\{\mathbf{x}; W_p > q\right\},$$

sendo q o quantil $1-\gamma$ da distribuição χ^2_k. No entanto, se n (tamanho da amostra) é pequeno ou moderado, a distribuição $\mathcal{X}^2_k$ pode não ser uma aproximação satisfatória para a verdadeira distribuição da estatística W_p.

Visando atenuar esses problemas, Sousa (2020) propôs uma correção na cauda da distribuição qui-quadrado por meio da distribuição qui-quadrado inf (proposta por Tablada (2017)).

Uma v.a X^{inf} segue distribuição qui-quadrado inf de parâmetros $k > 0$ e $c > 0$, se sua densidade é dada por

$$g^{\text{inf}}(x|k,c) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \left[1 - \left(1 - \frac{\gamma(\frac{k}{2},\frac{x}{2})}{\Gamma(\frac{k}{2})}\right)^c + c\frac{\gamma(\frac{k}{2},\frac{x}{2})}{\Gamma(\frac{k}{2})}\left(1 - \frac{\gamma(\frac{k}{2},\frac{x}{2})}{\Gamma(\frac{k}{2})}\right)^{c-1}\right],$$

em que, $\gamma(\alpha,t) = \int_0^t s^{\alpha-1}e^{-s}\mathrm{d}s$ a função gama incompleta inferior e $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}\mathrm{d}s$ a função gama.

Seguindo a abordagem discutida em Sousa (2020), com o objetivo de obter um TRV aperfeiçoado utiliza-se a região crítica modificada definida da seguinte forma

$$\mathbf{C}^*_p = \left\{\mathbf{x}; W_p > q_{\text{inf}}\right\},$$

em que q_{inf} é o quantil $1-\gamma$ da distribuição qui-quadrado inf de parâmetros k e $c = 0.5\log(n)$, o qual pode ser obtido ao resolver numericamente a equação

$$\int_0^{q_{\text{inf}}} g^{\text{inf}}(x|k,c)\,\mathrm{d}x = 1-\gamma. \tag{3.5}$$

Essa metodologia de aperfeiçoamento é consistente no sentido que, quando $n \to \infty$, a distribuição qui-quadrado inf de parâmetros k e $c = 0.5\log(n)$ converge para a distribuição χ^2_k. Neste trabalho, W^*_p denotará o TRV aperfeiçoado por

essa abordagem.

Para a obtenção dos resultados referentes a essa abordagem de aperfeiçoamento do TRV, utilizamos o pacote **LikRatioTest** [4], desenvolvido por Sousa (2020) utilizando a linguagem de programação R. Nesse pacote é possível obter o quantil $1-\gamma$ da distribuição qui-quadrado inf que satisfaz a equação (3.5), além de realizar simulações de Monte Carlo para estimar empiricamente os tamanhos dos testes.

Para calcular o quantil da distribuição qui-quadrado inf, utilizamos a função `est_q()` do pacote.

Exemplo: Seja X uma v.a com distribuição qui-quadrado inf. Para exemplificar, fixemos os parâmetros da distribuição em k = 1 e c = 1. Calculamos o quantil 1 - γ, dado $\gamma = 0,10$ tal que

$$P(X > q_{\text{inf}}) = 0,10$$

da forma

```
> # Função densidade da qui-quadrado inf
> fdp_chisq_inf <- function(par, x) {
+   k <- par[1]
+   c <- par[2]
+   dchisq(x = x, df = k) * (1 - (1 - pchisq(q = x, df = k)) ^ c +
+   c * pchisq(q = x, df = k) *(1 - pchisq(q = x, df = k)) ^ (c - 1))
+ }
> # Calculando o quantil por meio da função est_q()
> est_q(fn = fdp_chisq_inf, alpha = 0.10, bilateral = FALSE, c = 1,
+       k = 1)
[1] 3.798
```

3.4 Teste da Razão de Verossimilhanças Corrigidos via Bootstrap (Bootstrap Paramétrico e Bartlett-Bootstrap)

3.4.1 Teste da Razão de Verossimilhanças via Bootstrap Paramétrico

A técnica de reamostragem bootstrap paramétrico, proposta por Efron (1979) será utilizada neste trabalho para encontrar o quantil $1-\gamma$ da distribuição empírica da estatística W_p, sob hipótese nula, a partir da amostra observada $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$.

[4] O pacote pode ser consultado em: https://github.com/prdm0/LikRatioTest.

Para realizar a técnica de bootstrap seguimos os passos:

1. Calcular o valor da estatística W_p para a amostra observada $\mathbf{x}$;

2. Gerar B amostras pseudo-aleatórias $(\boldsymbol{x}^{*1}, \boldsymbol{x}^{*2}, \ldots, \boldsymbol{x}^{*B})$ a partir da distribuição de X, sob a hipótese nula;

3. Para cada pseudo amostra bootstrap, calcular a estatística W_p, ou seja, obtenha W_p^{*b}, com $b = 1, 2, \ldots, B$, da seguinte forma:

$$W_p^{*b} = 2\left[\ell_p(\hat{\boldsymbol{\xi}}_{\boldsymbol{p}}^{*\boldsymbol{b}}; x^{*b}) - \ell_p(\boldsymbol{\xi_0}; x^{*b})\right]$$

   Fixando o nível de significância γ, o percentil $1-\gamma$ é estimado pelo valor $\hat{q}_{1-\gamma}$, tal que

$$\frac{\#\{W_p^{*b} \leq \hat{q}_{(1-\gamma)}\}}{B} = 1 - \gamma,$$

   em que # indica o número de ocorrências do evento $\{W_p^{*b} \leq \hat{q}_{(1-\gamma)}\}$;

4. Determinar $\hat{q}_{1-\gamma}$, colocando em ordem crescente os valores obtidos das B estatísticas da razão de verossimilhanças bootstrap W_p^{*b}, $b = 1, 2, \ldots, B$, assim, a quantidade $\hat{q}_{1-\gamma}$ corresponde ao valor de W_p^{*b} que ocupa a posição $B \times (1-\gamma)$, supondo que este seja um número inteiro. Caso contrário, considere $k = \lfloor (B+1) \times \gamma \rfloor$ o maior inteiro menor ou igual a $(B+1) \times \gamma$, assim a quantidade $\hat{q}_{1-\gamma}$ corresponde ao valor W_p^{*b} que ocupa a posição $(B+1-k)$. Em R, use a função `quantile()`.

O teste da razão de verossimilhanças bootstrap (W_b) consiste em rejeitar H_0 se $W_p > \hat{q}_{1-\gamma}$.

3.4.2 Correção de Bartlett Bootstrap

Rocke (1989) propôs uma alternativa para o cálculo do fator de correção de Bartlett para a estatística da razão de verossimilhanças utilizando a técnica de bootstrap paramétrico apresentado anteriormente. Denotaremos essa estatística por W_{bb}, que pode ser obtida seguindo os passos:

1. Calcule a estatística W_p^{*b}, $b = 1, 2, \ldots, B$, como mostra o procedimento descrito na seção 3.4.1;

2. Faça $\overline{W}_p^{*b} = B^{-1} \sum_{b=1}^{B} W_p^{*b}$;

3. Determine $W_{bb} = k\dfrac{W_p}{\overline{W}_p^{*b}}$, em que W_p é o valor da estatística da razão de verossimilhanças (baseada na verossimilhança perfilada) obtida com base na amostra original e k é a quantidade de restrições sob a hipótese nula.

Segundo Rocke (1989) a estatística W_{bb}, sob H_0, possui distribuição assintótica ($\mathcal{X}^2_k$). Além disso, essa abordagem é computacionalmente mais eficiente do que a de bootstrap paramétrico, pois requer um menor número de reamostragens da amostra original (BAYER e CRIBARI-NETO, 2013).

Segundo Forero (2015) para a abordagem da correção de Bartlett bootstrap, é esperado que se obtenha resultados mais precisos com apenas 200 repetições, visto que estima a média da distribuição. Já para a estimativa do respectivo quantil crítico usando bootstrap paramétrico, é necessário pelo menos 1000 repetições para obter bons resultados.

Capítulo 4

Funções de Verossimilhança para o Parâmetro de Interesse em duas Distribuições de Probabilidade Contínuas

4.1 A Distribuição Lindley Ponderada

A distribuição Lindley Ponderada foi proposta por Ghitany et al. (2011) para a realização de modelagem de dados de tempo de vida.

Uma v.a X segue distribuição Lindley Ponderada com parâmetros $\mu > 0$ e $\beta > 0$ (denotada por $X \sim \text{LP}(\mu, \beta)$) se sua f.d.p for

$$f(x|\mu, \beta) = \frac{\mu^{\beta+1}}{(\mu + \beta)\Gamma(\beta)} x^{\beta-1}(1 + x)e^{-\mu x},$$

em que $x > 0$ e $\Gamma(\beta) = \int_0^\infty x^{\beta-1}e^{-x}\text{d}x$ é a função gama. Se $\beta = 1$, a distribuição Lindley Ponderada se reduz a distribuição Lindley.

Segundo Ramos et al. (2017) e Mazucheli et al. (2013), algumas características, tais como, o suporte nos reais positivos e o fato da função taxa de falha possuir diferentes formas (banheira, banheira invertida, crescente e decrescente), tornam a distribuição Lindley Ponderada altamente flexível e útil para modelar dados de tempo vida.

Na distribuição Lindley Ponderada, temos a facilidade para determinar alguns cumulantes em forma fechada, tais como, valores esperados das derivadas da função de log-verossimilhança, matriz de informação de Fisher, entre outros. Esse fato, favorece estudos relacionados à teoria assintótica de mais alta ordem, que configura o principal objetivo desse trabalho.

Seja $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ uma amostra aleatória de $X \sim \text{LP}(\mu, \beta)$. A função de

verossimilhança para os parâmetros μ e β é dada por

$$L(\mu,\beta)=\prod_{i=1}^{n} f(x_i|\mu,\beta)=\left[\frac{\mu^{\beta+1}}{(\mu+\beta)\Gamma(\beta)}\right]^n e^{-\mu T_0}\prod_{i=1}^{n} x_i^{\beta-1}(1+x_i),$$

em que $T_0=\sum_{i=1}^{n} x_i$.

A partir do resultado anterior segue que a função de log-verossimilhança para os parâmetros μ e β é da forma:

$$\ell(\mu,\beta)=n\left[(\beta+1)\log(\mu)-\log(\mu+\beta)-\log\Gamma(\beta)\right]-\mu T_0+(\beta-1)T_1+T_2,$$

com $T_1=\sum_{i=1}^{n}\log(x_i)$ e $T_2=\sum_{i=1}^{n}\log(1+x_i)$.

A matriz de informação observada da Lindley Ponderada é dada por

$$J(\mu,\beta)=n\begin{pmatrix} \dfrac{(\beta+1)}{\mu^2}-\dfrac{1}{(\mu+\beta)^2} & -\dfrac{1}{\mu}-\dfrac{1}{(\mu+\beta)^2} \\ -\dfrac{1}{\mu}-\dfrac{1}{(\mu+\beta)^2} & \psi'(\beta)-\dfrac{1}{(\mu+\beta)^2} \end{pmatrix}, \tag{4.1}$$

em que, $\psi'(\beta)=\dfrac{\partial^2}{\partial\beta^2}\log\Gamma(\beta)$ e as componentes da matriz são definidas da forma:

$$J_{11}(\mu,\beta)=-\frac{\partial^2}{\partial\mu^2}\ell(\mu,\beta),\quad J_{22}(\mu,\beta)=-\frac{\partial^2}{\partial\beta^2}\ell(\mu,\beta),$$

$$J_{12}(\mu,\beta)=-\frac{\partial^2}{\partial\mu\partial\beta}\ell(\mu,\beta),\quad J_{21}(\mu,\beta)=-\frac{\partial^2}{\partial\beta\partial\mu}\ell(\mu,\beta).$$

Para definir a matriz de informação esperada, calculamos o valor esperado dessas segundas derivadas apresentadas anteriormente. Veja que as componentes são constantes, então concluímos que:

$$K(\mu,\beta)=J(\mu,\beta)$$

ou seja, a matriz de informação esperada é igual a matriz de informação observada.

4.1.1 Geração de Números Pseudo-aleatórios

A geração de números pseudo-aleatórios é fundamental para a simulação de diferentes conjuntos de valores de uma v.a, que segue uma distribuição de acordo com determinada f.d.p com o intuito de avaliar cenários estatísticos similares aos observados.

Devido ao fato de que a função quantílica da distribuição Lindley Ponderada não tem forma fechada, iremos utilizar um resultado proposto por Ghitany et al. (2011),

segundo o qual a f.d.p da distribuição Lindley Ponderada pode ser expressada como combinação de dois componentes, da forma

$$f(x) = pf_1(x) + (1-p)f_2(x), \tag{4.2}$$

em que $p = \dfrac{\mu}{\mu+\beta}$ e

$$f_j(x) = \frac{\mu^{\beta+j-1}}{\Gamma(\beta+j-1)} x^{\beta+j-2} e^{-\mu x},\ x > 0, \quad \mu > 0, \quad \beta > 0, \quad j = 1, 2$$

é a f.d.p. da distribuição gama com parâmetro de forma $(\beta + j - 1)$ e de escala μ, denotada por $Gama(\beta + j - 1, \mu)$.

Com base no exposto, iremos gerar números pseudo-aleatórios da distribuição Lindley Ponderada por meio de distribuições Gama da seguinte forma:

1. Gerar uma observação $p \sim Bernoulli\left(\dfrac{\mu}{\mu+\beta}\right)$;
2. Gerar uma observação $y_1 \sim Gama(\beta, \mu)$;
3. Gerar $y_2 \sim Gama(\beta + 1, \mu)$;
4. Calcular $x = py_1 + (1-p)y_2$. Assim, x será uma observação proveniente da distribuição Lindley Ponderada com parâmetros μ e β.

Vale ressaltar que p, y_1 e y_2 devem ser gerados de forma independente.

4.1.2 Família Exponencial

Nesta seção veremos que a distribuição Lindley Ponderada pertence à família exponencial, o que nos permite obter alguns resultados interessantes.

Definição 4.1. *Dizemos que a distribuição da v.a X pertence à família exponencial de dimensão k se a f.d.p de X é dada por*

$$f(x|\boldsymbol{\theta}) = \exp\left[\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x) + d(\boldsymbol{\theta}) + S(x)\right], x \in A$$

em que c_j, T_j, d e S são funções reais, $j = 1, \ldots, k$, e A não depende de $\boldsymbol{\theta}$.

Para a f.d.p da distribuição Lindley Ponderada, podemos definir:

$$
\begin{aligned}
f(x|\mu,\beta) &= \frac{\mu^{\beta+1}}{(\mu+\beta)\Gamma(\beta)}x^{\beta-1}(1+x)e^{-\mu x},\\
\log[f(x|\mu,\beta)] &= \log\left[\frac{\mu^{\beta+1}}{(\mu+\beta)\Gamma(\beta)}\right]+\log(x^{\beta-1})+\log(1+x)+\log(e^{-\mu x})\\
&= (\beta+1)\log\mu-\log(\mu+\beta)-\log\Gamma(\beta)+(\beta-1)\log x\\
&\quad +\log(1+x)-\mu x\\
&= \beta\log\mu+\log\mu-\log(\mu+\beta)-\log\Gamma(\beta)+\beta\log x-\log x\\
&\quad +\log(1+x)-\mu x,\\
f(x|\mu,\beta) &= \exp(\beta\log\mu+\log\mu-\log(\mu+\beta)-\log\Gamma(\beta)+\beta\log x-\log x\\
&\quad +\log(1+x)-\mu x).
\end{aligned}
$$

Com base nesse resultado, temos que

$$
T_1(x)=x,\quad T_2(x)=\log x,\quad c_1(\boldsymbol{\theta})=-\mu,\quad c_2(\boldsymbol{\theta})=\beta,
$$

$$
d(\boldsymbol{\theta})=\log\left[\frac{\mu^{\beta+1}}{(\mu+\beta)\Gamma(\beta)}\right],\quad S(x)=\log\left(\frac{1+x}{x}\right).
$$

Dessa forma, concluímos que a distribuição de Lindley Ponderada pertence à família exponencial bidimensional.

Segundo Bolfarine e Sandoval (2001), as amostras de famílias exponenciais de dimensão k têm distribuições que são membros da família exponencial de dimensão k. Para uma amostra $X_1,\ldots,X_n$, temos que a f.d.p conjunta de $X_1,\ldots,X_n$ é dada por:

$$
f(x_1,\ldots,x_n|\boldsymbol{\theta})=\exp\left[\sum_{j=1}^{n}c_j^*(\boldsymbol{\theta})\sum_{i=1}^{n}T_j^*(x_i)+d^*(\boldsymbol{\theta})+S^*(x)\right],
$$

em que, para $j=1,\ldots,k$,

$$
T_j^*(x)=\textstyle\sum_{i=1}^{n}T_j(x_i),\quad c_j^*(\boldsymbol{\theta})=c_j(\boldsymbol{\theta}),\quad S^*(x)=\sum_{i=1}^{n}S(x_i),\quad d^*(\boldsymbol{\theta})=nd(\boldsymbol{\theta}),
$$

assim, $(T_1^*,T_2^*,\ldots,T_k^*)$ é conjuntamente suficiente para $\boldsymbol{\theta}$.

Portanto, a distribuição de uma amostra da densidade da Lindley Ponderada é

também da família exponencial com $T_1(X) = \sum_{i=1}^{n} X_i$ e $T_2(X) = \sum_{i=1}^{n} \log X_i$, que são estatísticas conjuntamente suficientes minimais para (μ, β).

4.1.3 Inferências

Nesta seção iremos utilizar os resultados obtidos em 4.1 para determinar a função de verossimilhança perfilada e suas versões modificadas para o parâmetro de interesse da distribuição Lindley Ponderada.

Inicialmente vamos determinar a função de verossimilhança perfilada. Para isso, iremos considerar os parâmetros β e μ sendo de interesse e de perturbação, respectivamente. A escolha de β como parâmetro de interesse se deve ao fato de que este é o parâmetro que generaliza a distribuição Lindley Ponderada, pois se $\beta = 1$, teremos a distribuição Lindley. Dessa forma, podemos testar $H_0 : \beta = 1$ *versus* $H_1 : \beta \neq 1$. Além disso, podemos determinar o EMV de μ em função de β em forma fechada.

Para definir a função de verossimilhança perfilada precisamos definir o EMV para μ que será o valor $\hat{\mu}_\beta$ que maximiza $\ell(\mu, \beta)$, e é obtido como solução da equação $\dfrac{\partial \ell(\mu, \beta)}{\partial \mu} = 0$ para μ fixo. Dessa forma, derivando $\ell(\mu, \beta)$ em relação a μ, teremos:

$$\frac{\partial \ell(\mu, \beta)}{\partial \mu} = n \left[\frac{\beta + 1}{\mu} - \frac{1}{(\mu + \beta)} \right] - T_0 = 0.$$

Como $\dfrac{\partial^2 \ell(\mu, \beta)}{\partial \mu^2} < 0$, então o EMV para μ, será:

$$\hat{\mu}_\beta = \frac{\beta(n - T_0) + \sqrt{\beta^2(n + T_0)^2 + 4n\beta T_0}}{2T_0}.$$

Portanto, a função de verossimilhança perfilada e a função de log-verossimilhança perfilada para o parâmetro β são dadas, respectivamente, por:

$$L_p(\beta) = L(\hat{\mu}_\beta, \beta) = \left[\frac{\hat{\mu}_\beta^{\beta+1}}{(\hat{\mu}_\beta + \beta)\Gamma(\beta)} \right]^n e^{-\hat{\mu}_\beta T_0} \prod_{i=1}^{n} x_i^{\beta-1}(1 + x_i)$$

e

$$\begin{aligned} \ell_p(\beta) &= \ell(\hat{\mu}_\beta, \beta) = n\left[(\beta + 1)\log(\hat{\mu}_\beta) - \log(\hat{\mu}_\beta + \beta) - \log\Gamma(\beta)\right] \\ &\quad -\hat{\mu}_\beta T_0 + (\beta - 1)T_1 + T_2. \end{aligned} \tag{4.3}$$

A partir da modificação de Barndorff-Nielsen (1983) e Severini (1998) definida em 3.2, vamos determinar a função de log-verossimilhança perfilada modificada para uma v.a X que segue distribuição Lindley Ponderada. Dessa forma, temos que:

$$\ell_{\text{BNS}}(\beta) = \ell_p(\beta) + \frac{1}{2}\log|j_{\mu\mu}(\hat{\mu}_\beta, \beta)| - \log|I_\mu(\hat{\mu}_\beta, \beta; \hat{\mu}, \hat{\beta})|,$$

em que $I_\mu(\mu,\beta;\mu_0,\beta_0) = E_{(\mu_0,\beta_0)}[\ell_\mu(\mu,\beta)\ell_\mu(\mu_0,\beta_0)^\top]$ e $\ell_\mu(\mu,\beta) = \frac{\partial \ell(\mu,\beta)}{\partial \mu}$.

Veja que já definimos $\ell_p(\beta)$ em (4.3). Portanto, precisamos determinar $j_{\mu\mu}(\hat{\mu}_\beta,\beta)$ e $I_\mu(\mu,\beta;\mu_0,\beta_0)$. Da matriz de informação observada (4.1), temos que:

$$j_{\mu\mu}(\hat{\mu}_\beta,\beta) = \frac{\beta+1}{\hat{\mu}_\beta^2} - \frac{1}{(\hat{\mu}_\beta+\beta)^2}.$$

Agora vamos determinar $I_\mu(\mu,\beta;\mu_0,\beta_0)$:

$$I_\mu(\mu,\beta,\mu_0,\beta_0) = E_{(\mu_0,\beta_0)}[l_\mu(\mu,\beta)l_\mu(\mu_0,\beta_0)^\top],$$

fazendo, $l_\mu(\mu,\beta)l_\mu(\mu_0,\beta_0)^\top = r$, temos:

$$\begin{aligned}
r &= \left[\frac{n(\beta+1)}{\mu} - \frac{n}{\mu+\beta} - T_0\right]\left[\frac{n(\beta_0+1)}{\mu_0} - \frac{n}{\mu_0+\beta_0} - T_0\right] \\
&= \left[\frac{n(\beta+1)(\mu+\beta)-n\mu}{\mu(\mu+\beta)} - T_0\right]\left[\frac{n(\beta_0+1)(\mu_0+\beta_0)-n\mu_0}{\mu_0(\mu_0+\beta_0)} - T_0\right] \\
&= \left[\frac{n((\beta+1)(\mu+\beta)-\mu)}{\mu(\mu+\beta)} - T_0\right]\left[\frac{n((\beta_0+1)(\mu_0+\beta_0)-\mu_0)}{\mu_0(\mu_0+\beta_0)} - T_0\right] \\
&= \frac{n^2[(\beta+1)(\mu+\beta)-\mu][(\beta_0+1)(\mu_0+\beta_0)-\mu_0]}{\mu\mu_0(\mu+\beta)(\mu_0+\beta_0)} \\
&\quad -\frac{n[(\beta+1)(\mu+\beta)-\mu]}{\mu(\mu+\beta)}T_0 - \frac{n[(\beta_0+1)(\mu_0+\beta_0)-\mu_0]}{\mu_0(\mu_0+\beta_0)}T_0 + T_0^2.
\end{aligned}$$

Aplicando a esperança, obtemos:

$$\begin{aligned}
E_{(\mu_0,\beta_0)}[l_\mu(\mu,\beta)l_\mu(\mu_0,\beta_0)^\top] &= \frac{n^2[(\beta+1)(\mu+\beta)-\mu][(\beta_0+1)(\mu_0+\beta_0)-\mu_0]}{\mu\mu_0(\mu+\beta)(\mu_0+\beta_0)} \\
&\quad -\frac{n[(\beta+1)(\mu+\beta)-\mu]}{\mu(\mu+\beta)}E_{(\mu_0,\beta_0)}(T_0) \\
&\quad -\frac{n[(\beta_0+1)(\mu_0+\beta_0)-\mu_0]}{\mu_0(\mu_0+\beta_0)}E_{(\mu_0,\beta_0)}(T_0) \\
&\quad +E_{(\mu_0,\beta_0)}(T_0^2).
\end{aligned} \tag{4.4}$$

Como observado em (4.4), precisamos determinar $E_{(\mu_0,\beta_0)}(T_0)$ e $E_{(\mu_0,\beta_0)}(T_0^2)$, então faremos:

$$E_{(\mu_0,\beta_0)}(T_0) = E_{(\mu_0,\beta_0)}(X_1 + X_2 + \cdots + X_n) = nE_{(\mu_0,\beta_0)}(X) \tag{4.5}$$

e

$$\begin{aligned} E_{(\mu_0,\beta_0)}(T_0^2) &= Var_{(\mu_0,\beta_0)}(T_0) + [E_{(\mu_0,\beta_0)}(T_0)]^2 \\ &= Var_{(\mu_0,\beta_0)}(X_1 + X_2 + \cdots + X_n) + [nE_{(\mu_0,\beta_0)}(X)]^2 \\ &= nVar_{(\mu_0,\beta_0)}(X) + n^2E^2_{(\mu_0,\beta_0)}(X). \end{aligned} \tag{4.6}$$

Além disso, segundo Ghitany et al. (2011) a esperança e variância de uma v.a que segue distribuição Lindley Ponderada são dadas, respectivamente, por:

$$E_{(\mu_0,\beta_0)}(X) = \frac{\beta_0(\mu_0+\beta_0+1)}{(\mu_0+\beta_0)\mu_0} \quad \text{e} \quad Var_{(\mu_0,\beta_0)}(X) = \frac{(\beta_0+1)(\mu_0+\beta_0)^2-\mu_0^2}{\mu_0^2(\mu_0+\beta_0)^2}. \tag{4.7}$$

Substituindo os resultados (4.5), (4.6) e (4.7) em (4.4), obtemos:

$$\begin{aligned} E_{(\mu_0,\beta_0)}[l_\mu(\mu,\beta)l_\mu(\mu_0,\beta_0)^\top] &= \frac{n^2[(\beta+1)(\mu+\beta)-\mu][(\beta_0+1)(\mu_0+\beta_0)-\mu_0]}{\mu\mu_0(\mu+\beta)(\mu_0+\beta_0)} \\ &\quad -\frac{n^2((\beta+1)(\mu+\beta)-\mu)\beta_0(\mu_0+\beta_0+1)}{\mu\mu_0(\mu+\beta)(\mu_0+\beta_0)} \\ &\quad -\frac{n^2((\beta_0+1)(\mu_0+\beta_0)-\mu_0)\beta_0(\mu_0+\beta_0+1)}{\mu_0^2(\mu_0+\beta_0)^2} \\ &\quad +\frac{n[(\beta_0+1)(\mu_0+\beta_0)^2-\mu_0^2]+n^2\beta_0^2(\mu_0+\beta_0+1)^2}{\mu_0^2(\mu_0+\beta_0)^2} \\ &= \frac{n\beta_0(\beta_0^2+\beta_0+2\beta_0\mu_0+\mu_0^2+2\mu_0)}{\mu_0^2(\mu_0+\beta_0)^2} \\ &= \frac{n\beta_0[(\mu_0+\beta_0)^2+2\mu_0+\beta_0]}{\mu_0^2(\mu_0+\beta_0)^2}. \end{aligned}$$

Portanto,

$$I_\mu(\mu,\beta,\mu_0,\beta_0) = \frac{n\beta_0[(\mu_0+\beta_0)^2+2\mu_0+\beta_0]}{\mu_0^2(\mu_0+\beta_0)^2}.$$

4.2 A Distribuição Weibull Exponencializada

A distribuição Weibull Exponencializada foi proposta por Mudholkar e Srivastava (1993) para análise de dados de taxa de falha.

Uma v.a X segue distribuição Weibull Exponencializada com parâmetros $a > 0$, $b > 0$ e $\lambda > 0$ (denotada por $X \sim \text{WE}(a, b, \lambda)$), se sua f.d.p for dada por

$$f(x|a, b, \lambda) = \frac{b\lambda}{a}\left(\frac{x}{a}\right)^{b-1}\left[1 - \exp\left(-\left(\frac{x}{a}\right)^{b}\right)\right]^{\lambda-1}\exp\left(-\left(\frac{x}{a}\right)^{b}\right), \tag{4.8}$$

em que $x > 0$. Além disso, sua f.d.a é dada por

$$F(x) = \left[1 - \exp\left(-\left(\frac{x}{a}\right)^{b}\right)\right]^{\lambda}. \tag{4.9}$$

Segundo Mudholkar e Srivastava (1993), a distribuição Weibull Exponencializada inclui não apenas distribuições com taxas de falha unimodal e de banheira, mas fornece uma classe mais ampla de taxas de falha monótonas.

É interessante observar que ao fixarmos o parâmetro $\lambda = 1$ da distribuição Weibull Exponencializada, esta apresentará a forma da distribuição Weibull de parâmetros a e b.

Seja $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ uma amostra aleatória de $X \sim \text{WE}(a, b, \lambda)$. A função de verossimilhança para os parâmetros a, b e λ é dada por

$$L(a, b, \lambda) = \left(\frac{b\lambda}{a}\right)^{n}\prod_{i=1}^{n}\left(\frac{x_i}{a}\right)^{b-1}\left[1 - \exp\left(-\left(\frac{x_i}{a}\right)^{b}\right)\right]^{\lambda-1}\exp\left(-\left(\frac{x_i}{a}\right)^{b}\right).$$

Com base nisso, segue que a função de log-verossimilhança para os parâmetros a, b e λ pode ser definida da forma

$$\begin{aligned}\ell(a, b, \lambda) &= n\log\left(\frac{b\lambda}{a}\right) + (b-1)\sum_{i=1}^{n}\log\left(\frac{x_i}{a}\right) \\ &\quad +(\lambda-1)\sum_{i=1}^{n}\log\left[1 - \exp\left(-\left(\frac{x_i}{a}\right)^{b}\right)\right] - \sum_{i=1}^{n}\left(\frac{x_i}{a}\right)^{b}.\end{aligned}$$

A matriz de informação observada é dada por

$$J(a,b,\lambda) = \begin{pmatrix} -\frac{\partial^2 \ell(a,b,\lambda)}{\partial a^2} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial a \partial b} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial a \partial \lambda} \\ -\frac{\partial^2 \ell(a,b,\lambda)}{\partial b \partial a} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial b^2} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial b \partial \lambda} \\ -\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda \partial a} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda \partial b} & -\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda^2} \end{pmatrix}, \tag{4.10}$$

em que

$$\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2},$$

$$\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda \partial a} = \frac{\partial^2 \ell(a,b,\lambda)}{\partial a \partial \lambda} = -\frac{b}{a^2} \sum_{i=1}^{n} \frac{e^{-\left(\frac{x_i}{a}\right)^b} x_i \left(\frac{x_i}{a}\right)^{-1+b}}{1 - e^{-\left(\frac{x_i}{a}\right)^b}},$$

$$\frac{\partial^2 \ell(a,b,\lambda)}{\partial \lambda \partial b} = \frac{\partial^2 \ell(a,b,\lambda)}{\partial b \partial \lambda} = \sum_{i=1}^{n} \frac{e^{-\left(\frac{x_i}{a}\right)^b} \log\left(\frac{x_i}{a}\right) \left(\frac{x_i}{a}\right)^b}{1 - e^{-\left(\frac{x_i}{a}\right)^b}},$$

$$\begin{aligned} \frac{\partial^2 \ell(a,b,\lambda)}{\partial a^2} &= \frac{n}{a^2} + \frac{(-1+b)n}{a} - \frac{2b}{a^3} \sum_{i=1}^{n} x_i \left(\frac{x_i}{a}\right)^{-1+b} \\ &\quad + (1-b)\frac{b}{a^4} \sum_{i=1}^{n} x_i^2 \left(\frac{x_i}{a}\right)^{-2+b} \\ &\quad + (\lambda - 1)\frac{2b}{a^3} \sum_{i=1}^{n} \frac{e^{-\left(\frac{x_i}{a}\right)^b} x_i \left(\frac{x_i}{a}\right)^{-1+b}}{1 - e^{-\left(\frac{x_i}{a}\right)^b}} \\ &\quad - (\lambda - 1)\frac{b^2}{a^4} \sum_{i=1}^{n} \frac{e^{-2\left(\frac{x_i}{a}\right)^b} x_i^2 \left(\frac{x_i}{a}\right)^{-2+2b}}{\left(1 - e^{-\left(\frac{x_i}{a}\right)^b}\right)^2} \\ &\quad - (\lambda - 1)\frac{b^2}{a^4} \sum_{i=1}^{n} \frac{e^{-\left(\frac{x_i}{a}\right)^b} x_i^2 \left(\frac{x_i}{a}\right)^{-2+2b}}{1 - e^{-\left(\frac{x_i}{a}\right)^b}} \\ &\quad - (\lambda - 1)(1-b)\frac{b^2}{a^4} \sum_{i=1}^{n} \frac{e^{-\left(\frac{x_i}{a}\right)^b} x_i^2 \left(\frac{x_i}{a}\right)^{-2+2b}}{1 - e^{-\left(\frac{x_i}{a}\right)^b}}, \end{aligned}$$

$$\frac{\partial^2\ell(a,b,\lambda)}{\partial a\partial b} = \frac{\partial^2\ell(a,b,\lambda)}{\partial b\partial a} = -\frac{n}{a} + \frac{1}{a^2}\sum_{i=1}^{n} x_i\left(\frac{x_i}{a}\right)^{-1+b} + \frac{b}{a^2}\sum_{i=1}^{n}\log\left(\frac{x_i}{a}\right)x_i\left(\frac{x_i}{a}\right)^{-1+b}$$
$$-(\lambda-1)\frac{1}{a^2}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{a}\right)^b}x_i\left(\frac{x_i}{a}\right)^{-1+b}}{1-e^{-\left(\frac{x_i}{a}\right)^b}}$$
$$-(\lambda-1)\frac{b}{a^2}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)x_i\left(\frac{x_i}{a}\right)^{-1+b}}{1-e^{-\left(\frac{x_i}{a}\right)^b}}$$
$$+(\lambda-1)\frac{b}{a^2}\sum_{i=1}^{n}\frac{e^{-2\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)x_i\left(\frac{x_i}{a}\right)^{-1+2b}}{\left(1-e^{-\left(\frac{x_i}{a}\right)^b}\right)^2}$$
$$+(\lambda-1)\frac{b}{a^2}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)x_i\left(\frac{x_i}{a}\right)^{-1+2b}}{1-e^{-\left(\frac{x_i}{a}\right)^b}}$$

e

$$\frac{\partial^2\ell(a,b,\lambda)}{\partial b^2} = -\frac{n}{b^2} - \sum_{i=1}^{n}\log\left(\frac{x_i}{a}\right)^2\left(\frac{x_i}{a}\right)^b$$
$$+(\lambda-1)\frac{1}{a^2}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)^2\left(\frac{x_i}{a}\right)^b}{1-e^{-\left(\frac{x_i}{a}\right)^b}}$$
$$-(\lambda-1)\frac{1}{a^2}\sum_{i=1}^{n}\frac{e^{-2\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)^2\left(\frac{x_i}{a}\right)^{2b}}{\left(1-e^{-\left(\frac{x_i}{a}\right)^b}\right)^2}$$
$$-(\lambda-1)\frac{1}{a^2}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{a}\right)^b}\log\left(\frac{x_i}{a}\right)^2\left(\frac{x_i}{a}\right)^{2b}}{1-e^{-\left(\frac{x_i}{a}\right)^b}}.$$

4.2.1 Geração de Números Pseudo-aleatórios

Para gerar números pseudo-aleatórios de uma v.a X que segue distribuição Weibull Exponencializada utilizamos o método da transformação inversa, em que a função quantílica (2.5) da distribuição Weibull Exponencializada é dada por

$$Q_X(u) = a\left(-\log(1-u)^{\frac{1}{\lambda}}\right)^{\frac{1}{b}}, \quad 0<u<1.$$

O método da transformação inversa é um dos procedimentos mais simples para gerar dados aleatórios a partir de uma determinada distribuição, sendo comumente utilizado quando a função quantílica da v.a pode ser expressa de forma fechada. Dessa forma, se $U \sim \mathcal{U}(0,1)$, então $Q_x(U) \sim WE(a,b,\lambda)$.

4.2.2 Inferências

Nesta seção iremos utilizar os resultados obtidos em 4.2 para determinar as funções de verossimilhança perfilada e perfilada modificada para o parâmetro de interesse da Weibull Exponencializada.

Primeiramente, iremos definir a função de verossimilhança perfilada. Para isso, consideramos os parâmetros λ e $\boldsymbol{\nu} = (a, b)^\top$, como sendo de interesse e perturbação, respectivamente. A escolha de λ como parâmetro de interesse se deve ao fato de que este generaliza a distribuição Weibull Exponencializada, permitindo testar esse modelo contra Weibull.

Para expressar a função de log-verossimilhança perfilada é necessário obter o EMV dos parâmetros a e b, fixando o parâmetro de interesse. Em outras palavras, $\hat{a}_\lambda$ e $\hat{b}_\lambda$ são obtidos resolvendo as equações não lineares $\dfrac{\partial \ell(a,b,\lambda)}{\partial a} = 0$ e $\dfrac{\partial \ell(a,b,\lambda)}{\partial b} = 0$, para λ fixo. No entanto, essas estimativas não podem ser obtidas analiticamente. Dessa forma, serão calculadas computacionalmente utilizando o algoritmo de otimização não-linear restrito SQP (Sequential Quadratic Programming). Neste trabalho usamos a função `slsqp()` do pacote nloptr, que esta implementado na linguagem R e foi proposto por Johnson (2020).

Denotaremos os estimadores de máxima verossimilhança de a e b para λ fixado por $\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}} = (\hat{a}_\lambda, \hat{b}_\lambda)^\top$, então a função de log-verossimilhança perfilada para o parâmetro λ é dada por

$$\begin{aligned}
\ell_p(\lambda) &= \ell(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda) \\
&= n \log\left(\frac{\hat{b}_\lambda \lambda}{\hat{a}_\lambda}\right) + (\hat{b}_\lambda - 1)\sum_{i=1}^n \log\left(\frac{x_i}{\hat{a}_\lambda}\right) \\
&\quad + (\lambda - 1)\sum_{i=1}^n \log\left[1 - \exp\left(-\left(\frac{x_i}{\hat{a}_\lambda}\right)^{\hat{b}_\lambda}\right)\right] - \sum_{i=1}^n \left(\frac{x_i}{\hat{a}_\lambda}\right)^{\hat{b}_\lambda}.
\end{aligned} \tag{4.11}$$

Ressaltamos que o valor esperado das segundas derivadas da função de log-verossimilhança da distribuição Weibull Exponencializada não podem ser obtidos em forma fechada. Portanto, para determinar a função de log-verossimilhança modificada usaremos uma aproximação baseada em covariâncias empíricas, apresentada em 3.2.3, dada por

$$\ell_{\mathrm{BNS}^*}(\lambda) = \ell_p(\lambda) + \frac{1}{2}\log|j_{\boldsymbol{\nu}\boldsymbol{\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)| - \log|\breve{I}_{\boldsymbol{\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda; \hat{\boldsymbol{\nu}}, \hat{\lambda})|,$$

em que $\breve{I}_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \lambda; \boldsymbol{\nu_0}, \lambda_0) = \sum_{j=1}^n \ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu}, \lambda)\ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu_0}, \lambda_0)^\top$.

O termo $\ell_p(\lambda)$ é dado por 4.11, $j_{\boldsymbol{\nu\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)$ é dada por

$$j_{\boldsymbol{\nu\nu}}(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda) = \begin{pmatrix} -\dfrac{\partial^2 \ell(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)}{\partial a^2} & -\dfrac{\partial^2 \ell(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)}{\partial a \partial b} \\ \\ -\dfrac{\partial^2 \ell(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)}{\partial b \partial a} & -\dfrac{\partial^2 \ell(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)}{\partial b^2} \end{pmatrix}, \tag{4.12}$$

sendo os componentes dessa matriz apresentados na Seção 4.2 e que neste caso devem ser avaliadas no vetor $(\hat{\boldsymbol{\nu}}_{\boldsymbol{\lambda}}, \lambda)^\top$.

Para determinar $\breve{I}_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \lambda; \boldsymbol{\nu_0}, \lambda_0)$, faremos:

$$\begin{aligned} \breve{I}_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \lambda; \boldsymbol{\nu_0}, \lambda_0) &= \textstyle\sum_{j=1}^n \ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu}, \lambda) \ell_{\boldsymbol{\nu}}^{(j)}(\boldsymbol{\nu_0}, \lambda_0)^\top \\ &= \textstyle\sum_{j=1}^n \begin{bmatrix} \ell_a^{(j)}(\boldsymbol{\nu}, \lambda) \\ \\ \ell_b^{(j)}(\boldsymbol{\nu}, \lambda) \end{bmatrix} \begin{bmatrix} \ell_a^{(j)}(\boldsymbol{\nu_0}, \lambda_0) & \ell_b^{(j)}(\boldsymbol{\nu_0}, \lambda_0) \end{bmatrix} \\ &= \textstyle\sum_{j=1}^n \begin{bmatrix} \ell_a^{(j)}(\boldsymbol{\nu}, \lambda)\ell_a^{(j)}(\boldsymbol{\nu_0}, \lambda_0) & \ell_a^{(j)}(\boldsymbol{\nu}, \lambda)\ell_b^{(j)}(\boldsymbol{\nu_0}, \lambda_0) \\ \\ \ell_b^{(j)}(\boldsymbol{\nu}, \lambda)\ell_a^{(j)}(\boldsymbol{\nu_0}, \lambda_0) & \ell_b^{(j)}(\boldsymbol{\nu}, \lambda)\ell_b^{(j)}(\boldsymbol{\nu_0}, \lambda_0) \end{bmatrix}, \end{aligned}$$

em que

$$\ell_a^{(j)}(\boldsymbol{\nu}, \lambda) = -\frac{n}{a} - \frac{(b-1)n}{a} + \frac{b}{a^2} x_j \left(\frac{x_j}{a}\right)^{-1+b} - (\lambda - 1)\frac{b e^{-\left(\frac{x_j}{a}\right)^b} x_j \left(\frac{x_j}{a}\right)^{-1+b}}{a^2 \left(1 - e^{-\left(\frac{x_j}{a}\right)^b}\right)}$$

e

$$\ell_b^{(j)}(\boldsymbol{\nu}, \lambda) = \frac{n}{b} + \log\left(\frac{x_j}{a}\right) - \log\left(\frac{x_j}{a}\right)\left(\frac{x_j}{a}\right)^b + (\lambda - 1)\frac{e^{-\left(\frac{x_j}{a}\right)^b} \log\left(\frac{x_j}{a}\right)\left(\frac{x_j}{a}\right)^b}{1 - e^{-\left(\frac{x_j}{a}\right)^b}}.$$

Capítulo 5

Resultados Numéricos

Neste capítulo objetivamos comparar e avaliar o desempenho do TRV baseado na verossimilhança perfilada (W_p) e suas versões modificadas (W_{BNS} e W_{BNS^*}) com o aperfeiçoamento de Sousa (2020) (W_p^*). Além disso, para efeito de comparação incluímos os testes que envolvem a correção de boostrap paramétrico (W_b) e Bartlett bootstrap (W_{bb}). Para isso, realizamos simulações de Monte Carlo por meio da Linguagem de Programação R (R CORE TEAM, 2019) e utilizamos como critério de comparação a taxa de rejeição empírica da hipótese nula.

Para as simulações selecionamos a distribuição Lindley Ponderada e a distribuição Weibull Exponencializada. Nos testes consideramos 15000 réplicas de Monte Carlo, adotamos os níveis nominais $\gamma = 1\%$, 5% e 10%, e tamanhos amostrais $n = 20, 25, 30, 40, 50, 100, 200$. Para os testes que envolvem bootstrap usamos 2000 réplicas bootstrap.

Para realizar as simulações, utilizamos o R versão 4.0 em uma máquina com a seguinte configuração: Intel Core Xeon 3.6GHz de 4 núcleos e 8 Threads, 16 GB de memória RAM e sistema operacional Ubuntu 20.04.

Ainda neste capítulo apresentaremos quatro exemplos numéricos utilizando conjunto de dados reais.

5.1 Testes para a Distribuição Lindley Ponderada

Nesta Seção apresentamos os resultados numéricos para o parâmetro de interesse β da distribuição Lindley Ponderada. Estamos interessados em testar a hipótese nula $H_0 : \beta = \beta_0$ *versus* $H_1 : \beta \neq \beta_0$. Nas Tabelas 5.1 e 5.2 dispomos dos resultados referentes aos cenários em que β é igual a 1,0 e 2,5, respectivamente, considerando o parâmetro de perturbação μ fixo em 1,5. É importante frisar que os valores das tabelas são exibidos em porcentagens.

Considerando cada tamanho amostral e nível nominal adotado, calculamos as taxas de rejeição dos testes baseados nas estatísticas W_p, W_p^*, W_{BNS}, W_b e W_{bb}, ou

Tabela 5.1: Taxas de rejeições nulas, $\beta = 1,0$.

n	γ	W_p	W_p^*	W_{BNS}	W_b	W_{bb}
20	1,0	1,167	1,080	1,013	1,053	1,013
	5,0	5,947	5,080	5,307	5,313	5,253
	10,0	11,653	9,427	10,467	10,227	10,080
25	1,0	1,287	1,213	0,987	1,033	1,046
	5,0	5,793	5,053	5,133	5,020	4,993
	10,0	11,200	9,447	10,320	10,200	10,147
30	1,0	1,180	1,100	0,993	1,127	1,040
	5,0	6,000	5,447	5,407	5,187	5,187
	10,0	11,147	9,693	10,320	10,400	10,413
40	1,0	1,220	1,180	1,073	0,987	0,940
	5,0	5,713	5,320	5,080	5,027	4,987
	10,0	11,047	9,940	10,467	10,173	10,080
50	1,0	1,227	1,200	1,020	1,147	1,133
	5,0	5,407	5,153	5,060	5,080	5,067
	10,0	10,667	9,840	10,200	10,280	10,280
100	1,0	1,093	1,093	1,053	1,180	1,160
	5,0	5,433	5,307	5,200	5,127	5,133
	10,0	10,620	10,180	10,440	10,113	10,133
200	1,0	1,193	1,193	1,107	1,053	1,013
	5,0	5,460	5,393	5,287	5,093	5,040
	10,0	10,667	10,480	10,460	10,093	10,033

seja, estimamos via simulação de Monte Carlo as probabilidades $P(W_p \geq \mathcal{X}^2_{(1-\gamma;1)})$, $P(W_p^* \geq \mathcal{X}^{\inf}_{(1-\gamma;1)})$, $P(W_{\text{BNS}} \geq \mathcal{X}^2_{(1-\gamma;1)})$, $P(W_{bb} \geq \mathcal{X}^2_{(1-\gamma;1)})$, em que $\mathcal{X}^2_{(1-\gamma;1)}$ e $\mathcal{X}^{\inf}_{(1-\gamma;1)}$ são o quantil $1-\gamma$ da distribuição $\mathcal{X}^2_1$ e $\mathcal{X}^{\inf}_1$, respectivamente. Para o teste bootstrap paramétrico estimamos $P(W_b \geq \hat{q}_{(1-\gamma)})$, conforme descrito na Seção 3.4.1.

Avaliando as Tabelas 5.1 e 5.2 é possível observar que para todos os níveis nominais e tamanhos amostrais, os testes baseados em W_p se mostraram liberais [1]. Os testes baseados na estatística W_p^* se apresentaram liberais para os níveis nominais $\gamma = 1\%$ e 5% e conservativo para o nível nominal de $\gamma = 10\%$.

Note que, em geral, o teste baseado na estatística W_p^* atenua a tendência liberal do teste baseado na estatística W_p, principalmente para amostras de tamanho pequeno a moderado. Por exemplo, ao nível nominal de $\gamma = 5\%$, $n = 20$ e $\beta = 1,0$, os testes baseados em W_p e W_p^* tiveram suas taxas de rejeições 5,947 e 5,080, respectivamente. Além disso, podemos observar que conforme o tamanho amostral cresce, as taxas de rejeições de ambos os testes aproximam-se do nível nominal especificado, como é esperado.

Analisando as Tabelas 5.1 e 5.2, observamos que de modo geral o teste baseado na estatística W_{BNS} apresenta caráter liberal. Observamos que, em geral os testes

[1] O teste é dito liberal quando taxas de rejeições estimadas estão acima do nível nominal especificado.

baseados em W_p^* e W_{BNS} possuem resultados semelhantes, próximos do nível nominal especificado e desempenho superior aos testes baseados em W_p. Por exemplo, ao nível nominal $\gamma = 5\%$, $n = 25$ e $\beta = 2,5$, os testes baseados em W_p^* e W_{BNS} tiveram suas taxas de rejeições 5,393 e 5,387, respectivamente, enquanto que a taxa baseada em W_p foi 6,027. Notamos também que para n grande os testes baseados em W_p, W_p^* e W_{BNS} são similares.

Já os testes baseados em W_b e W_{bb} demonstram resultados semelhantes e próximos ao nível nominal especificado, principalmente para os níveis nominais 1% e 5%. Por exemplo, com $n = 30$, $\beta = 1,0$ e ao nível nominal de 5%, as taxas baseadas em W_b e W_{bb}, foram iguais a 5,187, enquanto para $\gamma = 10$, foram 10,400 e 10,413, respectivamente.

Tabela 5.2: Taxas de rejeições nulas, $\beta = 2,5$.

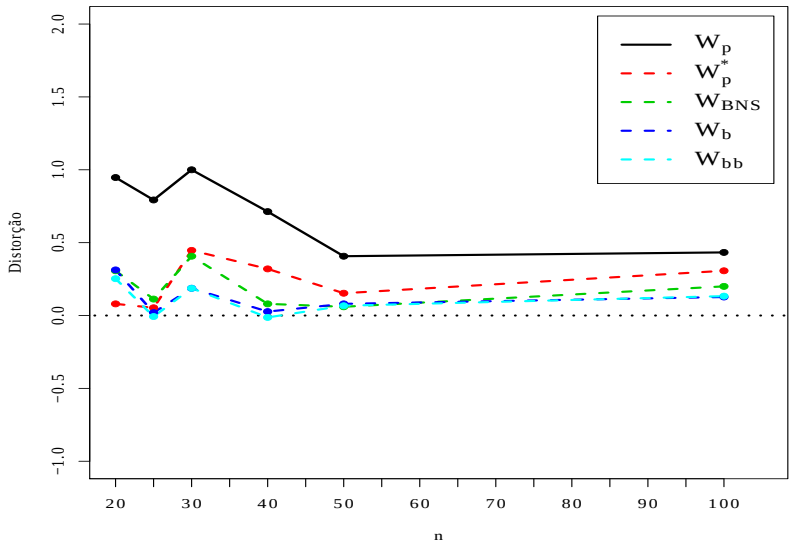
n	γ	W_p	W_p^*	W_{BNS}	W_b	W_{bb}
20	1,0	1,353	1,280	1,047	1,013	1,020
	5,0	5,853	4,887	4,973	4,880	4,813
	10,0	11,353	9,167	9,953	9,980	9,993
25	1,0	1,240	1,167	0,980	1,147	1,100
	5,0	6,027	5,393	5,387	5,127	5,093
	10,0	11,480	9,707	10,500	10,133	10,067
30	1,0	1,333	1,260	1,053	0,907	0,893
	5,0	5,853	5,247	5,233	4,580	4,533
	10,0	11,033	9,580	10,207	9,620	9,527
40	1,0	1,120	1,120	1,047	1,020	1,040
	5,0	5,407	5,067	4,933	5,080	5,013
	10,0	10,347	9,240	9,767	10,280	10,260
50	1,0	1,213	1,193	1,107	0,927	0,907
	5,0	5,373	5,167	4,960	4,960	4,913
	10,0	10,533	9,720	9,987	9,720	9,753
100	1,0	1,127	1,127	1,113	1,033	0,933
	5,0	5,287	5,187	5,133	5,027	5,067
	10,0	10,260	9,853	10,033	9,713	9,693
200	1,0	1,047	1,047	1,027	1,027	1,000
	5,0	4,887	4,853	4,753	4,913	4,900
	10,0	9,913	9,700	9,907	10,100	10,047

Vale enfatizar que os testes corrigidos baseados em W_{BNS} e W_p^* se apresentaram bastante competitivos. A partir dos experimentos vimos que W_{BNS} expôs resultados ligeiramente melhores do que W_p^*. No entanto, dependendo do modelo de interesse W_{BNS} pode ser de difícil obtenção e ter um grande custo computacional (conforme apresentado nas Tabelas A.1 e A.2). Já W_p^* é fácil de ser obtida para qualquer modelo, sem perda de desempenho.

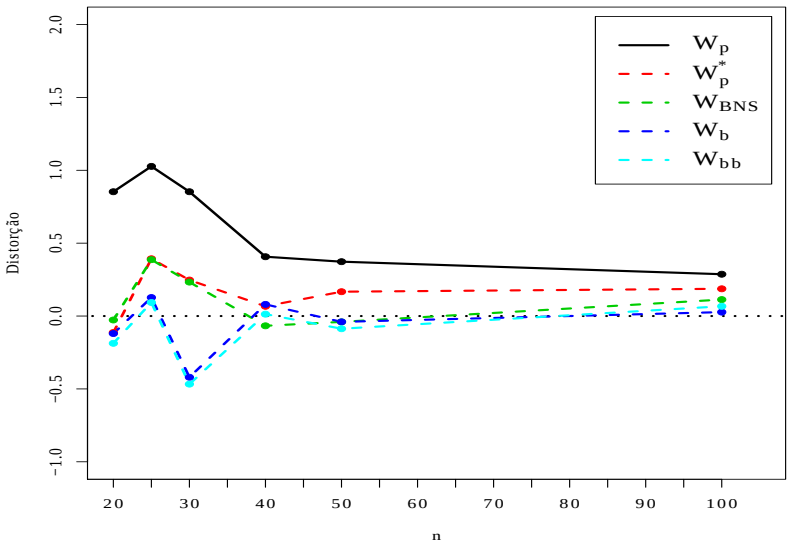
A Figura 5.1 apresenta as distorções dos testes, isto é, a diferença entre a taxa de rejeição estimada e o nível nominal correspondente. Neste caso, consideramos a

situação apresentada nas Tabelas 5.1 e 5.2, em que $\gamma = 5\%$. Note que em geral as taxas das estatísticas W_p^*, W_{BNS}, W_b e W_{bb} apresentam comportamento similares, sendo W_b e W_{bb} as que mais se aproximam da linha de referência, ordenada em zero. Já os testes baseados na estatística W_p expõe a maior distância entre eles e a linha de referência.

Figura 5.1: Distorções dos tamanhos dos testes: $\gamma = 5\%$.



(a) Parâmetro $\beta = 1,0$.



(b) Parâmetro $\beta = 2,5$.

A fim de avaliar o impacto do número de réplicas bootstrap, apresentamos na Tabela 5.3 as taxas de rejeições dos testes baseados em W_b e W_{bb}, considerando $n = 20$, níveis nominais $\gamma = 1\%$, 5% e 10% e o número de réplicas variando em

100, 500, 1000 e 2000. Observamos que para $B = 100$ os testes baseados em W_{bb} apresentaram os melhores resultados, pois as taxas de rejeições são mais próximas do nível nominal especificado. Isso ocorre porque W_{bb} necessita de menos réplicas bootstrap do que W_b para fornecer bons resultados. Nos demais casos, W_b e W_{bb} apresentam comportamento semelhantes.

Tabela 5.3: Comparação entre W_b e W_{bb}, $\beta = 2,5$ e $n = 20$.

B	γ (%)	W_b	W_{bb}
100	1,0	2,100	1,267
	5,0	6,393	5,793
	10,0	11,307	10,747
500	1,0	1,073	0,900
	5,0	4,747	4,747
	10,0	9,667	9,467
1000	1,0	1,073	1,013
	5,0	4,987	4,960
	10,0	9,960	9,927
2000	1,0	1,013	1,020
	5,0	4,880	4,813
	10,0	9,980	9,993

A Tabela 5.4 apresenta os resultados da simulação de poder do teste para $n = 50$, ao nível nominal $\gamma = 5\%$, e variando $\beta = 3,0$; 3,5; 4,0; 4,5; 5,0; 5,5; 6,0 e 6,5. Vale ressaltar que essas simulações correspondem ao cenário apresentado na Tabela 5.2 para $n = 50$ e $\beta = 2,5$. Avaliando os resultados, observamos que os testes baseados nas estatísticas W_p, W_p^*, W_b e W_{bb} são mais poderosos do que os testes baseados na estatística W_{BNS}. Em geral, podemos observar que os testes corrigidos W_p^*, W_{BNS}, W_b e W_{bb} não apresentaram perda significativa de poder do teste.

Tabela 5.4: Poder do teste $\beta = 2,5$, $n = 50$.

β	W_p	W_p^*	W_{BNS}	W_b	W_{bb}
3,0	15,49	15,00	12,97	14,92	14,54
3,5	34,53	33,55	30,42	34,92	34,17
4,0	58,51	57,59	54,08	58,98	58,10
4,5	77,29	76,58	73,61	77,49	76,97
5,0	89,13	88,62	87,28	89,02	88,51
5,5	95,57	95,31	94,66	95,38	95,32
6,0	98,41	98,29	97,91	98,55	98,49
6,5	99,40	99,34	99,13	99,45	99,44

5.2 Testes para a Distribuição Weibull Exponencializada

Nesta seção, apresentamos os resultados numéricos para o parâmetro de interesse λ da distribuição Weibull Exponencializada. Estamos interessados em testar a hipótese nula $H_0 : \lambda = \lambda_0$ *versus* $H_1 : \lambda \neq \lambda_0$. Nas Tabelas 5.5, 5.6 e 5.7 dispomos dos resultados referentes aos cenários em que λ é igual a 1,0, 1,5 e 2,0, respectivamente, considerando o vetor de parâmetros de perturbação $\boldsymbol{\nu}$ fixo em $\boldsymbol{\nu} = (2,0; 3,5)^\top$. É importante ressaltar que os valores das tabelas são exibidos em porcentagens.

Considerando cada tamanho amostral e nível nominal adotado, calculamos as taxas de rejeições empíricas dos testes baseados nas estatísticas W_p, W_p^*, W_{BNS^*}, W_b e W_{bb}, via simulação de Monte Carlo.

Analisando as Tabelas 5.5, 5.6 e 5.7 podemos observar que os testes baseados em W_p e W_p^* mostram-se liberais. Notamos que para tamanho de amostra pequeno a moderado as taxas baseadas em W_p^* atenuam a tendência liberal de W_p, ou seja, apresentam taxas mais próximas do nível nominal especificado. Por exemplo, ao nível nominal de $\gamma = 10\%$, $n = 20$ e $\lambda = 1,0$, os testes baseados em W_p e W_p^* apresentaram taxas de 15,413 e 12,407, respectivamente.

Tabela 5.5: Taxas de rejeições nulas, $\lambda = 1,0$.

n	γ	W_p	W_p^*	W_{BNS^*}	W_b	W_{bb}
20	1,0	0,967	0,807	0,900	1,053	0,820
	5,0	7,713	6,167	6,093	4,940	4,873
	10,0	15,413	12,407	13,093	9,993	10,660
25	1,0	1,107	1,000	1,033	1,147	1,140
	5,0	7,653	6,707	6,173	5,040	5,073
	10,0	14,360	12,100	12,320	10,413	11,040
30	1,0	1,387	1,333	1,120	0,920	1,100
	5,0	7,627	6,927	6,307	4,733	4,920
	10,0	13,873	12,193	11,913	9,593	10,113
40	1,0	1,440	1,387	1,080	1,073	1,147
	5,0	6,773	6,347	5,713	5,047	5,227
	10,0	12,733	11,673	11,333	9,747	9,920
50	1,0	1,380	1,380	1,073	1,107	1,113
	5,0	6,740	6,380	5,873	5,253	5,340
	10,0	12,680	11,707	11,513	10,260	10,340
100	1,0	1,153	1,153	1,047	1,113	1,067
	5,0	5,560	5,500	5,287	5,233	5,167
	10,0	10,707	10,353	10,333	10,327	10,307
200	1,0	0,987	0,987	0,933	1,147	1,120
	5,0	5,360	5,313	5,053	5,067	4,993
	10,0	10,360	10,160	10,173	10,040	9,980

Tabela 5.6: Taxas de rejeições nulas, $\lambda = 1,5$.

n	γ	W_p	W_p^*	W_{BNS^*}	W_b	W_{bb}
20	1,0	1,347	1,180	1,087	1,013	0,813
	5,0	7,900	6,500	6,420	4,873	4,820
	10,0	14,947	12,253	12,940	10,067	10,853
25	1,0	1,427	1,353	1,113	1,153	0,873
	5,0	7,507	6,733	6,347	5,047	5,247
	10,0	13,787	11,873	12,127	10,373	10,913
30	1,0	1,607	1,520	1,180	0,907	0,807
	5,0	7,340	6,687	6,173	4,773	4,933
	10,0	13,140	11,633	11,500	9,787	10,067
40	1,0	1,533	1,480	1,160	1,053	1,027
	5,0	6,460	6,027	5,480	5,087	5,193
	10,0	12,140	11,013	10,993	9,760	9,820
50	1,0	1,347	1,327	1,087	1,093	1,140
	5,0	6,553	6,113	5,687	5,340	5,427
	10,0	12,207	11,153	11,100	10,300	10,267
100	1,0	1,213	1,207	1,087	1,153	1,080
	5,0	5,693	5,553	5,293	5,247	5,207
	10,0	10,633	10,240	10,333	10,347	10,287
200	1,0	1,013	1,033	0,973	1,120	1,060
	5,0	5,240	5,220	5,073	5,067	4,967
	10,0	10,213	9,993	10,020	10,147	10,060

Tabela 5.7: Taxas de rejeições nulas, $\lambda = 2,0$.

n	γ	W_p	W_p^*	W_{BNS^*}	W_b	W_{bb}
20	1,0	1,440	1,307	1,087	1,033	0,863
	5,0	7,607	6,413	6,293	4,900	5,020
	10,0	14,387	11,800	12,607	10,153	10,807
25	1,0	1,487	1,400	1,193	1,147	0,993
	5,0	7,287	6,460	6,160	5,060	5,280
	10,0	13,433	11,487	11,780	10,220	10,753
30	1,0	1,613	1,547	1,207	0,887	0,813
	5,0	7,007	6,440	6,047	4,673	4,893
	10,0	12,700	11,113	11,247	9,833	10,080
40	1,0	1,493	1,467	1,233	1,060	1,027
	5,0	6,233	5,813	5,480	5,040	5,073
	10,0	11,880	10,780	10,807	9,893	9,953
50	1,0	1,313	1,307	1,040	1,107	1,153
	5,0	6,253	5,973	5,507	5,347	5,420
	10,0	11,873	10,860	10,893	10,353	10,340
100	1,0	1,173	1,173	1,093	1,187	1,107
	5,0	5,513	5,413	5,220	5,267	5,167
	10,0	10,587	10,107	10,087	10,273	10,180
200	1,0	1,167	1,167	1,147	1,167	1,107
	5,0	5,233	5,220	5,113	5,127	5,047
	10,0	10,067	9,887	9,947	10,187	10,160

Avaliando as Tabelas 5.5, 5.6 e 5.7 notamos que as taxas baseadas em W_{BNS^*} também atenuaram a tendência liberal de W_p. Embora o desempenho baseado em W_{BNS^*} ainda seja liberal, de modo geral, as taxas baseadas em W_{BNS^*} apresentaram desempenho superior a W_p^*. Por exemplo, para o nível nominal $\gamma = 5\%$, $n = 40$ e $\lambda = 1,5$, os testes baseados em W_p, W_p^* e W_{BNS^*} apresentaram taxas iguais a 6,460, 6,027 e 5,480, respectivamente.

Já para os testes baseados W_b e W_{bb} apresentaram taxas semelhantes. Além disso, observamos que para todos os valores de λ testados, W_b e W_{bb} obtiveram desempenho superior aos testes baseados em W_p, W_p^* e W_{BNS^*}, principalmente, para os níveis nominais 5% e 10%. Por exemplo, para $n = 25$, $\gamma = 5\%$ e $\lambda = 1,5$, W_b e W_{bb}, apresentaram taxas iguais a 5,047 e 5,270, respectivamente, enquanto, W_p, W_p^* e W_{BNS^*} obtiveram taxas iguais a 7,507, 6,733 e 6,340, respectivamente.

É possível observar também que conforme aumentamos o tamanho amostral, as taxas baseadas em W_p, W_p^* e W_{BNS^*} se aproximam do nível nominal especificado. Além disso, em todos os cenários testados, podemos perceber que os testes corrigidos (W_p^*, W_{BNS^*}, W_b e W_{bb}) demonstraram melhores resultados do que o não corrigido (W_p).

Figura 5.2: Distorções dos tamanhos dos testes, $\gamma = 5\%$ e $\lambda = 1,0$.

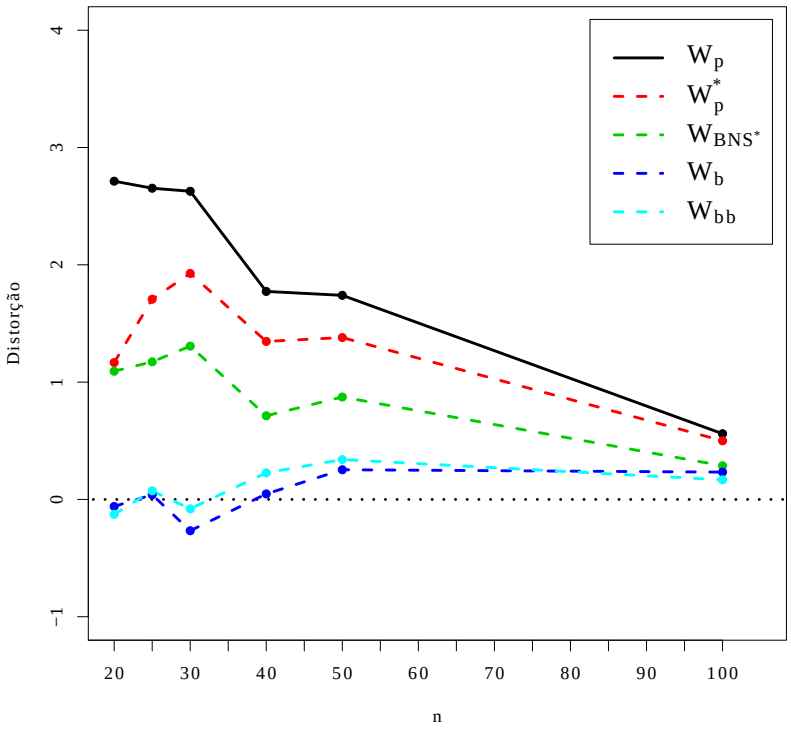


Figura 5.3: Distorções dos tamanhos dos testes, $\gamma = 5\%$ e $\lambda = 1, 5$.

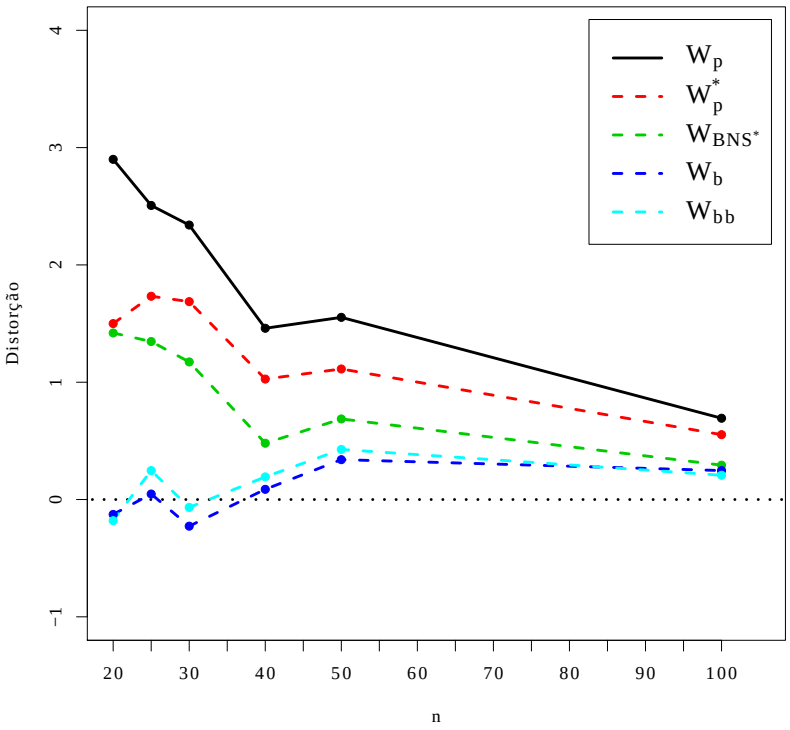
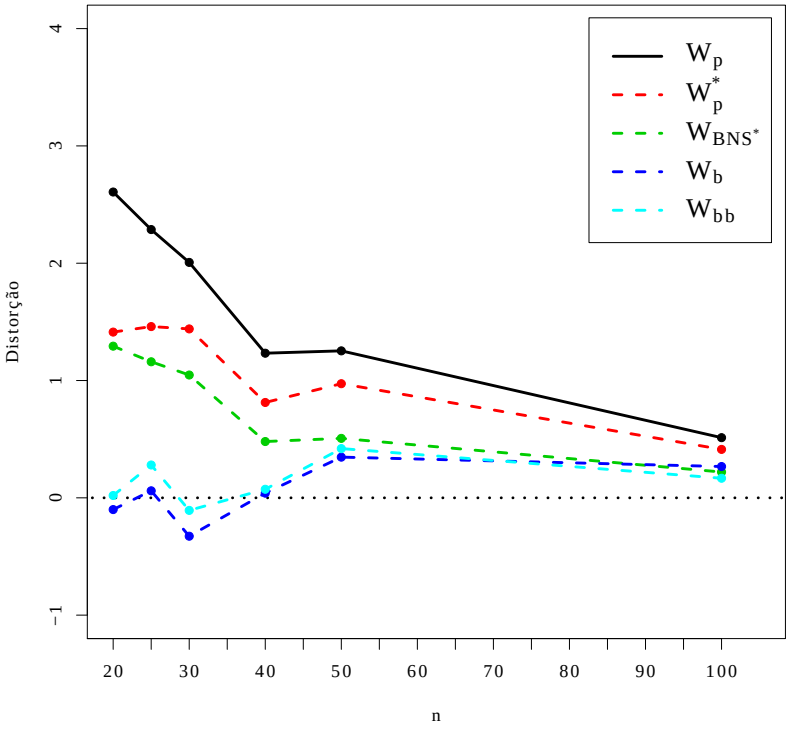


Figura 5.4: Distorções dos tamanhos dos testes, $\gamma = 5\%$ e $\lambda = 2, 0$.



As Figuras 5.2, 5.3 e 5.4, apresentam as distorções dos testes, que caracteriza a diferença entre a taxa de rejeição estimada e o nível nominal especificado. Para isso, consideramos os cenários expostos nas Tabelas 5.5, 5.6 e 5.7, em que $\gamma = 5\%$. Observe que para n de tamanho pequeno a moderado os testes baseados nas estatísticas corrigidas apresentaram melhores resultados, sendo W_b e W_{bb} as que mais se aproximam da linha de referência, ordenada em zero.

A Tabela 5.8 apresenta os resultados da simulação de poder do teste para $n = 200$, ao nível nominal $\gamma = 10\%$, e variando $\lambda = 1{,}5;\ 2{,}0;\ 2{,}5;\ 3{,}0;\ 3{,}5;\ 4{,}0$ e $4{,}5$. Essas simulações correspondem ao cenário exposto na Tabela 5.5 para $n = 200$ e $\lambda = 1,0$. Notamos que, de modo geral, os testes baseados em W_p, $\mathrm{W}_{\mathrm{BNS}^*}$ e W_{bb} apresentaram poderes semelhantes e ligeiramente superiores a W_p^* e W_b. Além disso, é possível notar que os testes corrigidos W_p^*, $\mathrm{W}_{\mathrm{BNS}^*}$, W_b e W_{bb} não apresentaram perda significativa de poder do teste.

Tabela 5.8: Poder do teste $\lambda = 1,0$, $n = 200$.

λ	W_p	W_p^*	$\mathrm{W}_{\mathrm{BNS}^*}$	W_b	W_{bb}
1,5	32,060	31,713	32,487	32,053	31,873
2,0	61,793	61,447	62,060	61,507	61,433
2,5	82,700	82,480	82,713	80,100	80,013
3,0	92,453	92,340	92,520	89,300	89,333
3,5	94,153	94,047	94,193	94,053	94,040
4,0	96,793	96,727	96,827	96,587	96,560
4,5	98,067	98,047	98,120	97,973	97,967

5.3 Exemplos Numéricos

Nesta seção iremos realizar quatro exemplos numéricos a partir de quatro amostras completas. No primeiro e segundo caso, iremos considerar que as observações são independentes e seguem distribuição $\mathrm{LP}(\mu, \beta)$. Já no terceiro e quarto caso iremos assumir que as observações são independentes e seguem distribuição $\mathrm{WE}(a, b, \lambda)$.

Com intuito de verificar a aderência dos dados à distribuição que assumimos, realizamos o teste de Kolmogorov - Smirnov (KS) para cada conjunto de dados. Para isso, utilizamos o pacote Adequacy Model proposto por Marinho et al. (2019), disponível na linguagem R.

Para o primeiro exemplo, considerados o conjunto de dados dispostos na Tabela 5.9, que representa 24 observações referentes a tempos de falha de componentes mecânicos disponível em Murthy et al. (2004). No segundo exemplo, utilizamos os dados expostos na tabela 5.10, que corresponde a 23 observações de tempos de falha por fadiga de rolamento de esferas disponível em Lawless (2011).

A partir dos dados apresentados em 5.9 e 5.10 realizamos o teste de KS considerando as seguintes hipóteses H_0: os dados seguem distribuição Lindley Ponderada *versus* H_1: H_0 é falso. Para o primeiro conjunto de dados obtivemos a estatística e o p-valor, iguais a 0,141 e 0,676, respectivamente, isso indica que aos níveis nominais usuais (1%, 5% e 10%), H_0 não é rejeitada. Já para o segundo conjunto de dados obtivemos a estatística e o p-valor, iguais a 0,123 e 0,879, respectivamente, isso significa que aos níveis nominais usuais, H_0 não é rejeitada.

No primeiro e segundo exemplo (Tabelas 5.9 e 5.10) o interesse é testar a hipótese nula H_0: $\beta = 1$ *versus* H_1: $\beta \neq 1$, ou seja, estamos testando o modelo Lindley *versus* Lindley Ponderada.

No primeiro exemplo (Tabela 5.9), obtivemos W_p, W_p^* e W_{BNS^*} iguais a 11,832, 11,832 e 10,729, respectivamente, e seus correspondentes p-valores iguais a 0,001. Dessa forma, aos níveis nominais usuais os testes baseados em W_p, W_p^* e W_{BNS^*} rejeitam a hipótese nula, ou seja, o modelo Lindley é rejeitado.

Tabela 5.9: Dados I - Tempos de falha de componentes mecânicos.

30,94	18,51	16,62	51,56	22,85	22,38	19,08	49,56
17,12	10,67	25,43	10,24	27,47	14,70	14,10	29,93
27,98	36,02	19,40	14,97	22,57	12,26	18,14	18,84

No segundo exemplo (Tabela 5.10), obtivemos W_p, W_p^* e W_{BNS^*} iguais a 5,296, 5,296 e 4,597, respectivamente, e seus correspondentes p-valores são 0,0214, 0,0237 e 0,032. Dessa forma, ao nível de nominal de 1% os testes baseados em W_p, W_p^* e W_{BNS^*} não rejeitam a hipótese nula, ou seja, o modelo Lindley não é rejeitado. Já aos níveis nominais de 5% e 10% os testes rejeitam a hipótese nula, ou seja, o modelo Lindley é rejeitado.

Tabela 5.10: Dados II - Tempos de falha por fadiga de rolamentos de esferas.

17,23	28,92	33,00	41,52	42,12	45,60	48,80	51,84
51,96	54,12	55,56	67,80	68,64	68,64	68,88	84,12
93,12	98,64	105,12	105,84	127,92	128,04	173,40	

Para o terceiro exemplo, consideramos o conjunto de dados expostos na Tabela 5.11, que consiste em 15 observações referentes a tempos entre falhas sucessivas de equipamentos de ar condicionado em uma aeronave Boeing 720 disponível em Lawless (2011). Já para o quarto exemplo, consideramos o conjunto de dados expostos na Tabela 5.12, referente a 30 observações de fluxos anuais do rio Weldon em Mill Grove disponível em Vujica (1972).

Com os dados expostos nas Tabelas 5.11 e 5.12 realizamos o teste de KS considerando as seguintes hipóteses H_0: os dados seguem distribuição Weibull Exponencializada *versus* H_1: H_0 é falso. Para o terceiro conjunto de dados obtivemos a

Tabela 5.11: Dados III - Tempos entre falhas sucessivas de equipamentos de ar condicionado em uma aeronave Boeing 720.

74,0	57,0	48,0	29,0	502,0	12,0	70,0	21,0
29,0	386,0	59,0	27,0	153,0	26,0	326,0	

estatística e o p-valor, iguais a 0,167 e 0,800, respectivamente, isso indica que aos níveis nominais usuais, H_0 não é rejeitada. Já para o quarto conjunto de dados obtivemos a estatística e o p-valor, iguais a 0,117 e 0,761, respectivamente, isso significa que aos níveis nominais usuais, H_0 não é rejeitada.

No terceiro e quarto conjunto de dados (Tabelas 5.11 e 5.12) o interesse é testar a hipótese nula H_0: $\lambda = 1$ *versus* H_1: $\lambda \neq 1$, ou seja, estamos testando o modelo Weibull *versus* Weibull Exponencializada.

Tabela 5.12: Dados IV - Fluxos anuais do rio Weldon em Mill Grove.

108,0	472,0	143,0	441,0	244,0	132,0	53,6	96,5	93,7	386,0
400,0	44,0	585,0	217,0	398,0	567,0	245,0	72,5	98,1	42,7
298,0	122,0	114,0	135,0	40,6	208,0	248,0	151,0	659,0	635,0

Para o terceiro exemplo (Tabela 5.11), temos W_p = 5,014, W_p^* = 5,014 e W_{BNS^*} = 3,659 e os correspondentes p-valores são 0,025, 0,031 e 0,055. Dessa forma, ao nível de significância de 5%, os testes baseados em W_p e W_p^*, rejeitam a hipótese nula, enquanto o teste baseado em W_{BNS^*} não rejeita a hipótese nula.

Para o quarto exemplo (Tabela 5.12), temos W_p = 0,792, W_p^* = 0,792 e W_{BNS^*} = 0,420 e os correspondentes p-valores são 0,373, 0,491 e 0,517. Dessa forma, aos níveis de significância usuais , os testes baseados em W_p e W_p^* e W_{BNS^*}, não rejeitam a hipótese nula, ou seja, o modelo Weibull não é rejeitado.

Capítulo 6

Conclusões

Neste trabalho, foram realizadas inferências por meio de testes de hipóteses para um parâmetro de interesse em duas distribuições de probabilidade contínuas (Lindley Ponderada e Weibull Exponencializada). A partir disso, foram derivados ajustes para a função de verossimilhança perfilada, a fim de atenuar o impacto dos parâmetros de perturbação em amostras de tamanho pequeno a moderado.

O estudo foi realizado via simulação de Monte Carlo para comparar o desempenho da função de verossimilhança perfilada e suas versões ajustadas com o método de aperfeiçoamento proposto por Sousa (2020), em amostras de tamanho pequeno a moderado. Para isso, foram feitos testes de hipóteses como base na estatística da razão de verossimilhanças. Além disso, foram considerados também os testes aperfeiçoados via bootstrap paramétrico e Bartlett bootstrap.

Os resultados apontaram que para amostras de tamanho pequeno a moderado as taxas de rejeições nulas obtidas por meio dos testes baseados na metodologia de Sousa (2020) e na estatística da razão de verossimilhanças que utiliza as versões modificadas apresentaram melhores desempenhos. No entanto, para grandes amostras todas as estatísticas de teste se mostraram equivalentes. Já os testes baseados em bootstrap paramétrico e Bartlett bootstrap obtiveram resultados satisfatórios em todos os cenários utilizados.

Por fim, foram apresentados dois exemplos numéricos para cada distribuição de probabilidade considerada, com base em conjunto de dados reais.

Referências Bibliográficas

BARNDORFF-NIELSEN, O., 1983, *On a formula for the distribution of the maximum likelihood estimator*, v. 70. Oxford University Press.

BARNDORFF-NIELSEN, O. E., 1995, *Stable and invariant adjusted profile likelihood and directed likelihood for curved exponential models*, v. 82. Oxford University Press.

BAYER, F. M., CRIBARI-NETO, F., 2013, "Bartlett corrections in beta regression models", *Journal of Statistical Planning and Inference*, v. 143, n. 3, pp. 531–547.

BOLFARINE, H., SANDOVAL, M. C., 2001, *Introdução à inferência estatística*, v. 2. SBM.

CORDEIRO, G. M., CRIBARI-NETO, F., 2014, *An introduction to Bartlett correction and bias reduction*. Springer.

CORDEIRO, G., 1999, *Introdução à Teoria Assintótica*. 22o Colóquio Brasileiro de Matemática, IMPA.

COX, D. R., REID, N., 1987, *Parameter orthogonality and approximate conditional inference*, v. 49. Wiley Online Library.

COX, D., REID, N., 1993, *A note on the calculation of adjusted profile likelihood*, v. 55. Wiley Online Library.

EFRON, B., 1979, *Bootstrap methods: another look at the jackknife*. The annals of Statistics 7(1).

FORERO, S. L., 2015, *Teste de razão de verossimilhanças Bootstrap e técnicas de diagnóstico em modelos em séries de potências não-lineares generalizados*. Dissertação de Mestrado, Universidade Federal de Pernambuco.

GHITANY, M., OTHERS, 2011, *A two-parameter weighted Lindley distribution and its applications to survival data*, v. 81. Elsevier.

JAMES, B. R., 1996, *Probabilidade: um curso em nível intermediário*. N. 519.2.

JOHNSON, S. G., 2020, "The NLopt nonlinear-optimization package", .

LAWLESS, J. F., 2011, *Statistical models and methods for lifetime data*, v. 362. John Wiley & Sons.

MAGALHÃES, M. N., 2006, *Probabilidade e variáveis aleatórias*. Edusp.

MARINHO, P. R. D., OTHERS, 2019, "AdequacyModel: An R package for probability distributions and general purpose optimization", *PloS one*, v. 14, n. 8, pp. e0221487.

MAZUCHELI, J., OTHERS, 2013, *Comparison of estimation methods for the parameters of the weighted Lindley distribution*, v. 220. Elsevier.

MUDHOLKAR, G. S., SRIVASTAVA, D. K., 1993, *Exponentiated Weibull family for analyzing bathtub failure-rate data*, v. 42. IEEE.

MURTHY, D. P., OTHERS, 2004, *Weibull models*, v. 505. John Wiley & Sons.

R CORE TEAM, 2019, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org/>.

RAMOS, P., OTHERS, 2017, *Maximum Likelihood Estimation for the Weight Lindley Distribution Parameters under Different Types of Censoring*. Revista Brasileira de Biometria.

ROCKE, D. M., 1989, *Bootstrap Bartlett adjustment in seemingly unrelated regression*, v. 84. Taylor & Francis.

SEVERINI, T. A., 1998, *An approximation to the modified profile likelihood function*, v. 85. Oxford University Press.

SEVERINI, T. A., 2000, *Likelihood methods in statistics*. Oxford University Press.

SOUSA, A. R. D., 2020, *Distribuição qui-Quadrado inf: Uma nova abordagem para o aperfeiçoamento do teste da razão de verossimilhanças.* Dissertação de mestrado, Universidade Federal da Paraíba.

TABLADA, C. J., 2017, *Generalized probability distributions for lifetime applications.* Tese de doutorado, Universidade Federal de Pernambuco.

VUJICA, Y., 1972, *Probability and Statistics in Hydrology.* Water Resources Publication.

Apêndice A

Tempo de simulações

A.1 Lindley Ponderada

Na Tabela A.1 apresentamos os tempos de simulações referente ao caso $\beta = 1,0$. Para o caso $\beta = 2,5$, os tempos de simulações se mostraram equivalentes aos obtidos em A.1.

Tabela A.1: Tempo de simulações em segundos para $\beta = 1,0$.

n	γ	W_p e W_p^*	W_{BNS}	W_b	W_{bb}
20	1,0	9,053	4,406	2376,175	2301,775
	5,0	8,167	4,121	2337,431	2407,301
	10,0	7,910	4,043	2366,176	2748,409
25	1,0	8,931	4,129	2407,749	2310,648
	5,0	8,303	4,111	2379,284	2387,424
	10,0	7,925	4,144	2351,116	2794,367
30	1,0	8,563	4,240	2341,069	2294,175
	5,0	8,256	4,271	2342,343	2419,252
	10,0	7,963	4,174	2377,056	2791,643
40	1,0	8,493	4,413	2441,239	2393,753
	5,0	8,423	4,422	2399,759	2501,340
	10,0	8,319	4,381	2427,765	2860,493
50	1,0	9,091	4,523	2438,640	2358,362
	5,0	8,469	4,499	2407,866	2469,008
	10,0	8,536	4,469	2400,972	2857,351
100	1,0	10,097	6,266	3177,713	3111,471
	5,0	9,747	6,271	3152,338	3248,511
	10,0	9,819	6,244	3139,753	3673,560
200	1,0	12,148	9,860	4282,489	4067,209
	5,0	11,071	9,239	4183,902	4554,343
	10,0	11,757	9,147	4177,654	4888,588

A.2 Weibull Exponencializada

Na Tabela A.2 apresentamos os tempos de simulações referente ao caso $\lambda = 1,0$. Para o caso $\lambda = 1,5$ e $\lambda = 2,0$, os tempos de simulações se mostraram equivalentes aos obtidos em A.2.

Tabela A.2: Tempo de simulações em segundos para λ = 1,0.

n	γ	W_p, W_p^* e W_{BNS^*}	W_b e W_{bb}
20	1,0	2080,473	21503,788
	5,0	2099,757	21164,291
	10,0	2101,598	21405,407
25	1,0	1705,181	20060,192
	5,0	1717,380	20159,308
	10,0	1728,070	20741,643
30	1,0	1533,789	19916,255
	5,0	1579,027	20614,175
	10,0	1550,563	20354,046
40	1,0	1489,676	20143,651
	5,0	1543,991	20473,619
	10,0	1506,754	20455,328
50	1,0	1620,477	20994,549
	5,0	1677,417	21422,461
	10,0	1660,927	21083,803
100	1,0	2816,080	28462,687
	5,0	2845,541	28863,614
	10,0	2884,392	28478,291
200	1,0	3036,390	45823,919
	5,0	3411,660	46771,573
	10,0	3848,170	45750,521