



## ESCOLHA ENTRE MODELOS SEPARADOS: UMA COMPARAÇÃO ENTRE O TESTE DE SIGNIFICÂNCIA BAYESIANO E O TESTE DE COX

Cachimo Combo Assane

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Produção.

Orientadores: Basílio de Bragança Pereira  
Carlos Alberto de Bragança  
Pereira  
Júlio Michael Stern

Rio de Janeiro  
Março de 2017

ESCOLHA ENTRE MODELOS SEPARADOS: UMA COMPARAÇÃO ENTRE  
O TESTE DE SIGNIFICÂNCIA BAYESIANO E O TESTE DE COX

Cachimo Combo Assane

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR  
EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Examinada por:

---

Prof. Basílio de Bragança Pereira, Ph.D.

---

Prof. Carlos Alberto de Bragança Pereira, Ph.D.

---

Prof. Hélio dos Santos Migon, Ph.D.

---

Prof. Emília Matos do Nascimento, D.Sc.

---

Prof. Edilson Fernandes Arruda, D.Sc.

---

Prof. Thais Cristina Oliveira da Fonseca, Ph.D.

---

Prof. Claudio José Struchiner, Ph.D.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2017

Assane, Cachimo Combo

Escolha entre modelos separados: uma comparação entre o teste de significância Bayesiano e o teste de Cox/Cachimo Combo Assane. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIII, 74 p.: il.; 29, 7cm.

Orientadores: Basílio de Bragança Pereira

Carlos Alberto de Bragança Pereira

Júlio Michael Stern

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Produção, 2017.

Referências Bibliográficas: p. 66 – 74.

1. Seleção de modelos. 2. Hipóteses separadas. 3. Modelo de misturas. 4. Teste de significância. 5. O teste FBST. 6. Medida de evidência e-valor. 7. Teste de Cox. I. Pereira, Basílio de Bragança *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

*Ao meu filho, Akil Assane, e à  
minha esposa, Lindacir Schwarz.  
Que ALLAH os abençoe!!*

***“Lemma:*** *We have to learn to  
live with uncertainty.*

***Theorem:*** *We can model  
uncertainty.*

***Corollary:*** *We can make a  
living out of uncertainty”.*

*(Yaakov Bar-Shalom and X. Li)*

# Agradecimentos

Em primeiro lugar, quero louvar e agradecer a Deus, o Clemente, o Misericordioso, por ter-me concedido saúde, inteligência e autoconfiança suficientes para concretizar o desejo de ser Doutor.

Ao professor Basílio de Bragança Pereira, a quem manifesto a minha gratidão e o meu profundo apreço pela receptividade, amizade e, principalmente, por todo suporte concedido durante os seis anos que estou aqui no Brasil. Agradeço por ter acreditado no meu potencial e por estar contribuindo imensamente no meu crescimento acadêmico e pessoal, através das suas orientações e conselhos.

Ao professor Carlos Alberto de Bragança Pereira, pelo incentivo e por compartilhar os seus conhecimentos através da sua imprescindível orientação.

À toda minha família, em especial aos meus pais, Combo Assane e Nema Buana, e ao meu avô Assane Combo, pela compreensão, apoio e, certamente, pelos preces em prol do meu sucesso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento dos meus estudos.

Aos professores da PO/PEP/COPPE e do Departamento de Métodos Estatísticos do IM/UFRJ. Em especial, ao Hélio dos Santos Migon e ao Edilson Fernandes Arruda, pelo incentivo e pelas relevantes sugestões que contribuíram para a melhoria da tese.

À professora Emília Matos do Nascimento, por quem tenho grande respeito e admiração. Agradeço pela amizade, carinho e, principalmente, pelo apoio incondicional que sempre me prestou.

Aos colegas da seção de Estatística, Departamento de Matemática e Informática da Universidade Eduardo Mondlane, Moçambique, em especial, ao Dr Lino Marques, pelo incentivo e voto de confiança, pelo suporte e, acima de tudo, pela sua flexibilidade nos momentos em que precisei de auxílio para resolver meus problemas profissionais.

À coordenação do programa, em especial à Andréia, à Roberta, ao Pedrinho e à Cleudete, pelo suporte e auxílio nas questões burocráticas.

Ao Brian Alvarez Ribeiro de Melo por ter me ajudado a implementar o modelo de mistura.

Ao Dr. Mauro Alves por ter cedido o banco de dados de hemodiálise. Aos meus amigos Manuel Ramon e Rubens Oliveira, pelo apoio e importantes contribuições. Ao meu amigo, irmão, colega de batalha, Miranda Muaualo, pelo companheirismo e pelos ensinamentos compartilhados nos últimos seis anos. Ao meu tio Mussa Mbeve, a quem manifesto o meu especial agradecimento, pelo seu imprescindível suporte durante o período da minha formação.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## ESCOLHA ENTRE MODELOS SEPARADOS: UMA COMPARAÇÃO ENTRE O TESTE DE SIGNIFICÂNCIA BAYESIANO E O TESTE DE COX

Cachimo Combo Assane

Março/2017

Orientadores: Basílio de Bragança Pereira  
Carlos Alberto de Bragança Pereira  
Júlio Michael Stern

Programa: Engenharia de Produção

O problema de testes para famílias separadas de hipóteses foi inicialmente considerado por COX (1961, 1962). Neste trabalho, propõe-se o teste de significância totalmente Bayesiano, o FBST, para discriminar entre os modelos lognormal, gama e Weibull cujas famílias de distribuições são separadas. Aqui, o problema é abordado num contexto de mistura linear dos modelos candidatos. A novidade nesta abordagem é que a relação entre os parâmetros desses modelos é explorada através da reparametrização das respectivas funções densidades em termos de parâmetros comuns: média,  $\mu$ , e variância,  $\sigma^2$ , da população. Uma vantagem desta reparametrização, do ponto de vista computacional, é a de reduzir a quantidade de parâmetros a serem estimados. O procedimento do FBST é utilizado para testar as hipóteses sobre os pesos da mistura, com o objetivo de calcular as evidências a favor de cada modelo.

Para avaliar o desempenho do método proposto, alguns resultados numéricos baseados em simulações de pontos amostrais são apresentados. Nessas simulações, os resultados do FBST são comparados com os do teste de Cox. Três exemplos de aplicações ilustrando os procedimentos para conjunto de dados não censurados também são apresentados. Além disso, o modelo de mistura lognormal-gama-Weibull foi aplicado à base de dados de pacientes com insuficiência renal crônica terminal submetidos à hemodiálise no Rio de Janeiro, e a medida de evidência do FBST utilizada para escolher o modelo com melhor ajuste a esses dados.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

MODEL CHOICE IN SEPARATE FAMILIES: A COMPARISON BETWEEN  
THE BAYESIAN SIGNIFICANCE TEST AND THE COX TEST

Cachimo Combo Assane

March/2017

Advisors: Basílio de Bragança Pereira  
Carlos Alberto de Bragança Pereira  
Júlio Michael Stern

Department: Production Engineering

Tests of separate families of hypotheses were initially considered by COX (1961, 1962). In this work, the Fully Bayesian Significance Test, FBST, is proposed for discriminating between the lognormal, gamma and Weibull models whose families of distributions are separate. This problem is analyzed in the context of the linear mixture of the candidate models. The novelty of this approach is that the relationship between the parameters of these models is explored by reparametrizing their respective density functions in terms of common parameters: the mean,  $\mu$ , and the variance,  $\sigma^2$ , of the population. An advantage of this reparametrization, from a computational point of view, is that it allows one to reduce the number of the parameters to be estimated. The FBST procedure is used to test the hypotheses on the mixture weights in order to calculate the evidence in favor of each model.

In order to evaluate the performance of the proposed method, some numerical results based on simulations of sample points are given. In these simulations, the results of FBST are compared with those of the Cox test. Three application examples illustrating the procedures for uncensored dataset are also presented. In addition, the lognormal-gamma-Weibull mixture model was applied to the database of patients with end-stage chronic kidney failure who received hemodialysis in Rio de Janeiro and the Bayesian evidence measure was used to choose the best fit model.

# Sumário

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	4
1.2 Organização da tese . . . . .	4
<b>2 Testes para famílias separadas de hipóteses</b>	<b>6</b>
2.1 Introdução . . . . .	6
2.2 Testes de Cox . . . . .	8
2.3 Fator de Bayes . . . . .	9
2.3.1 Fator de Bayes a posteriori . . . . .	10
2.3.2 Fator de Bayes parcial . . . . .	11
2.3.3 Fator de Bayes fracionário . . . . .	11
2.3.4 Fator de Bayes intrínseco . . . . .	12
2.4 Notas Bibliográficas . . . . .	12
<b>3 O FBST–Teste de Significância Totalmente Bayesiano</b>	<b>14</b>
3.1 Introdução . . . . .	14
3.2 Modelo Estatístico Bayesiano . . . . .	15
3.3 Formulação original do FBST . . . . .	16
3.4 Formulação invariante do FBST . . . . .	18
3.5 Nível crítico para rejeição da hipótese $H$ . . . . .	19
3.5.1 Funções de perda . . . . .	19
3.5.2 Análise assintótica . . . . .	20
3.6 Notas Bibliográficas . . . . .	22
<b>4 Mistura de modelos separados</b>	<b>24</b>
4.1 Introdução . . . . .	24
4.2 Análise de sobrevivência e suas funções . . . . .	25
4.2.1 Definições . . . . .	25

4.2.2	Distribuições de probabilidade de interesse . . . . .	26
4.3	Distribuições Multinomial e Dirichlet . . . . .	27
4.4	Modelo de mistura finita . . . . .	30
4.5	O FBST para mistura de modelos separados . . . . .	34
<b>5</b>	<b>Simulações e Exemplos de aplicações</b>	<b>38</b>
5.1	Simulações . . . . .	38
5.1.1	Esquema para simulação de pontos amostrais . . . . .	38
5.1.2	Resultados das simulações . . . . .	40
5.2	Exemplos de aplicações . . . . .	46
<b>6</b>	<b>Análise de modelo de mistura finita com dados censurados</b>	<b>54</b>
6.1	Introdução . . . . .	54
6.2	Simulações . . . . .	55
6.2.1	Esquema para simulação de pontos amostrais . . . . .	55
6.2.2	Avaliação do desempenho do FBST . . . . .	56
6.2.3	Resultados das simulações . . . . .	57
6.3	Aplicação: Escolha de um modelo de sobrevivência para pacientes com insuficiência renal crônica terminal . . . . .	59
6.3.1	Base de dados . . . . .	59
6.3.2	Resultados . . . . .	60
<b>7</b>	<b>Conclusões</b>	<b>63</b>
7.1	Propostas para trabalhos futuros . . . . .	65
7.2	Trabalhos resultantes da tese . . . . .	65
	<b>Referências Bibliográficas</b>	<b>66</b>

# Lista de Figuras

5.1	Curvas de sobrevivência estimadas pelo estimador PEXE e pelos modelos Weibull e de mistura LGW para dados de ciclos até a falha do fio de poliéster . . . . .	49
5.2	Curvas de sobrevivência estimadas por PEXE e pelos modelos lognormal e gama para dados de larvas do grupo A . . . . .	51
5.3	Curvas de sobrevivência estimadas por PEXE e pelos modelos Weibull e de mistura LGW para dados de latência do HIV em adultos . . . . .	53
6.1	Densidades das distribuições Gama, Weibull e Lognormal para $\mu = 20$ e $\sigma^2 = 50$ . . . . .	59
6.2	Curvas de sobrevivência estimadas pelo estimador PEXE e pelos modelos lognormal e de mistura LGW para dados de hemodiálise . . . . .	62

# Lista de Tabelas

5.1	Valores dos parâmetros das distribuições utilizadas para geração das amostras . . . . .	39
5.2	Taxas de aceitação da hipótese nula verdadeira, $H_L$ . Testes para lognormal <i>vs.</i> Weibull usando dados gerados da lognormal(0,1). . . . .	41
5.3	Taxas de rejeição da hipótese nula falsa, $H_W$ . Testes para lognormal <i>vs.</i> Weibull usando dados gerados da lognormal(0,1). . . . .	41
5.4	Taxas de aceitação da hipótese nula verdadeira, $H_W$ . Testes para lognormal <i>vs.</i> Weibull usando dados gerados da Weibull(1,1). . . . .	41
5.5	Taxas de Rejeição da hipótese nula falsa, $H_L$ . Testes para lognormal <i>vs.</i> Weibull usando dados gerados da Weibull(1,1). . . . .	41
5.6	Taxas de aceitação da hipótese nula $H_L$ com base no FBST e no teste de Cox, usando dados gerados da lognormal. . . . .	43
5.7	Taxas de rejeição da hipótese nula $H_G$ com base no FBST e no teste de Cox, usando dados gerados da lognormal. . . . .	43
5.8	Taxas de aceitação da hipótese nula $H_G$ com base no FBST e no teste de Cox, usando dados gerados da gama. . . . .	43
5.9	Taxas de rejeição da hipótese nula $H_L$ com base no FBST e no teste de Cox, usando dados gerados da gama. . . . .	44
5.10	Taxas de aceitação da hipótese nula $H_W$ com base no FBST e no teste de Cox, usando dados gerados da Weibull. . . . .	45
5.11	Taxas de rejeição da hipótese nula $H_G$ com base no FBST e no teste de Cox, usando dados gerados da Weibull. . . . .	46
5.12	Taxas de aceitação da hipótese nula $H_G$ com base no FBST e no teste de Cox, usando dados gerados da gama. . . . .	46
5.13	Taxas de rejeição da hipótese nula $H_W$ com base no FBST e no teste de Cox, usando dados gerados da Weibull. . . . .	46
5.14	Evidências fornecidas pelos dados de ciclos até a falha do fio de poliéster . . . . .	47
5.15	Sumário da distribuição a posteriori dos parâmetros do modelo LGW, aplicado aos dados de ciclos até a falha do fio de poliéster . . . . .	48

5.16	Testes de hipóteses para os pesos da mistura LGW usando dados de ciclos até a falha do fio de poliéster . . . . .	49
5.17	Sumário da distribuição a posteriori dos parâmetros do modelo Weibull, aplicado aos dados de ciclos até a falha do fio de poliéster . . . .	49
5.18	Evidências fornecidas pelos dados de larvas do grupo A . . . . .	50
5.19	Sumário da distribuição a posteriori dos parâmetros do modelo lognormal, aplicado aos dados de larvas do grupo A. . . . .	51
5.20	Sumário da distribuição a posteriori dos parâmetros do modelo gama, aplicado aos dados de larvas do grupo A . . . . .	51
5.21	Evidências fornecidas pelos dados de latência do HIV em adultos . . . . .	52
5.22	Sumário da distribuição a posteriori dos parâmetros do modelo LGW, aplicado aos dados de latência do HIV em adultos . . . . .	53
5.23	Testes de hipóteses para os pesos da mistura LGW aplicado aos dados de latência do HIV em adultos . . . . .	53
5.24	Sumário da distribuição a posteriori dos parâmetros do modelo Weibull, aplicado aos dados de latência do HIV em adultos . . . .	53
6.1	Estimativas do modelo LGW e taxas de acerto do FBST na escolha da distribuição geradora dos dados, para amostras com diferentes porcentagens de censura . . . . .	58
6.2	Evidências fornecidas pelos dados de hemodiálise . . . . .	60
6.3	Sumário da distribuição a posteriori dos parâmetros do modelo LGW aplicado aos dados de hemodiálise . . . . .	61
6.4	Testes de hipóteses para os pesos da mistura do modelo LGW aplicado aos dados de hemodiálise . . . . .	61
6.5	Sumário da distribuição a posteriori dos parâmetros do modelo lognormal aplicado aos dados de hemodiálise . . . . .	61

# Capítulo 1

## Introdução

A escolha entre modelos estatísticos alternativos é um problema importante em análise estatística, especialmente na modelagem da distribuição do tempo até a ocorrência de uma falha (KUNDU e MANGLICK, 2004, 2005; LAWLESS, 2002; PEREIRA, 1978, 1979). A teoria de teste de hipóteses de Neyman-Pearson pode ser aplicada para comparar modelos pertencentes à mesma família de distribuições (BUSE, 1982). Alternativamente, procedimentos especiais são necessários quando os modelos candidatos pertencem a famílias separadas, no sentido de que um membro arbitrário de uma família não pode ser obtido como limite dos membros de uma outra. Problemas desse tipo são denominados problemas de *modelos separados* ou de *hipóteses separadas*. O conjunto de distribuições de probabilidade pertencente a famílias separadas inclui as distribuições lognormal, gama e Weibull (PEREIRA, 1978, 1979).

Testes para discriminar modelos separados foram inicialmente considerados por COX (1961, 1962), que desenvolveu resultados gerais e alguns exemplos. Trata-se de testes assintóticos baseados na modificação da razão de máxima verossimilhança de Neyman-Pearson. Nesses trabalhos fundamentais de Cox, foram sugeridas abordagens alternativas, que serviram como base para o desenvolvimento de diversos procedimentos de testes para modelos separados. Para revisão geral e referências, vide ARAUJO *et al.* (2005), ARAÚJO (1998), ROJAS (2001), ARAUJO e PEREIRA (2007), PEREIRA (2010) e PEREIRA e PEREIRA (2017).

O problema abordado neste trabalho é descrito como se segue. Considere uma amostra aleatória de  $n$  observações independentes,  $y = (y_1, \dots, y_n)$ , e  $m$  distribuições de probabilidades. Sem perda de generalidade, suponha que a hipótese  $H_f$  especifica que  $y$  provém de uma distribuição  $F$ , que pertence a uma família de distribuições de probabilidade  $\mathcal{F} = \{f(y|\alpha) : \alpha \in \Omega_\alpha\}$ . Suponha ainda que a hipótese  $H_g$  especifica que  $y$  provém da distribuição  $G \in \mathcal{G} = \{g(y|\beta) : \beta \in \Omega_\beta\}$ . Aqui,  $\alpha$  e  $\beta$  são vetores de parâmetros desconhecidos e supõe-se que as famílias das distribuições são separadas, i.e, os espaços de parâmetros  $\Omega_\alpha$  e  $\Omega_\beta$  são tais que  $\Omega_\alpha \cap \Omega_\beta \neq \Omega_\alpha$  e  $\Omega_\beta \cap \Omega_\alpha \neq \Omega_\beta$

(DASTOOR, 1983). O problema consiste em construir um procedimento de seleção para decidir qual dentre as distribuições  $F$  e  $G$  se ajusta adequadamente aos dados disponíveis. Teste de significância é considerado como procedimento estatístico para medir a consistência dos dados com a hipótese sendo testada, por vezes chamada de hipótese nula (COX, 1977; KEMPTHORNE, 1976).

Na estatística frequentista, o p-valor é usualmente considerado como uma medida de evidência a favor da hipótese nula (BERGER e DELAMPADY, 1987). A idéia geral desse procedimento é computar a probabilidade de, ao repetir um experimento aleatório sob um dado modelo estatístico e uma dada hipótese nula, obter resultado de um teste estatístico igual ou mais extremo (menos provável) do que aquele que foi efetivamente observado (STERN e PEREIRA, 2014). Portanto, trata-se de um conceito estabelecido sobre o espaço amostral (os dados) e não sobre o espaço paramétrico onde a hipótese é formulada.

No contexto Bayesiano, BERGER e DELAMPADY (1987) apresentam o fator de Bayes e a probabilidade a posteriori da hipótese nula como medidas de evidência alternativas do p-valor clássico. Estas medidas, no entanto, apresentam diversas limitações quando são utilizadas para resolver o problema de seleção de modelos (KAMARY *et al.*, 2014). Particularmente, sabe-se que o fator de Bayes, quando utilizado para testar hipóteses precisas ou para discriminar modelos separados, é muito sensível a suposições a priori. Por exemplo, quando a informação a priori sobre os parâmetros dos modelos candidatos é fraca e prioris impróprias são usadas, o fator de Bayes torna-se indefinido (AITKIN, 1991; O'HAGAN, 1995). Para contornar as dificuldades com relação ao uso de prioris impróprias, diversas modificações do fator de Bayes têm sido propostas na literatura (ARAUJO e PEREIRA, 2007; ARAUJO *et al.*, 2005; GHOSH *et al.*, 2002, 2006).

O Teste de Significância Totalmente Bayesiano (FBST, do inglês *Fully Bayesian Significance Test*) foi apresentado por PEREIRA e STERN (1999) como alternativa ao fator de Bayes, à probabilidade a posteriori da hipótese nula e aos p-valores clássicos, para calcular a medida de evidência a favor da hipótese nula precisa. O teste é considerado totalmente Bayesiano por fazer cálculos probabilísticos no espaço paramétrico original, em sua plena dimensão. O e-valor, a medida de evidência utilizada para definir o FBST, representa a probabilidade a posteriori relacionada aos pontos menos prováveis do espaço paramétrico. Essa medida favorece a hipótese nula quando seu valor é elevado. Uma vantagem do FBST é a de não necessitar de premissas adicionais, como probabilidade positiva para hipóteses precisas.

A medida de evidência do FBST alcança a invariância, uma propriedade fundamental de medidas de significância bem definidas (STERN e PEREIRA, 2014). A definição original do teste já é invariante com respeito a parametrizações da hipótese nula. MADRUGA *et al.* (2003) apresentaram uma formulação do FBST

explicitamente invariante sob transformações do sistema de coordenadas escolhido para o espaço paramétrico. Adicionalmente, o procedimento do FBST é compatível com a teoria Bayesiana da decisão (MADRUGA *et al.*, 2001), ou seja, é definido através da minimização da esperança a posteriori de uma função de perda. Para uma revisão geral e referências a respeito do FBST, vide PEREIRA *et al.* (2008) e STERN e PEREIRA (2014). Algumas aplicações ilustrando o uso do FBST em problemas relevantes de inferência estatística podem ser encontradas em PEREIRA e STERN (1999), DINIZ *et al.* (2012), LAURETTO *et al.* (2003) e LAURETTO *et al.* (2009).

Nesta tese, considerou-se o FBST para a escolha entre as distribuições lognormal, gama e Weibull cujas famílias são separadas. Estas distribuições estão entre as mais utilizadas em análise de sobrevivência para a modelagem do tempo até a ocorrência de uma falha (KLEIN e MOESCHBERGER, 2003; LAWLESS, 2002). Conforme sugerido por COX (1961), aqui, o problema é abordado num contexto de mistura linear dos modelos candidatos, ou seja, os modelos a serem comparados são considerados como componentes de um modelo de mistura finita. O procedimento do FBST é utilizado para testar as hipóteses sobre os pesos da mistura, com o objetivo de calcular as evidências a favor de cada modelo.

A novidade nesta abordagem é que as funções densidades dos componentes da mistura são reparametrizadas em termos da média,  $\mu$ , e da variância,  $\sigma^2$ , da população, para que todos os modelos em comparação compartilhem dos mesmos parâmetros. O mais comum na análise Bayesiana de um modelo de mistura finita é considerar pares de parâmetros diferentes para cada um desses modelos e assumir prioris independentes para cada par de parâmetros e priori Dirichlet para os pesos da mistura (LAURETTO *et al.*, 2007). No entanto, uma vez que a comparação entre os modelos é baseada no mesmo conjunto de observações, isto é, na mesma população, entende-se que seria inconveniente considerar, nesses modelos, médias e variâncias diferentes e, ainda por cima, com prioris diferentes. Do ponto de vista de inferência, esta reparametrização é vantajosa, na medida em que reduz a quantidade de parâmetros a serem estimados.

Uma vantagem de uso do procedimento do FBST para discriminar entre modelos separados através da mistura é a de permitir o uso de distribuições a priori impróprias (PEREIRA e PEREIRA, 2017; PEREIRA *et al.*, 2008). Outras vantagens de abordar o problema de seleção de modelos, de um modo geral, utilizando procedimentos de testes para os pesos de um modelo de mistura finita, são discutidas em KAMARY *et al.* (2014).

## 1.1 Objetivos

Este trabalho tem os seguintes objetivos principais:

- Formular o teste FBST para discriminar entre as distribuições lognormal, gama e Weibull, num contexto de mistura linear dos modelos candidatos. A idéia geral é explorar a relação entre os parâmetros desses modelos através da reparametrização das respectivas funções densidades em termos da média,  $\mu$ , e variância,  $\sigma^2$ , da população. Dessa forma, obtém-se um modelo de mistura com um número reduzido de parâmetros a serem estimados.
- Comparar, através de um estudo de simulação, os resultados empíricos do FBST com os do teste de Cox na discriminação entre as distribuições citadas, tomando como medidas de desempenho as taxas de decisões corretas (aceitação do modelo verdadeiro e rejeição do modelo falso). Ainda no âmbito desta comparação, três conjuntos de dados não censurados são usados para ilustrar os dois procedimentos. A escolha do teste de Cox como *benchmark* deveu-se, principalmente, ao fato de ser um procedimento consistente, assintoticamente normal e, além disso, comparado com outros testes desenvolvidos na econometria, como por exemplo, testes J e JA, ele é o único que pode ser generalizado para o caso multivariado (PESARAN, 1982).
- Aplicar o modelo de mistura lognormal-gama-Weibull à base de dados de pacientes com insuficiência renal crônica terminal submetidos à hemodiálise no estado do Rio de Janeiro. E, através do procedimento do FBST, examinar os pesos da mistura para escolher o modelo com melhor ajuste dos dados. O modelo escolhido pode, eventualmente, servir como a distribuição da variável resposta (no caso, o tempo de sobrevivência em hemodiálise) num modelo de regressão, para avaliar os fatores prognósticos individuais.

## 1.2 Organização da tese

Este trabalho está organizado como se segue: O capítulo 2 revê conceitos básicos de hipóteses separadas; sintetiza os resultados do teste de Cox para comparação entre os pares de modelos de famílias separadas; e apresenta uma breve revisão sobre soluções Bayesianas alternativas do fator de Bayes usual, que permitem o uso de distribuições a priori impróprias para os parâmetros dos modelos. O Capítulo 3 apresenta a revisão de conceitos básicos do FBST. O Capítulo 4 desenvolve a metodologia referente à formulação do FBST para hipóteses separadas, considerando os modelos sob hipóteses como componentes de um modelo de mistura finita. O Capítulo 5 apresenta alguns resultados numéricos baseados em simulações de pontos

amostrais, com o objetivo de avaliar o desempenho do FBST comparado com o do teste de Cox. Três exemplos de ilustração também são discutidos neste capítulo. O capítulo 6 analisa o modelo de mistura com dados censurados e mostra sua aplicação para um estudo de caso sobre a escolha de um modelo de sobrevivência para pacientes com insuficiência renal crônica terminal. Finalmente, o Capítulo 7 apresenta as considerações finais e propostas para estudos futuros.

# Capítulo 2

## Testes para famílias separadas de hipóteses

### 2.1 Introdução

Em qualquer área do conhecimento científico, pesquisadores lidam freqüentemente com o problema de escolha entre os modelos estatísticos alternativos. Nesse sentido, surgem várias questões relevantes (ATKINSON, 1970; LAURETTO, 2007):

- (i) Existe alguma evidência de que os ajustes fornecidos pelos modelos candidatos sejam significativamente diferentes para os dados disponíveis?
- (ii) Assumindo que um dos modelos seja o verdadeiro, qual a evidência fornecida pelos dados a favor desse modelo?
- (iii) Se um modelo representa a hipótese atualmente mantida, existe alguma evidência de não adequação dos dados àquele modelo, mas sim a um outro?
- (iv) Se nenhum dos modelos é verdadeiro, qual é a evidência fornecida pelos dados?

Algumas destas questões podem ser respondidas aplicando a teoria de teste de hipótese de Neyman-Pearson, caso os modelos candidatos pertençam a mesma família de distribuições. Geralmente, os procedimentos são baseados nos testes da razão de verossimilhança, de Wald e de multiplicador de Lagrange, vide BUSE (1982) para revisão geral destes testes. Alternativamente, procedimentos especiais são necessários quando o problema envolve modelos que pertencem a famílias separadas de hipóteses, no sentido de que um membro arbitrário de uma família não pode ser obtido como limite dos membros da outra.

Uma grande variedade de artigos discutindo os testes para famílias separadas de hipóteses surgiu a partir do trabalho fundamental de COX (1961, 1962), que desenvolveu um procedimento geral conhecido como teste de Cox, o qual generaliza

a razão de verossimilhança de Neyman-Pearson usada para hipóteses encaixadas. Para revisões gerais e referências, vide ARAUJO e PEREIRA (2007,b), PEREIRA (2005, 1981) e PEREIRA e PEREIRA (2017).

O problema de testar hipóteses separadas pode ser visto como uma avaliação da função densidade de probabilidade (fdp) de um vetor aleatório,  $y = (y_1, \dots, y_n)$ , de observações condicionalmente (aos parâmetros) independentes e identicamente distribuídas (c.i.i.d), com respeito a duas hipóteses:

$$\begin{aligned} H_f : y \text{ tem fdp } f(y|\alpha), \quad \alpha \in \Omega_\alpha, \text{ e} \\ H_g : y \text{ tem fdp } g(y|\beta), \quad \beta \in \Omega_\beta, \end{aligned}$$

onde os espaços de parâmetros  $\Omega_\alpha$  e  $\Omega_\beta$  são tais que  $\Omega_\alpha \cap \Omega_\beta \neq \Omega_\alpha$  e  $\Omega_\beta \cap \Omega_\alpha \neq \Omega_\beta$ ;  $\alpha$  e  $\beta$  são vetores de parâmetros desconhecidos das fdps  $f$  e  $g$ , respectivamente. Se  $\Omega_\alpha \cap \Omega_\beta = \Omega_\alpha$  (ou  $\Omega_\beta \cap \Omega_\alpha = \Omega_\beta$ ), então as hipóteses são consideradas encaixadas (DASTOOR, 1983).

A definição formal de hipóteses separadas usa o conceito das discrepâncias como em GHOSH e SUBRAMANYAM (1975):

$$\begin{aligned} d(f, g) &= E_\alpha\{|f - g|\} \\ &= \int |f(y|\alpha) - g(y|\beta)|f(y|\alpha)dy, \end{aligned} \tag{2.1}$$

ou da medida de proximidade entre as hipóteses  $H_f$  e  $H_g$ , o critério de informação de Kullback-Leibler, como em PESARAN (1994) e GOURIEROUX e MONFORT (1994):

$$\begin{aligned} I(f, g) &= E_\alpha\{\log f - \log g\} = E_\alpha\{\ell_{fg}\} \\ &= \int_{\Omega_f} \log \frac{f(y|\alpha)}{g(y|\beta)} f(y|\alpha)dy, \end{aligned} \tag{2.2}$$

Portanto,  $H_f$  and  $H_g$  são separadas se

$$\begin{aligned} \inf_{\Omega_f, \Omega_g} d(f, g) &> 0 \text{ ou} \\ \inf_{\Omega_f, \Omega_g} I(f, g) &> 0. \end{aligned}$$

GOURIEROUX e MONFORT (1994) também definem hipóteses parcialmente separadas, quando  $\inf_{\Omega_f, \Omega_g} I(f, g) = 0$ , para alguns (não para todos) parâmetros. Expressões análogas a (2.1) e (2.2) são definidas quando  $H_f$  e  $H_g$  são invertidas.

Neste capítulo são apresentados os resultados do teste de Cox para comparação entre os pares de modelos de famílias separadas e mencionados alguns enfoques

alternativos para o problema. Em seguida, são introduzidas algumas soluções Bayesianas alternativas ao fator de Bayes usual, que permitem o uso de distribuições a priori impróprias para os parâmetros dos modelos. Finalmente, são apresentadas algumas notas bibliográficas sobre o desenvolvimento e aplicações recentes desses procedimentos.

## 2.2 Testes de Cox

Os testes assintóticos (LEHMANN, 1999) desenvolvidos por COX (1961, 1962) são baseados em uma modificação da razão de máxima verossimilhança de Neyman-Pearson. Suponhamos que  $H_f$  e  $H_g$ , definidas na seção anterior, sejam, respectivamente, as hipóteses nula e alternativa. A estatística considerada por Cox para testar  $H_f$  contra  $H_g$  é

$$T_{fg} = \ell_f(\hat{\alpha}) - \ell_g(\hat{\beta}) - n \left[ \text{plim}_{n \rightarrow \infty} \frac{\ell_f(\hat{\alpha}) - \ell_g(\hat{\beta})}{n} \right]_{\alpha=\hat{\alpha}} \quad (2.3)$$

onde, para uma amostra aleatória de tamanho  $n$ ,  $\hat{\alpha}$  e  $\hat{\beta}$  são, respectivamente, os estimadores de máxima verossimilhança de  $\alpha$  e  $\beta$ ,  $\ell_f(\hat{\alpha}) = \log f(y|\hat{\alpha})$ ,  $\ell_g(\hat{\beta}) = \log g(y|\hat{\beta})$ . O operador “plim” simboliza convergência em probabilidade e o subscrito  $\alpha$  indica que o limite em probabilidade é calculado sob  $H_f$ .

Cox mostrou que, assintoticamente, sob a hipótese alternativa,  $T_{fg}$  tem média negativa e que, sob a hipótese nula,  $T_{fg}$  tem distribuição normal com média zero e variância

$$V_\alpha(T_{fg}) = V_\alpha \{ \ell_f(\alpha) - \ell_g(\beta_\alpha) \} - C'_\alpha I_\alpha^{-1} C_\alpha \quad (2.4)$$

onde  $\beta_\alpha$  é o limite em probabilidade de  $\hat{\beta}$ , quando  $n \rightarrow \infty$ , supondo  $H_f$  verdadeira,  $C_\alpha \equiv n \frac{\partial}{\partial \alpha} \left[ \text{plim}_{n \rightarrow \infty} \frac{\ell_f(\hat{\alpha}) - \ell_g(\hat{\beta})}{n} \right]$  e  $I_\alpha$  é a matriz de informação de  $\alpha$ .

Quando  $H_g$  é a hipótese nula e  $H_f$  é a hipótese alternativa, resultados análogos são obtidos para uma estatística  $T_{gf}$ . Portanto,  $T_{fg}^* = T_{fg} \{V_\alpha(T_{fg})\}^{-1/2}$  e  $T_{gf}^* = T_{gf} \{V_\beta(T_{gf})\}^{-1/2}$ , sob  $H_f$  e  $H_g$ , respectivamente, podem ser consideradas, aproximadamente, como variáveis aleatórias com distribuição normal padrão, e testes bilaterais podem ser efetuados, de acordo com as recomendações de COX (1962).

Assim, fixado um nível de significância (assintótico)  $\gamma$ , a região de rejeição do teste de Cox será definida como  $R = \{Y : |T_{(\bullet)}^*| > z_{\gamma/2}\}$ , sendo  $T_{(\bullet)}^* \sim N(0, 1)$ .

A aplicação dos testes  $T_{fg}^*$  e  $T_{gf}^*$  inclui a possibilidade de rejeitar ambas as hipóteses (atendendo a questão (iv)), sugerindo a necessidade de formulação de uma nova hipótese não considerada previamente, ou não rejeitar nenhuma das hipóteses,

o que implica a falta de evidência sobre qual dos modelos deve ser escolhido. Se  $T_{fg}^*$  é significativamente negativo, implica a rejeição de  $H_f$  na direção (em favor) da hipótese nula  $H_g$ . Se  $T_{fg}^*$  é significativamente positivo, implica rejeitar  $H_f$  na direção oposta de  $H_g$ . Interpretações sobre outros possíveis resultados, quando ambos os testes são aplicados, podem ser encontradas em PEREIRA (2005).

As estatísticas do teste de Cox para comparações entre pares de distribuições de famílias separadas são apresentados no capítulo 5. Demonstrações completas encontram-se em PEREIRA (1978, 1979) e PEREIRA e PEREIRA (2017).

Uma abordagem alternativa introduzida por COX (1961) considera a formulação de um modelo geral que inclui os modelos em  $H_f$  e  $H_g$  como casos particulares. A função de densidade de cada observação poderia ser proporcional a uma combinação exponencial dos modelos candidatos, isto é,

$$h_c(y|\alpha, \beta, \lambda) \propto [f(y|\alpha)]^\lambda [g(y|\beta)]^{1-\lambda}, \quad \lambda \in [0, 1]. \quad (2.5)$$

Esta idéia foi desenvolvida por ATKINSON (1970) para algumas distribuições de família exponencial. Outro modelo composto mencionado por ATKINSON (1970) e aprofundado por QUANDT (1974) considera uma combinação convexa dos modelos em  $H_f$  e  $H_g$ , isto é,

$$h_c(y|\alpha, \beta, \lambda) \propto \lambda f(y|\alpha) + (1 - \lambda)g(y|\beta), \quad \lambda \in [0, 1]. \quad (2.6)$$

Em todas essas abordagens, a escolha entre os modelos é feita com base nos testes de hipóteses sobre o parâmetro da mistura  $\lambda$ . A hipótese  $H_f$  é obtida impondo  $\lambda = 1$ , enquanto que a  $H_g$  obtém-se impondo  $\lambda = 0$ . Assim, em princípio, testar  $\lambda = 1$  ou  $\lambda = 0$  implica testar a significância de  $H_f$  ou  $H_g$ , respectivamente.

Nesta tese, o procedimento do FBST, proposto para discriminar entre modelos de famílias separadas, é baseado na mistura linear dos modelos candidatos. Detalhes sobre este tópico são discutidos no capítulo 4.

## 2.3 Fator de Bayes

COX (1961) também sugeriu uma formulação geral para o problema de famílias separadas de hipóteses sob perspectiva de inferência Bayesiana. A razão a posteriori para  $H_f$  versus  $H_g$  é

$$\frac{\pi_f \int_{\Omega_\alpha} f(\mathbf{y}|\boldsymbol{\alpha})\pi_f(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\pi_g \int_{\Omega_\beta} g(\mathbf{y}|\boldsymbol{\beta})\pi_g(\boldsymbol{\beta})d\boldsymbol{\beta}} = \frac{\pi_f q_f(\mathbf{y})}{\pi_g q_g(\mathbf{y})} = \frac{\pi_f}{\pi_g} B_{fg}(\mathbf{y}), \quad (2.7)$$

onde  $\pi_f$  e  $\pi_g$  são, respectivamente, as probabilidades a priori de  $H_f$  e  $H_g$ ,  $q_f(\mathbf{y})$  e

$q_g(\mathbf{y})$  denotam as distribuições preditivas com probabilidades a priori  $\pi_f(\boldsymbol{\alpha})$  e  $\pi_g(\boldsymbol{\beta})$  para os parâmetros sob  $H_f$  e  $H_g$ .  $B_{fg}(\mathbf{y})$  é o fator de Bayes e representa o peso de evidência nos dados em favor a  $H_f$  contra  $H_g$ .

É bem sabido que o fator de Bayes em (2.7) possui, no entanto, muitas limitações, como conseqüências de sua sensibilidade a suposições a priori (BERGER e PERICCHI, 1996; GHOSH *et al.*, 2002; KAMARY *et al.*, 2014; O'HAGAN, 1995). Esta abordagem exige que a especificação das prioris  $\pi_f$  e  $\pi_f(\boldsymbol{\alpha})$  seja coerente com as de  $\pi_g$  e  $\pi_g(\boldsymbol{\beta})$ . Se os espaços de parâmetros possuem dimensões diferentes, por exemplo, não existe relações simples entre os parâmetros. Quando a informação a priori é fraca e prioris impróprias são usadas, o fator de Bayes torna-se indefinido. Por exemplo, caso se use  $\pi_f(\boldsymbol{\alpha})$  e  $\pi_g(\boldsymbol{\alpha})$  impróprias proporcionais a constante  $c_f$  e  $c_g$ , respectivamente, e  $\pi_f = \pi_g$ , o fator de Bayes em (2.7) dependerá da razão  $c_f/c_g$  e, portanto, não está determinado. Assim, mesmo que a posteriori seja própria, a distribuição preditiva a priori (distribuição marginal dos dados), usada no cálculo do fator de Bayes, é imprópria (PAULINO *et al.*, 2003). Para solucionar o problema com o uso de prioris impróprias, fatores de Bayes alternativos foram propostos (ARAUJO e PEREIRA, 2007; ARAUJO *et al.*, 2005; GHOSH *et al.*, 2002).

A seguir, são descritos os principais fatores de Bayes modificados que permitem o uso de prioris impróprias.

### 2.3.1 Fator de Bayes a posteriori

AITKIN (1991) propôs o fator de Bayes a posteriori que compara as médias a posteriori das funções de verossimilhança sob  $H_f$  e  $H_g$ . Mais formalmente, a densidade a posteriori sob  $H_f$  e  $H_g$  são, respectivamente,

$$\pi_f(\boldsymbol{\alpha}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\alpha})\pi_f(\boldsymbol{\alpha})}{\int_{\Omega_\alpha} f(\mathbf{y}|\boldsymbol{\alpha})\pi_f(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \quad \text{e} \quad \pi_g(\boldsymbol{\beta}|\mathbf{y}) = \frac{g(\mathbf{y}|\boldsymbol{\beta})\pi_g(\boldsymbol{\beta})}{\int_{\Omega_\beta} g(\mathbf{y}|\boldsymbol{\beta})\pi_g(\boldsymbol{\beta})d\boldsymbol{\beta}}. \quad (2.8)$$

As médias a posteriori das funções de verossimilhança são, respectivamente,

$$q_f^{PO}(\mathbf{y}) = \int_{\Omega_\alpha} f(\mathbf{y}|\boldsymbol{\alpha})\pi_f(\boldsymbol{\alpha}|\mathbf{y})d\boldsymbol{\alpha} = \frac{\int_{\Omega_\alpha} \{f(\mathbf{y}|\boldsymbol{\alpha})\}^2 \pi_f(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int_{\Omega_\alpha} f(\mathbf{y}|\boldsymbol{\alpha})\pi_f(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \quad (2.9)$$

e

$$q_g^{PO}(\mathbf{y}) = \int_{\Omega_\beta} g(\mathbf{y}|\boldsymbol{\beta})\pi_g(\boldsymbol{\beta}|\mathbf{y})d\boldsymbol{\beta} = \frac{\int_{\Omega_\beta} \{g(\mathbf{y}|\boldsymbol{\beta})\}^2 \pi_g(\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\Omega_\beta} g(\mathbf{y}|\boldsymbol{\beta})\pi_g(\boldsymbol{\beta})d\boldsymbol{\beta}}. \quad (2.10)$$

O fator de Bayes a posteriori é então definido como a razão das médias a posteriori, ou seja,

$$B_{fg}^{PO}(\mathbf{y}) = \frac{q_f^{PO}(\mathbf{y})}{q_g^{PO}(\mathbf{y})}. \quad (2.11)$$

### 2.3.2 Fator de Bayes parcial

O cálculo formal do fator de Bayes parcial foi apresentado por O'HAGAN (1995). A ideia é utilizar parte dos dados como uma amostra de treinamento para atualizar as distribuições a priori, e o remanescente para comparar os modelos. Mais formalmente, considere a divisão da amostra em duas partes,  $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ , de tamanho  $m$  e  $n - m$ , respectivamente, com  $0 < m < n$ . Primeiro, a subamostra  $\mathbf{x}$  é usada para obter as distribuições a posteriori próprias,  $\pi_f(\boldsymbol{\alpha}|\mathbf{x})$  e  $\pi_g(\boldsymbol{\beta}|\mathbf{x})$ ; na segunda etapa, tomando essas distribuições como prioris, os dados remanescentes,  $\mathbf{z}$ , são usados para calcular o fator de Bayes, conhecido como fator de Bayes parcial,

$$B_{fg}^p(\mathbf{z}|\mathbf{x}) = \frac{q_f(\mathbf{z}|\mathbf{x})}{q_g(\mathbf{z}|\mathbf{x})} = \frac{\int \pi_f(\boldsymbol{\alpha}|\mathbf{x}) f(\mathbf{z}|\boldsymbol{\alpha}, \mathbf{x}) d\boldsymbol{\alpha}}{\int \pi_g(\boldsymbol{\beta}|\mathbf{x}) g(\mathbf{z}|\boldsymbol{\beta}, \mathbf{x}) d\boldsymbol{\beta}}. \quad (2.12)$$

Note que

$$q_j(\mathbf{z}|\mathbf{x}) = \frac{q_j(\mathbf{z}, \mathbf{x})}{q_j(\mathbf{x})} = \frac{q_j(\mathbf{y})}{q_j(\mathbf{x})}, \quad j = f, g, \quad (2.13)$$

então

$$B_{fg}^p(\mathbf{z}|\mathbf{x}) = \frac{q_f(\mathbf{y})/q_g(\mathbf{y})}{q_f(\mathbf{x})/q_g(\mathbf{x})} = \frac{B_{fg}(\mathbf{y})}{B_{fg}(\mathbf{x})}, \quad (2.14)$$

ou seja,

$$B_{fg}(\mathbf{y}) = B_{fg}^p(\mathbf{z}|\mathbf{x}) B_{fg}(\mathbf{x}). \quad (2.15)$$

Da equação (2.14), observa-se que as prioris impróprias afetam  $B_{fg}(\mathbf{y})$  e  $B_{fg}(\mathbf{x})$  da mesma maneira e, portanto, seus efeitos se cancelam.

### 2.3.3 Fator de Bayes fracionário

A grande dificuldade em relação ao fator de Bayes parcial é a seleção da amostra de treinamento,  $\mathbf{x}$ , a partir dos dados. Para evitar a arbitrariedade da escolha de uma amostra de treinamento ou ter que considerar todas subamostras possíveis de um dado tamanho, O'HAGAN (1995) desenvolveu uma forma simplificada do fator de Bayes parcial. Formalmente, seja  $b = m/n$ ; se  $n$  e  $m$  são suficientemente grandes, as verossimilhanças  $f(\mathbf{x}, \boldsymbol{\alpha})$  e  $g(\mathbf{x}, \boldsymbol{\beta})$ , baseadas apenas na amostra de treinamento  $\mathbf{x}$ , são aproximadas a verossimilhanças para a amostra completa,  $f(\mathbf{y}, \boldsymbol{\alpha})$  e  $g(\mathbf{y}, \boldsymbol{\beta})$ , elevadas a  $b$ , respectivamente.

Por analogia às equações (2.14) e (2.15), o fator de Bayes fracionário é definido como

$$B_{fg}^b(\mathbf{y}) = \frac{q_f(b, \mathbf{y})}{q_g(b, \mathbf{y})}, \quad (2.16)$$

onde

$$q_f(b, \mathbf{y}) = \frac{\int_{\Omega_\alpha} \pi_f(\boldsymbol{\alpha}) f(\mathbf{y}|\boldsymbol{\alpha}) d\boldsymbol{\alpha}}{\int_{\Omega_\alpha} \pi_f(\boldsymbol{\alpha}) \{f(\mathbf{y}|\boldsymbol{\alpha})\}^b d\boldsymbol{\alpha}} \quad \text{e} \quad q_g(b, \mathbf{y}) = \frac{\int_{\Omega_\beta} \pi_g(\boldsymbol{\beta}) f(\mathbf{y}|\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int_{\Omega_\beta} \pi_g(\boldsymbol{\beta}) \{f(\mathbf{y}|\boldsymbol{\beta})\}^b d\boldsymbol{\beta}}. \quad (2.17)$$

Se  $\pi_f(\boldsymbol{\alpha}) = c_f h_f(\boldsymbol{\alpha})$  e  $\pi_g(\boldsymbol{\beta}) = c_g h_g(\boldsymbol{\beta})$ , as constantes indeterminadas,  $c_f$  e  $c_g$ , cancelam-se. O'HAGAN (1995) mostrou que o fator de Bayes fracionário é consistente, dado que  $b$  tende a zero quando  $n$  cresce.

### 2.3.4 Fator de Bayes intrínseco

BERGER e PERICCHI (1996) definiram que a amostra de treinamento  $\mathbf{x}$  da partição  $\mathbf{y} = (\mathbf{x}, \mathbf{z})$  é mínima se as posteriores para  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$  são próprias e não existe subconjunto de  $\mathbf{x}$  que leve a uma posteriori própria. Existem usualmente muitas, digamos  $R$ , partições caracterizando uma amostra de treinamento minimal. Para evitar a arbitrariedade da escolha de uma amostra de treinamento, o fator de Bayes intrínseco de Berger e Pericchi toma a média, geométrica ou aritmética, ou a mediana dos fatores de Bayes parciais  $\{B_{fg}^p(\mathbf{z}_r|\mathbf{x}_r); r = 1, \dots, R\}$  obtidos de todas as  $R$  amostras de treinamento minimais.

O fator de Bayes intrínseco para a média geométrica é

$$B_{fg}^{IG}(\mathbf{y}) = \left\{ \prod_{r=1}^R B_{fg}^p(\mathbf{z}_r|\mathbf{x}_r) \right\}^{1/R}, \quad (2.18)$$

$$B_{fg}^{IA}(\mathbf{y}) = \frac{1}{R} \sum_{r=1}^R B_{fg}^p(\mathbf{z}_r|\mathbf{x}_r) \quad (2.19)$$

e

$$B_{fg}^{IM}(\mathbf{y}) = \text{med}\{B_{fg}^p(\mathbf{z}_r|\mathbf{x}_r); r = 1, \dots, R\}. \quad (2.20)$$

Uma vez que todas essas versões são baseadas no fator de Bayes parcial, a indeterminação devido à constante multiplicativa  $c_f/c_g$  desaparece.

## 2.4 Notas Bibliográficas

Uma grande variedade de pesquisas a respeito de testes para famílias separadas de hipóteses tem sido desenvolvida desde os trabalhos fundamentais de COX (1961, 1962). A relevância do tema e sua influência sobre os últimos procedimentos de testes desenvolvidos na literatura estatística foram novamente discutidas em COX (2013). Outras contribuições de Cox sugerindo enfoques alternativos são apresentadas em ARAUJO *et al.* (2005). Revisões gerais e referências bibliográficas a respeito do assunto podem ser encontradas em GOURIEROUX e MONFORT

(1994), MCALEER (1995), PEREIRA (2005, 2010, 1977a, 1981), PEREIRA e PEREIRA (2017) e PESARAN (1982).

ROJAS (2001) desenvolveu outros procedimentos de testes para hipóteses separadas, a saber: Critério de Verossimilhança Relativa (procedimento de Lindsey), Teste de Divergência Direcional, Teste da Função Geradora de Momentos e Critério de Informação de Kullback-Leibler. Através de estudos de simulação, o autor avaliou a eficiência do Critério de verossimilhança relativa estimando a probabilidade da correta identificação da distribuição da qual são geradas as amostras. Os resultados indicam alta eficiência deste critério na escolha da distribuição verdadeira quando esta é bastante simétrica e o tamanho da amostra é maior que 50. Para os demais procedimentos de seleção, os resultados da avaliação da distribuição da estatística de cada teste, sob a hipótese nula, e o poder são obtidos por simulações. Essas estatísticas são desenvolvidas para selecionar entre as famílias de distribuições de probabilidade exponencial, lognormal, gama e Weibull.

Revisões de literatura focando o caso particular de modelos de regressão aparecem em MCALEER e PESARAN (1987, 1986), MACKINNON (1983) e PESARAN (1974). Outras referências em séries temporais são BURKE *et al.* (1990), CHAMBERS (1993), GODFREY e TREMAYNE (1988) e SILVAPULLE e KING (1993). Abordagens Bayesianas foram desenvolvidas por ALLENBY (1993); DE JONG (1993); ROSSI (1985); RUST e SCHMITTLEIN (1985); WIGINTON (1993); ZELLNER (1984). Outras aplicações em econometria são apresentadas por AURIKKO (1985), HARVEY (1985) e PESARAN e PESARAN (1995).

Na década de 90, estatísticos Bayesianos desenvolveram métodos alternativos para contornar as dificuldade com o fator de Bayes usual. Vide ARAUJO e PEREIRA (2007) e PEREIRA e PEREIRA (2017) para revisão geral e referências. Resultados de simulação de alguns fatores de Bayes modificados foram apresentados na tese de ARAÚJO (1998). Para discussões sobre o fator de Bayes a posteriori de AITKIN (1991), vide AITKIN (1992), AITKIN *et al.* (2005) e LINDLEY (1993). Os fatores de Bayes fracionário e intrínseco são amplamente discutidos por O'HAGAN (1997) e em vários trabalhos citados nos artigos *reviews* de BERGER e PERICCHI (2001) e PERICCHI (2005). Uma expressão geral a partir da qual são derivados os fatores de Bayes pode ser vista em GELFAND e DEY (1994). Outros resultados de simulação sobre os fatores de Bayes fracionário, intrínseco e a posteriori são apresentados em ARAÚJO (1998). PEREIRA e WECHSLER (1993) usaram o fator de Bayes na ordenação do espaço amostral em qualquer dimensão com objetivo de definir uma nova medida de “P-valor”.

# Capítulo 3

## O FBST–Teste de Significância Totalmente Bayesiano

### 3.1 Introdução

O FBST, *Fully Bayesian Significance Teste*, ou Teste de Significância Totalmente Bayesiano, foi apresentado por PEREIRA e STERN (1999) como um teste intuitivo e teoricamente coerente, com caracterização geométrica, e que pode ser implementado usando técnicas modernas de otimização e integração numérica. O método é conceitualmente simples e objetivo, sendo exclusivamente baseado em avaliações da probabilidade a posteriori de conjuntos do espaço paramétrico. Trata-se de um teste de significância Bayesiano desenhado particularmente para computar um valor de evidência a favor de uma hipótese precisa.

A medida de evidência do FBST, também tratado como e-valor ou *valor epistêmico*, apresenta diversas propriedades desejáveis para uma função de suporte estatístico a hipóteses precisas. Por exemplo, PEREIRA *et al.* (2008), STERN e PEREIRA (2014) e BORGES e STERN (2007) destacaram as seguintes propriedades:

- (i) Fornece uma medida intuitiva de significância estatística da hipótese sendo testada como uma probabilidade definida diretamente no espaço paramétrico original, em sua plena dimensão;
- (ii) Tem uma definição geométrica explicitamente invariante, independente da particular parametrização da hipótese sendo testada ou particular sistema de coordenadas escolhido para o espaço paramétrico;
- (iii) Fornece uma medida de significância suave, i.e., contínua e diferenciável, nos parâmetros da hipótese e nas estatísticas da amostra, sob condições apropriadas de regularidade topológica e analítica do modelo;

- (iv) Obedece ao princípio de verossimilhança, i.e, a informação obtida a partir das observações deve ser representada pela, e somente pela, função de verossimilhança;
- (v) Não requer qualquer artifício *ad hoc* como atribuir probabilidades positivas a conjuntos de medida nula, ou estabelecer razões de crença iniciais arbitrárias entre as hipóteses;
- (vi) É uma função de suporte *possibilística*, onde o suporte a uma disjunção lógica é o máximo entre o suporte dos disjuntos;
- (vii) Fornece uma medida consistente com a medida subjacente de probabilidade e conforme (similaridade em forma) com a função surpresa subjacente;
- (viii) É um procedimento exato, i.e, não se utiliza no cálculo do e-valor qualquer aproximação assintótica;
- (ix) Fornece operações de Composicionalidade em modelos complexos;
- (x) Permite a incorporação de experiência prévia ou opiniões de especialistas via distribuições a priori.

A seguir serão apresentados os conceitos básicos do modelo estatístico Bayesiano sobre o qual é definida a medida de evidência do FBST.

## 3.2 Modelo Estatístico Bayesiano

O modelo padrão da estatística Bayesiana paramétrica concerne a um parâmetro (vetorial)  $\theta \in \Theta$ , e uma variável aleatória (vetorial),  $y \in \mathcal{Y}$ , que tem uma distribuição amostral com uma forma funcional especificada,  $p(y|\theta)$ , indexada por  $\theta$ . A mesma forma funcional, considerada como função da variável livre  $\theta$  com argumento fixo  $y$ , passa a ser denominada função de verossimilhança do modelo,  $l(\theta; y) = p(y|\theta)$ . O conjunto de todos os valores possíveis de  $\theta$ ,  $\Theta \subset \mathcal{R}^q$ ,  $q \geq 1$ , é denominado espaço paramétrico; e o conjunto de todos os valores possíveis de  $y$ ,  $\mathcal{Y} \subset \mathcal{R}^p$ ,  $p \geq 1$ , é denominado espaço amostral.

No contexto Bayesiano, o parâmetro  $\theta$  é considerado como uma variável aleatória latente (não observável) e, portanto, a teoria de probabilidade é utilizada em ambos os espaços, amostral e paramétrico. Conseqüentemente, a distribuição conjunta,  $p(\theta, y)$  deveria resumir toda a informação<sup>1</sup> disponível em um modelo estatístico, vide WECHSLER *et al.* (2008).

---

<sup>1</sup>Informação, neste contexto, é um conceito intuitivo. “Informação é o que ela provoca: a mudança do nível da incerteza sobre quantidades ou estado da natureza desconhecidos”. Uma discussão detalhada sobre o conceito de informação estatística sob perspectiva Bayesiana pode ser vista em STERN e PEREIRA (2012).

Seguindo as regras do cálculo de probabilidades (Regra de Bayes), a distribuição conjunta de  $y$  e  $\theta$  pode ser fatorizada, tanto como a função de verossimilhança do parâmetro dada a observação vezes a distribuição a priori de  $\theta$ , quanto como a distribuição a posteriori do parâmetro vezes a densidade marginal da observação,

$$p(\theta, y) = l(\theta; y)p(\theta) = p(\theta|y)p(y).$$

A distribuição de probabilidade a priori,  $p_0(\theta)$ , representa a informação inicial disponível sobre o parâmetro. Se, em seguida, for observado um único evento,  $y$ , a distribuição a posteriori para  $\theta$ , representando a informação disponível sobre o parâmetro após a observação, é dada por

$$p_1(\theta) \propto l(\theta; y)p_0(\theta).$$

Caso se queira substituir o símbolo de “proporcional a”,  $\propto$ , pelo de igualdade,  $=$ , basta dividir o lado direito pela constante de normalização,  $p_1(y) = \int_{\Theta} p(y|\theta)p_0(\theta)d\theta$ .

O aprendizado Bayesiano é um processo recursivo, onde a distribuição a posteriori após um passo no aprendizado torna-se a distribuição a priori para o passo seguinte. Assumindo que as observações são c.i.i.d., a distribuição a posteriori após  $n$  observações,  $\mathbf{y} = (y_1, \dots, y_n)$ , será,

$$p_n(\theta) \propto l(\theta; \mathbf{y}_n)p_{n-1}(\theta) \propto \left[ \prod_{i=1}^n l(\theta; y_i) \right] p_0(\theta). \quad (3.1)$$

### 3.3 Formulação original do FBST

Nesta seção é apresentada a versão original intuitiva do FBST, introduzida por PEREIRA e STERN (1999), com objetivo de fornecer uma alternativa Bayesiana aos testes de significância de hipóteses precisas. A medida de evidência do FBST, o e-valor, é definida em um modelo estatístico Bayesiano, apresentado na última seção, com densidade a posteriori,  $p_n(\theta)$ .

**Definição 3.1** (STERN e PEREIRA, 2014) Uma hipótese estatística,  $H$ , (por vezes chamada de hipótese nula) afirma que o parâmetro  $\theta$  do modelo estatístico encontra-se dentro do conjunto da hipótese,  $\Theta_H$ .

O conjunto da hipótese é geralmente definido por restrições de desigualdade e igualdade definidas por funções vetoriais,  $g = [g_1(\theta), g_2(\theta), \dots, g_l(\theta)]'$  e  $h = [h_1(\theta), h_2(\theta), \dots, h_k(\theta)]'$  no espaço paramétrico,

$$H : \theta \in \Theta_H, \quad \Theta_H = \{\theta \in \Theta | g(\theta) \leq 0 \wedge h(\theta) = 0\}. \quad (3.2)$$

Note que a dimensão da hipótese é especificada pelo número de restrições de igualdade. A seguinte notação será utilizada:  $t = \dim(\Theta)$  é a dimensão do espaço paramétrico,  $k$ , a codimensão de  $\Theta_H$ , conta o número de equações escalares restringindo  $H$ ;  $\dim(\Theta_H) = h = t - k$ , a dimensão de  $\Theta_H$ , conta o número de graus de liberdade para uma partícula movendo-se em  $H$ .

O domínio de interesse do FBST são as hipóteses *precisas*, i.e., aquelas nas quais  $\dim(\Theta_H) < \dim(\Theta)$ , ou seja,  $k > 0$ . Ao longo deste trabalho, serão assumidas, sempre que necessárias, condições apropriadas de regularidade topológica e analítica, tais como continuidade, diferenciabilidade, e a existência de argumentos maximais, restritos (ao conjunto  $\Theta_H$ ) e irrestritos (ao conjunto  $\Theta$ ). A bem da simplicidade, de agora em diante, será utilizada uma notação relaxada, escrevendo  $H$  no lugar de  $\Theta_H$ .

O cálculo da medida de evidência do FBST é realizado em duas etapas, nomeadamente a otimização e integração numéricas. O passo de otimização consiste em encontrar  $\theta^*$  como sendo o ponto de máxima densidade a posteriori sob a hipótese precisa  $H$ , e respectivo valor na imagem de  $p$ . O passo de integração consiste em avaliar a integral da densidade a posteriori no conjunto tangente,  $T(\mathbf{y})$ , um subconjunto do espaço paramétrico formado por todos os pontos que possuem valores da densidade a posteriori maiores que  $p_n(\theta^*)$ . Mais formalmente,

**Definição 3.2** Considere uma hipótese precisa  $H : \theta \in \Theta_H$  e seja

$$\theta^* \in \arg \max_H \{p_n(\theta)\} \quad \text{e} \quad p^* = \max_H \{p_n(\theta)\} = p_n(\theta^*).$$

A medida de evidência contra  $H$  é a probabilidade a posteriori do conjunto  $T(\mathbf{y})$ , i.e,

$$\bar{ev}(H) = \Pr(\theta \in T(\mathbf{y}) | \mathbf{y}) = \int_{T(\mathbf{y})} p_n(\theta) d\theta, \quad \text{onde} \quad (3.3)$$

$$T(\mathbf{y}) = \{\theta \in \Theta : p_n(\theta) \geq p^*\}.$$

Note que a medida de evidência a favor de  $H$ ,  $ev(H) = 1 - \bar{ev}(H)$ , não é uma evidência contra a hipótese alternativa  $\bar{H} = \Theta \setminus H$  (que não é precisa). Equivalentemente,  $\bar{ev}(H)$  não é evidência a favor de  $\bar{H}$  embora seja contra  $H$ . Observe também que, de acordo com a definição acima, o conjunto  $T(\mathbf{y})$  (tangencial a  $H$  no ponto  $\theta^*$ ) é uma região de máxima densidade de probabilidade a posteriori (HPD, do inglês *Highest posterior density*), cujo nível de credibilidade é  $\bar{ev}(H)$ . Intuitivamente, um grande volume de  $T(\mathbf{y})$  é indício de que a região sobre a hipótese (incluindo  $\theta^*$ ) possui baixa densidade de probabilidade, i.e,  $\theta$  está entre os pontos menos prováveis a posteriori, indicando “forte” evidência contra  $H$ .

**Definição 3.3** O FBST é um procedimento que rejeita a hipótese  $H$  quando  $ev(H)$  é pequeno.

Importa destacar que, no caso em que o subconjunto que define a hipótese tem a mesma dimensão do espaço paramétrico, a decisão sobre aceitação/rejeição da hipótese pode ser baseada tão somente na probabilidade a posteriori daquele conjunto.

### 3.4 Formulação invariante do FBST

Nesta seção é apresentada uma formulação explicitamente invariante da medida de evidência do FBST, sob transformações regulares do sistema de coordenadas, proposta por MADRUGA *et al.* (2003). Para isso, o e-valor de uma hipótese  $H \subseteq \Theta$  é definido em um modelo estatístico Bayesiano com densidade a posteriori,  $p_n(\theta)$ , e uma densidade de referência,  $r(\theta)$ , definida em  $\Theta$ .

**Definição 3.4** Considere uma hipótese precisa  $H$  e seja

$$\theta^* \in \arg \max_H \left\{ \frac{p_n(\theta)}{r(\theta)} \right\}, \quad \text{e} \quad s^* = \max_H \left\{ \frac{p_n(\theta)}{r(\theta)} \right\} = \left\{ \frac{p_n(\theta^*)}{r(\theta^*)} \right\}.$$

A função de surpresa,  $s(\theta) = p_n(\theta)/r(\theta)$ , indica a mudança na densidade a posteriori relativamente à densidade de referência, representando uma situação inicial de informação mínima (STERN e PEREIRA, 2014). Seu papel no FBST é tornar  $ev(H)$  explicitamente invariante sob transformações regulares do sistema de coordenadas do espaço paramétrico.

Agora, o conjunto  $T(\mathbf{y}) \subset \Theta$  é uma região de máxima surpresa relativa (HRSS, do inglês *Highest relative surprise set*), formado por todos os pontos que possuem valores da função surpresa maiores que  $s^*$ . Assim, a medida de evidência invariante contra  $H$  é definida como probabilidade a posteriori deste conjunto, i.e.,

$$\begin{aligned} \bar{ev}(H) &= \Pr(\theta \in T(\mathbf{y}) | \mathbf{y}) = \int_{T(\mathbf{y})} p_n(\theta) d\theta, \quad \text{onde} \\ T(\mathbf{y}) &= \{\theta \in \Theta : s(\theta) \geq s^*\} \end{aligned} \tag{3.4}$$

**Definição 3.5** A versão invariante do FBST é um procedimento que rejeita a hipótese precisa  $H$  quando  $ev(H) = 1 - \bar{ev}(H)$  é pequeno.

Observe que a definição original da medida de evidência do FBST é um caso particular da definição invariante tomando a densidade de referência como uma priori não-informativa uniforme,  $r(\theta) \propto 1$ . Aliás, uma priori não-informativa, interpretada como uma representação de pouca ou mínima informação no espaço paramétrico,

é a escolha natural para uma densidade de referência (STERN e PEREIRA, 2014; STERN, 2005). Nesta tese será utilizada a formulação original, considerando a densidade de referência trivial não-informativa.

Para uma discussão detalhada sobre o papel desempenhado no FBST pela densidade de referência e pela função surpresa, e como elas são usadas para se alcançar a propriedade fundamental de invariância da resultante medida de significância, vide STERN e PEREIRA (2014).

Por definição, o e-valor de uma hipótese precisa  $H$  é uma função de conjunto, porém, baseado no caso mais favorável, isto é,  $ev(H) = ev(\{\theta^*\})$ , uma propriedade que caracteriza  $ev(H)$  como cálculo abstrato de crença de tipo *possibilístico* (BORGES e STERN, 2007). Ao permitir o uso de uma densidade de referência não trivial,  $r(\theta)$ , e uma função surpresa,  $s(\theta)$ , que tem uma “forma distinta” da densidade a posteriori,  $p_n(\theta)$ , o e-valor,  $ev(H)$ , torna-se mais flexível, gerando uma medida de possibilidade explicitamente invariante de significância estatística (STERN e PEREIRA, 2014). Deste modo,  $ev(H)$  pode ser mantida *consistente* com a medida subjacente de  $p_n(H)$ , enquanto  $ev(\{\theta\})$  mantém-se *conforme* (similaridade em forma) com a função surpresa subjacente,  $s(\theta)$ , isto é,

$$\text{Consistência: } ev(H) \geq p_n(H), \forall H \subseteq \Theta;$$

$$\text{Conformidade: } ev(\{\theta\}) \geq ev(\{\tau\}) \Leftrightarrow s(\theta) \geq s(\tau), \forall \theta, \tau \in \Theta.$$

## 3.5 Nível crítico para rejeição da hipótese $H$

Como em qualquer teste de significância, o procedimento de FBST requer a escolha do nível crítico,  $c$ , para rejeição/aceitação da hipótese  $H$  a um nível de significância  $\alpha$ , ou seja, um ponto de corte que vai indicar até que ponto o  $ev(H)$  é “pequeno”. Para um certo valor  $c \in (0, 1)$ , um teste baseado no  $ev(H)$  rejeita a hipótese  $H$  se  $ev(H) < c$ , e aceita  $H$  caso contrário. Existem várias abordagens alternativas que podem ser utilizadas para estabelecer o nível crítico, vide LAURETTO *et al.* (2009) para lista de referências dos métodos. Neste trabalho, serão apresentados dois desses critérios: funções de perda e análise assintótica.

### 3.5.1 Funções de perda

Do ponto de vista da teoria estatística Bayesiana da decisão, um teste de significância é considerado compatível se, e somente se, ele pode ser caracterizado como um procedimento de decisão sobre aceitação ou rejeição da hipótese definido pela minimização da esperança a posteriori de uma função de perda (BORGES e STERN, 2007). MADRUGA *et al.* (2001) mostraram a existência de funções de perdas

específicas que caracterizam o FBST e o tornam um procedimento de teste de significância compatível com a teoria Bayesiana da decisão.

Considere a hipótese precisa  $H$  e o espaço de decisão  $D = \{d_0 = \text{Aceitar } H, d_1 = \text{Rejeitar } H\}$ .

**Definição 3.6** A função de perda  $L$ , também denominada  $LP_1$ , definida sob o espaço  $D \times \Theta \times \mathcal{Y}$ , é dada por

$$\begin{aligned} L(d_1, \theta) &= a\{1 - \mathbf{I}[\theta \in T(\mathbf{y})]\}, \quad a > 0 \quad \text{e} \\ L(d_0, \theta) &= b + c\{\mathbf{I}[\theta \in T(\mathbf{y})]\}, \quad b \geq 0, c > 0. \end{aligned}$$

Os autores demonstram que o risco a posteriori tendo como decisão aceitar  $H$  e o risco a posteriori de rejeitar  $H$  são, respectivamente, dadas por

$$\begin{aligned} \mathbf{E}[L(d_0, \theta)] &= b + c[1 - \text{ev}(H)] \\ \mathbf{E}[L(d_1, \theta)] &= a \text{ev}(H). \end{aligned}$$

Portanto, a aceitação da hipótese  $H$  se dará se, e somente se,  $\mathbf{E}[L(d_0, \theta)] < \mathbf{E}[L(d_1, \theta)]$ , ou seja, se

$$\text{ev}(H) > \frac{b + c}{a + c}.$$

### 3.5.2 Análise assintótica

Embora se acredite que, sob perspectiva Bayesiana, a propriedade da convergência assintótica (ou da distribuição amostral) da medida de evidência do FBST seja irrelevante (devido às propriedades do e-valor anunciadas na seção 3.1), existem aspectos matemáticos interessantes que podem ser explorados, uma vez que a distribuição a posteriori,  $p_n(\theta)$ , depende do ponto da amostra a ser observada, ou seja, o e-valor é uma estatística típica, uma função da amostra observada (PEREIRA *et al.*, 2008).

A convergência assintótica da medida de evidência do FBST, aqui apresentada, está demonstrada nos trabalhos de STERN (2005) e BORGES e STERN (2007). Considere a função de distribuição da evidência contra uma hipótese,  $\bar{V}(c) = \Pr(\bar{\text{ev}} \leq c)$ . Dado  $\theta^0$ , o valor verdadeiro do parâmetro, sob condições apropriadas de regularidade, para tamanho de amostra crescente,  $n \rightarrow \infty$ , pode-se afirmar que:

- Se  $H$  é falsa,  $\theta^0 \notin H$ , então  $\bar{\text{ev}}(H)$  converge (em probabilidade) para 1, i.e.,  $\bar{V}(0 < c < 1) \rightarrow 0$ . Logo  $\bar{\text{ev}}(H)$  é uma estatística consistente;
- Se  $H$  é verdadeira,  $\theta^0 \in H$ , então  $\bar{V}(c)$ , o nível de confiança, também denotado por  $1 - \alpha$ , converge (em distribuição) para função

$$Q(t, h, c) = F_{t-h}[F_t^{-1}(c)],$$

onde

$$F_g(y) = \frac{\Gamma(g/2, y/2)}{\Gamma(g/2, \infty)}, \quad \Gamma(g, y) = \int_0^y x^{g-1} e^{-x} dx;$$

$t = \dim(\Theta)$ ,  $h = \dim(H)$ ,  $F_g(x)$  é a função densidade acumulada da distribuição qui-quadrada com  $g$  graus de liberdade e  $F^{-1}$  é o quantil da qui-quadrado.

Sob as mesmas condições de regularidade, uma escolha apropriada do nível crítico,  $c(n)$ , fornece um teste consistente,  $\tau_c$ , que rejeita a hipótese nula  $H$  se a evidência contra  $H$  estiver acima do nível crítico,  $\bar{e}v(H) > c$ , e aceita  $H$  caso contrário. O nível crítico,  $c$ , que conduz a rejeição de  $H$  com um nível de confiança,  $1 - \alpha$ , usando o critério assintótico, é estabelecido como  $c = F_t[F_{t-h}^{-1}(1 - \alpha)]$ , ou seja, escolhe-se  $c$  tal que  $Q(t, h, c) = 1 - \alpha$ .

DINIZ *et al.* (2012) mostram que, assintoticamente, existe uma relação entre  $\bar{e}v(H)$  e o p-valor baseado no teste da razão de verossimilhança. Assim, pela propriedade de normalidade assintótica,  $\bar{e}v(H) \approx F_t[F_{t-h}^{-1}(1 - \text{p-valor})]$ . Invertendo esta equação, tem-se, então, que

$$\text{p-valor} = 1 - F_{t-h}[F_t^{-1}(1 - \text{ev}(H))]. \quad (3.5)$$

Observe que, no cálculo da probabilidade do conjunto tangente, utiliza-se a função de distribuição. Mas, pelo teorema fundamental da probabilidade/estatística, sob a hipótese nula sendo verdadeira, esta função aplicada na variável aleatória tem distribuição uniforme  $(0, 1)$ . Assim, ao fixar um corte de  $c \times 100\%$  (uma proporção para indicar que houve erro), a distribuição amostral da medida de evidência converge, sob a hipótese nula verdadeira, para a mesma proporção de erro.

Outro critério de definição do nível crítico, desenvolvido em STERN e ZACKS (2002) e LAURETTO *et al.* (2003), é baseado na análise de poder empírico, que busca minimizar uma combinação entre os dois tipos de erros: *Erro do tipo I*- a taxa de rejeição da hipótese  $H$  quando esta é verdadeira, e *Erro do tipo II*- a taxa de aceitação de  $H$  quando esta é falsa. Este critério fornece níveis críticos que são consistentes e também efetivos para amostras pequenas. Por outro lado, seu custo computacional pode ser elevado quando a dimensão do espaço paramétrico é alta.

Ainda segundo STERN e ZACKS (2002), a função poder do teste FBSST para  $\theta \in \Theta \setminus \Theta_H$  é

$$\psi(\theta) = \Pr_\theta(\bar{e}v(H) > c_\alpha),$$

onde

$$c_\alpha = \sup_{\theta \in \bar{H}} q_\alpha(\theta),$$

$q_\alpha(\theta)$  é o quantil  $(1 - \alpha)$  da distribuição aproximada de  $\bar{e}\bar{v}(H)$ .

### 3.6 Notas Bibliográficas

Os trabalhos de WEST e HARRISON (1997) e BASU (1996) podem ser vistos como precursores mais próximos do FBST. No entanto, estes autores se restringem às hipóteses pontuais e não apresentam uma solução invariante dos testes. A invariância é uma propriedade fundamental de medidas de significância bem definidas (STERN e PEREIRA, 2014). West e Harrison definem um conjunto de credibilidade similar ao conjunto tangente de PEREIRA e STERN (1999), porém, o fazem sobre uma futura observação, ou seja, trabalham no espaço amostral e não no espaço paramétrico, o que também é válido se a predição for considerada como uma avaliação de um novo parâmetro, a futura observação invisível (variável aleatória latente).

Ao contrário dos trabalhos citados, a medida de evidência do FBST é calculada no espaço paramétrico original, em sua plena dimensão, sendo, por isso, uma medida possibilística que satisfaz ótimas propriedades (como, por exemplo, a consistência e a conformidade) (STERN e PEREIRA, 2014). A grande vantagem do e-valor é o fato de ser uma medida invariante sob transformações gerais do sistema de coordenadas no espaço paramétrico e não só lineares (MADRUGA *et al.*, 2003). O FBST é aplicável não só quando a hipótese precisa é um ponto na reta, mas também quando ela é definida por um subespaço (de dimensão menor do que o espaço de parâmetro original), e assim, a definição do e-valor compara conjuntos de credibilidade com o máximo deste subespaço. A revisão geral do FBST e referências bibliográficas constam nos trabalhos de PEREIRA *et al.* (2008) e STERN e PEREIRA (2014). Algumas aplicações interessantes ilustrando o uso do e-valor e do FBST em problemas relevantes de inferência estatística podem ser encontradas em DINIZ *et al.* (2012), LAURETTO *et al.* (2003), LAURETTO *et al.* (2009), LAURETTO (2007), LAURETTO e STERN (2005) e PEREIRA e STERN (1999).

Em sua tese de doutorado, DE ARRUDA (2012) apresentou abordagens analítica e computacional para a extensão do FBST no contexto da teoria da Decisão sequencial, tendo em vista a redução do custo computacional associado à operação das etapas de otimização numérica e integração multidimensional. Recentemente, CABRAS *et al.* (2015) propuseram uma aproximação (assintótica) da medida de evidência fornecida pelo FBST para hipótese nula precisa na presença de parâmetros molestos ou perturbadores. Esta abordagem é baseada na aproximação da área da

cauda de ordem superior (HOTA) da distribuição a posteriori marginal do parâmetro de interesse. D’CUNHA e RAO (2016) compararam o poder frequentista do FBST com outros testes de significância Bayesiano para teste da mediana da distribuição lognormal.

O trabalho de LAURETTO e STERN (2005) foi o primeiro a explorar o procedimento do FBST para o problema de seleção de modelos através da mistura. Com objetivo de determinar o número de componentes (ou classes) da população a partir de uma amostra, os autores formularam um modelo de mistura finita de componentes normais multivariadas. No que se refere a hipóteses separadas, LAURETTO *et al.* (2007) formularam e implementaram o FBST, num contexto de mistura, para discriminação entre as distribuições Weibull e Gompertz. No entanto, nesse trabalho, a análise Bayesiana do modelo de mistura é feita sob o ponto de vista de abordagem “padrão”, ou seja, considera-se pares de parâmetros diferentes para cada uma das distribuições e assume-se prioris independentes para cada par de parâmetros.

# Capítulo 4

## Mistura de modelos separados

### 4.1 Introdução

A mistura finita de modelos é vista como uma representação conveniente da densidade de probabilidade, quando não se encontram distribuições *standard* de probabilidades capazes de se ajustar suficientemente bem aos dados (por exemplo, quando a distribuição dos dados possui vários máximos locais). Os modelos de mistura são frequentemente aplicados a problemas de classificação, clusterização e reconhecimento de padrões (BÖHNING *et al.*, 2007; ZHANG e HUANG, 2015).

Em testes para famílias separadas de hipóteses, a mistura de modelos foi sugerida por COX (1961). Esta abordagem permite a utilização de um procedimento de teste de hipótese sobre os pesos da mistura para decidir entre modelos alternativos candidatos (KAMARY *et al.*, 2014). Neste trabalho, o procedimento do FBST é formulado para discriminar entre as distribuições de famílias separadas a partir da mistura linear de modelos (QUANDT, 1974), ou seja, os modelos das hipóteses a serem confrontadas são considerados como componentes de um modelo de mistura.

Neste capítulo é desenvolvida toda metodologia referente à formulação do FBST para hipóteses separadas num contexto de mistura dos modelos candidatos. Na seção 4.2 são apresentadas, brevemente, as definições básicas de análise de sobrevivência e suas funções, destacando as famílias de distribuições de probabilidade consideradas neste trabalho. Na sequência, a seção 4.3 apresenta as distribuições Multinomial e Dirichlet, que compõem a classe de distribuições a priori usadas neste trabalho para o cálculo do e-valor. Na seção 4.4 apresenta-se a formulação geral do modelo de mistura finita e o cálculo da distribuição a posteriori com base na estrutura de dados aumentados. Finalmente, são discutidos os detalhes envolvendo a formulação e implementação do FBST para a mistura de modelos separados. O foco principal na implementação numérica do e-valor são as etapas de otimização e integração.

## 4.2 Análise de sobrevivência e suas funções

### 4.2.1 Definições

O termo análise de sobrevivência refere-se ao estudo de dados relacionados ao tempo até a ocorrência de um determinado evento de interesse, a partir de um tempo inicial até um tempo final, devidamente pré-definidos, conhecido como tempo de estudo ou janela de observação. Em estudos médicos, em geral, os eventos de interesse, que marcam término de acompanhamento de um paciente, são conhecidos como falhas. Como exemplo, cita-se o tempo até a morte de um paciente, o tempo até a ocorrência de uma determinada doença ou complicação da mesma, tempo de duração da doença, tempo até a cura, entre outros.

Seja  $T$  uma variável aleatória que representa o tempo de sobrevivência de um indivíduo, relacionado a um determinado evento de interesse. A distribuição de  $T$  pode ser caracterizada por três funções importantes, matematicamente equivalentes (LEE e WANG, 2003): a função de densidade de probabilidade, a função de sobrevivência e a função de risco.

A função de densidade de probabilidade,  $f(t)$ , é definida como sendo a probabilidade de um indivíduo sofrer um evento de interesse em um intervalo instantâneo de tempo, e pode ser escrita como

$$f(t) = \frac{dF(t)}{dt},$$

onde  $F(t)$  representa a função de distribuição de  $T$ . A função de sobrevivência, denotada por  $S(t)$ , representa a probabilidade de um indivíduo sobreviver até um determinado tempo,  $t$ , e pode ser escrita como

$$S(t) = 1 - F(t).$$

Note que  $S(t)$  é uma função contínua em  $t$  e monótona decrescente, tal que  $S(0) = 1$  e  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ . A função de risco,  $\lambda(t)$ , é definida como sendo a probabilidade condicional de um indivíduo sofrer o evento de interesse dentro de um intervalo de tempo pequeno  $(t, t + \Delta t)$ , sabendo que o mesmo sobreviveu até o início deste intervalo, ou seja,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (4.1)$$

A função de risco acumulado do tempo de sobrevivência  $T$ , denotada por  $\Lambda(t)$ , é definida por

$$\Lambda(t) = \int_0^t \lambda(u) d(u) \quad (4.2)$$

Portanto, quando  $t = 0$ ,  $S(t) = 1$  e  $\Lambda(t) = 0$ , e quando  $t \rightarrow \infty$ ,  $S(t) = 0$  e  $\Lambda(t) = \infty$ , isto é, a função de risco acumulado pode assumir qualquer valor de zero a infinito.

## 4.2.2 Distribuições de probabilidade de interesse

Os métodos estatísticos paramétricos são ferramentas muito úteis em análise de sobrevivência, desde que a distribuição de probabilidade selecionada se ajuste adequadamente aos dados (MOHAMMED *et al.*, 2013). Nesta tese, o procedimento do FBST é proposto para discriminar entre os modelos estatísticos das distribuições lognormal, Weibull e gama, as mais utilizadas na literatura para modelagem de dados de sobrevivência (LAWLESS, 2002; LEE e WANG, 2003; PEREIRA, 1978; PEREIRA e PEREIRA, 2017). As funções de densidade de probabilidade,  $f(y|\cdot)$ ; de sobrevivência,  $S(y|\cdot)$ ; e de risco ou taxa de falha,  $h(y|\cdot)$ , dessas famílias de distribuições de probabilidade são apresentadas a seguir.

a) Para a distribuição lognormal com vetor de parâmetros  $\alpha = (\alpha_1, \alpha_2)$ ,

$$f_L(y|\alpha) = \frac{1}{y\sqrt{2\pi\alpha_2}} \exp\left\{-\frac{(\log y - \alpha_1)^2}{2\alpha_2}\right\}, \quad -\infty < \alpha_1 < \infty, \alpha_2, y > 0;$$

$$S_L(y|\alpha) = \frac{1}{\sqrt{2\pi\alpha_2}} \int_t^\infty \frac{1}{y} \exp\left\{-\frac{(\log y - \alpha_1)^2}{2\alpha_2}\right\} dy$$

$$= 1 - \Phi\left[\frac{(\log y - \alpha_1)}{\sqrt{\alpha_2}}\right];$$

$$h_L(y|\alpha) = \frac{f_{LN}(y|\alpha)}{S_{LN}(y|\alpha)}.$$

b) Para a distribuição Weibull com vetor de parâmetros  $\beta = (\beta_1, \beta_2)$ ,

$$f_W(y|\beta) = \frac{\beta_1}{\beta_2^{\beta_1}} y^{\beta_1-1} \exp\left\{-\left(\frac{y}{\beta_2}\right)^{\beta_1}\right\}, \quad \beta_1, \beta_2, y > 0;$$

$$S_W(y|\beta) = \exp\left\{-\left(\frac{y}{\beta_2}\right)^{\beta_1}\right\};$$

$$h_W(y|\beta) = \frac{\beta_1}{\beta_2^{\beta_1}} y^{\beta_1-1}.$$

c) Para a distribuição gama com vetor de parâmetros  $\gamma = (\gamma_1, \gamma_2)$

$$\begin{aligned} f_G(y|\gamma) &= \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}} y^{\gamma_1-1} \exp\left\{-\frac{y}{\gamma_2}\right\}, \quad \gamma_1, \gamma_2, y > 0; \\ S_G(y|\gamma) &= 1 - \int_0^y \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}} u^{\gamma_1-1} \exp\left\{-\frac{u}{\gamma_2}\right\} du; \\ h_G(y|\gamma) &= \frac{f_G(y|\gamma)}{S_G(y|\gamma)}. \end{aligned}$$

### 4.3 Distribuições Multinomial e Dirichlet

As distribuições Multinomial e Dirichlet, apresentadas nesta seção, compõem a classe de distribuições a priori que serão usadas para a modelagem da incerteza a priori dos parâmetros da mistura, no cálculo da medida de evidência do FBST.

#### a) Distribuição Multinomial

Seja  $y = (y_1, y_2, \dots, y_m)$  um vetor aleatório cujos componentes tomam valores inteiros não negativos tal que  $\sum_{k=1}^m y_k = n$  está fixado. Diz-se que  $y$  possui uma distribuição Multinomial com parâmetros  $n$  e  $w$ , simbolicamente,

$$y|n, w \sim M_m(n, w),$$

se qualquer subconjunto de  $m - 1$  componentes de  $y$  tiver a seguinte a função de densidade

$$f(y|n, w) = \frac{n!}{y_1! \dots y_m!} \prod_{k=1}^m w_k^{y_k} \quad (4.3)$$

onde  $w = (w_1, \dots, w_m)$  com  $w_k > 0$  e  $\sum_{k=1}^m w_k = 1$ .

Na prática, esta distribuição tem dimensionalidade  $m - 1$ , embora, por abuso de notação (praticado por conveniência), se considere todo o vetor  $y$ . Quando  $m = 2$ , a função densidade (4.3) representa a família de distribuições Binomial, isto é,  $y_1 \sim Bi(n, w_1)$ .

A função geradora de momentos da distribuição Multinomial é

$$M_y(t) = \mathbf{E}(e^{ty}|n, w) = \left( \sum_{k=1}^{m-1} w_k e^{t_k} + w_m \right)^n, \quad (4.4)$$

onde  $t = (t_1, \dots, t_{m-1}, 0)$ . A partir de (4.4), os momentos são definidos da seguinte

forma:

$$\text{Expectâncias : } \mathbf{E}(y_k|n, w_k) = nw_k, \quad k = 1, \dots, m,$$

$$\text{Variâncias : } \text{Var}(y_k|n, w_k) = nw_k(1 - w_k), \quad k = 1, \dots, m,$$

$$\text{Covariâncias : } \text{Cov}(y_i, y_j|n, w_i, w_j) = -nw_iw_j, \quad i \neq j; \quad i, j = 1, \dots, m,$$

$$\text{Correlação : } \text{Cor}(y_i, y_j|w_i, w_j) = -\sqrt{\frac{w_iw_j}{(1-w_i)(1-w_j)}}, \quad \text{para } i \neq j.$$

Na forma matricial, os dois primeiros momentos do vetor  $y$  são dados por

$$\begin{aligned} \mathbf{E}(y|n, w) &= nw \quad \text{e} \\ \text{Var}(y|n, w) &= n(D_w - ww^\top) \end{aligned}$$

onde  $D_w = \text{diag}(w_1, \dots, w_m)$ .

## b) Distribuição Dirichlet

Seja  $c$  um número positivo. O simplex (fechado)  $m$ -dimensional em  $\mathbf{R}^m$  e simplex (aberto)  $(m-1)$ -dimensional em  $\mathbf{R}^{m-1}$  são definidos por

$$\begin{aligned} S_m(c) &= \left\{ (w_1, \dots, w_m) : w_i > 0 \wedge \sum_{k=1}^m w_k = c \right\} \quad \text{e} \\ S_{m-1}(c) &= \left\{ (w_1, \dots, w_{m-1}) : w_i > 0 \wedge \sum_{k=1}^{m-1} w_k < c \right\}, \end{aligned}$$

respectivamente. A bem da simplicidade, será usada uma notação relaxada, definindo  $S_m = S_m(1)$  e  $S_{m-1} = S_{m-1}(1)$ .

Um vetor aleatório  $w = (w_1, \dots, w_m) \in S_m$  possui uma distribuição Dirichlet com parâmetro  $a = (a_1, \dots, a_m) \in \mathbf{R}_+^m$ , representada simbolicamente por

$$w|a \sim D_m(a),$$

se a função de densidade de qualquer subconjunto de  $m-1$  componentes de  $w$  definido em  $S_{m-1}$  é expressa por (de novo numa notação sobredimensionada, por conveniência)

$$h(w|a) = [B(a)]^{-1} \prod_{k=1}^m w_k^{a_k-1},$$

onde

$$B(a) = \int_{S_{m-1}} \prod_{k=1}^m w_k^{a_k-1} dw = \frac{1}{\Gamma(A)} \prod_{k=1}^m \Gamma(a_k),$$

tendo  $A = \sum_{k=1}^m a_k$ , como fator de normalização da distribuição – a função beta multivariada (integral de Dirichlet  $(m - 1)$ -dimensional). A família de distribuições Dirichlet é uma generalização da família distribuições Beta, que se reduz quando  $m = 2$ , isto é,  $w_1 \sim B(a_1, a_2)$ . Os momentos de  $w$  podem ser obtidos a partir da fórmula

$$\mathbf{E} \left[ \prod_{k=1}^m w_k^{r_k} | a \right] = \frac{B(a + r)}{B(a)},$$

onde  $r = (r_1, \dots, r_m)$ . Em particular, denotando  $A = \sum_{k=1}^m a_k$ , tem-se:

$$\text{Expectâncias : } \mathbf{E}(w_k | a) = \frac{a_k}{A}, \quad k = 1, \dots, m,$$

$$\text{Variâncias : } \text{Var}(w_k | a) = \frac{a_k(A - a_k)}{A^2(1 + A)}, \quad k = 1, \dots, m$$

$$\text{Covariâncias : } \text{Cov}(w_i, w_j | a) = -\frac{a_i a_j}{A^2(1 + A)}, \quad i \neq j; \quad i, j = 1, \dots, m,$$

$$\text{Correlação : } \text{Cor}(w_i, w_j | a) = -\sqrt{\frac{a_i a_j}{(A - a_i)(A - a_j)}}, \quad i \neq j; \quad i, j = 1, \dots, m.$$

Na forma matricial, os dois primeiros momentos do vetor  $w$  são dados por

$$\mathbf{E}(w | a) = \frac{a}{A} \quad \text{e}$$

$$\text{Var}(w | a) = \frac{1}{A(1 + A)} \left( \text{diag}(a) - \frac{1}{A} a a^\top \right)$$

### c) Distribuição Multinomial-Dirichlet

Se  $y | n, w \sim M_m(n, w)$  e  $w | a \sim D_m(a)$ , então a distribuição marginal de  $y$  é denominada distribuição Multinomial-Dirichlet (Binomial-Dirichlet quando  $m = 2$ ), denotando-se por  $y | n, a \sim MD_m(n, a)$ . Formalmente, um vector aleatório  $y = (y_1, y_2, \dots, y_m) \in S_m(n)$  possui uma distribuição Multinomial-Dirichlet se a função de densidade de qualquer subconjunto de  $m - 1$  componentes de  $y$  definido em  $S_{m-1}(n)$  é expressa (usando denotação relaxada) por

$$f(y | n, a) = \frac{n!}{\prod_{k=1}^m y_k!} \frac{B(a + n)}{B(a)}.$$

Para maiores detalhes sobre a distribuição Multinomial-Dirichlet e outras distribuições relacionadas, vide NG *et al.* (2011).

Observe que a distribuição Dirichlet é uma distribuição a priori conjugada natural para os parâmetros de uma distribuição Multinomial. Portanto, a distribuição a

posteriori condicional de  $w$  é uma distribuição Dirichlet, representada por

$$\begin{aligned} p(w|y, a) &= \frac{f(y|w)\pi(w|a)}{\int f(y|w)\pi(w|a)dw} \\ &\propto \frac{\Gamma(A+n)}{\prod_{k=1}^m \Gamma(y_k + a_k)} \prod_{k=1}^m w_k^{y_k + a_k - 1}. \end{aligned}$$

Ou seja, se  $y|n, w \sim M_m(n, w)$  e  $w|a \sim D_m(a)$ , então

$$w|y, a \sim D_m(a + y).$$

## 4.4 Modelo de mistura finita

Considere uma variável aleatória (vetorial),  $\mathbf{y} = \{y_1, \dots, y_n\}$ , assumindo valores no espaço amostral  $\mathcal{Y} \subset \mathcal{R}^q$ ,  $q \geq 1$ , e  $m$  distribuições de probabilidade,  $F_1(\mathbf{y}|\psi_1), \dots, F_m(\mathbf{y}|\psi_m)$ , tal que  $F_k \in \mathfrak{F}_k$ , onde  $\mathfrak{F}_k$  é uma família de distribuições com densidade  $f_k(\mathbf{y}|\psi_k)$  e  $\psi_k$  é o parâmetro (possivelmente um vetor) desconhecido. Diz-se que  $y$  provem da mistura finita de distribuições se a função densidade de probabilidade de cada observação  $y_j$  é representada como uma combinação linear de várias densidades, ou seja,

$$f(y_j|\boldsymbol{\theta}) = p_1 f_1(y_j|\psi_1) + \dots + p_m f_m(y_j|\psi_m), \quad p_k \geq 0, \quad \sum_{k=1}^m p_k = 1, \quad (4.5)$$

onde o vetor  $\boldsymbol{\theta} = (\mathbf{p}, \boldsymbol{\psi})$ ,  $\mathbf{p} = (p_1, \dots, p_m)$  e  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_m)$ , contém todos os parâmetros desconhecidos do modelo de mistura que, portanto, devem ser inferidos a partir dos dados disponíveis. As quantidades  $p_1, \dots, p_m$  são chamadas de pesos ou proporções da mistura e  $m$  o número especificado de componentes da mistura.

Assumindo que as observações são condicionalmente (aos parâmetros) independentes, a verossimilhança correspondente a todas as partições possíveis da amostra  $\mathbf{y}$  para  $m$  componentes é dada por

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{j=1}^n f(y_j|\boldsymbol{\theta}) = \prod_{j=1}^n \sum_{k=1}^m p_k f_k(y_j|\psi_k). \quad (4.6)$$

A estimação dos parâmetros do modelo de mistura pode ser efetuada de forma direta ou através da distribuição a posteriori obtida com base no método de dados aumentados (também chamado de método de dados completos). Na inferência Bayesiana, a estratégia de aumento de dados é empregada juntamente com métodos de Monte Carlo via Cadeia de Markov (MCMC) para facilitar, do ponto de vista de implementação computacional, o processo de estimação dos parâmetros envolvidos

no modelo (DIEBOLT e ROBERT, 1994; FRUHWIRTH-SCHNATTER, 2006).

Na formulação de um modelo de mistura com base nos dados aumentados, cada observação  $y_j$  é associada a um determinado componente (distribuição de probabilidade). Dessa forma, para cada  $y_j$ , introduz-se um vetor  $m$ -dimensional de classificação,  $z_j$ , tal que  $z_j^k \in \{0, 1\}$  e  $\sum_{k=1}^m z_j^k = 1$ , sendo que  $z_j^k = 1$  se, e somente se,  $y_j$  pertence à classe  $k$ . Uma observação  $y_j$  de classe  $k = c(j)$  tem densidade de probabilidade  $f(y_j|\psi_k)$ . Dado o vetor  $\boldsymbol{\theta}$ , a especificação do modelo assume que as variáveis de classificação são independentes e  $\Pr(z_j^k = 1|\boldsymbol{\theta}) = p_k$  (FRUHWIRTH-SCHNATTER, 2006; TIAN *et al.*, 2010). Denota-se por  $\mathbf{z} = \{z_j^k\}$  a matriz de classificação composta por dados de variáveis não observáveis, por vezes denominadas *variáveis latentes* ou *missing data*.

A amostra  $\mathbf{y}$ , composta por dados observados, é aumentada com dados latentes  $\mathbf{z}$  formando uma matriz de dados completos,  $(\mathbf{y}, \mathbf{z})$ , também chamada de matriz de dados aumentados, cuja verossimilhança é dada por

$$\begin{aligned} f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) &= f(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta})f(\mathbf{z}|\boldsymbol{\theta}) \\ &= \prod_{j=1}^n f(y_j|z_j, \boldsymbol{\theta})f(z_j|\boldsymbol{\theta}). \end{aligned}$$

Visto que  $f(y_j|z_j^k = 1, \boldsymbol{\theta}) = f_k(y_j|\psi_k)$  e  $\Pr(z_j^k = 1|\boldsymbol{\theta}) = p_k$ , então

$$\begin{aligned} f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) &= \prod_{j=1}^n \prod_{k=1}^m [p_k f_k(y_j|\psi_k)]^{z_j^k} \\ &= \prod_{k=1}^m \left[ p_k^{x_k} \prod_{j|c(j)=k} f_k(y_j|\psi_k) \right], \end{aligned} \quad (4.7)$$

onde,  $x_k$  representa o número de exemplos  $y_j$  de classe  $k$ , ou seja,  $x_k = |\{y_j : c(j) = k\}| = \sum_j z_j^k$ , ou  $x = \mathbf{z}\mathbf{1}$ , sendo  $\mathbf{1}$  vetor unitário  $n$ -dimensional.

Sob perspectiva Bayesiana,  $\mathbf{z}$  é tratada como um vetor de parâmetros, uma vez que não é observada. Portanto, assim como o vetor dos parâmetros  $\boldsymbol{\theta}$ ,  $\mathbf{z}$  deve ser estimada a partir dos dados.

Dados os parâmetros da mistura,  $\boldsymbol{\theta}$ , e os dados observados,  $\mathbf{y}$ , e assumindo independência entre as classificações individuais das observações, a densidade a posteriori de  $\mathbf{z}$ ,  $S = f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$ , é escrita como

$$f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}) = \prod_{j=1}^n f(z_j|\boldsymbol{\theta}, y_j), \quad (4.8)$$

onde  $f(z_j|\boldsymbol{\theta}, y_j)$  é a densidade da distribuição multinomial de classificação a

posteriori individual, i.e,  $z_j|\boldsymbol{\theta}, y_j \sim M(1, s_j^1, \dots, s_j^m)$ , sendo

$$s_j^k = \Pr(z_j^k = 1|y_j, \boldsymbol{\theta}) = \frac{f_k(y_j, z_j^k|\boldsymbol{\theta})}{f(y_j|\boldsymbol{\theta})} = \frac{p_k f_k(y_j|\psi_k)}{\sum_{k=1}^m p_k f_k(y_j|\psi_k)} \quad (4.9)$$

a probabilidade condicional de classificação.

A distribuição a posteriori conjunta  $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$  é obtida através da combinação da verossimilhança para os dados completos,  $f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})$ , com a distribuição a priori  $\pi(\boldsymbol{\theta})$ ,

$$f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) = \frac{f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \propto f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Assumindo independência a priori entre os parâmetros da mistura,  $\boldsymbol{\theta} = (\mathbf{p}, \boldsymbol{\psi})$  (FRUHWIRTH-SCHNATTER, 2006; TIAN *et al.*, 2010), a densidade a priori  $\pi(\boldsymbol{\theta})$  é definida como

$$\pi(\boldsymbol{\theta}) = \pi_0(\mathbf{p}) \prod_{k=1}^m \pi_k(\psi_k). \quad (4.10)$$

Substituindo (4.7) e (4.10) em (4.4) obtém-se

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) &\propto \prod_{k=1}^m \left[ p_k^{x_k} \prod_{j|c(j)=k} f_k(y_j|\psi_k) \right] \pi(\mathbf{p}) \prod_{k=1}^m \pi(\psi_k) \\ &= \prod_{k=1}^m f(\psi_k|\mathbf{y}, \mathbf{z}) f(\mathbf{p}|\mathbf{z}), \end{aligned} \quad (4.11)$$

onde

$$f(\psi_k|\mathbf{y}, \mathbf{z}) \propto \prod_{j|c(j)=k} f_k(y_j|\psi_k)\pi(\psi_k) \quad (4.12)$$

e

$$f(\mathbf{p}|\mathbf{z}) \propto \prod_{k=1}^m p_k^{x_k} \pi(\mathbf{p}) \quad (4.13)$$

são as densidades das condicionais completas dos blocos de parâmetros. Dessa forma, o processo de estimação dos parâmetros de cada componente,  $\psi_k$ , e da distribuição dos pesos da mistura,  $\mathbf{p}$ , pode ser realizada de forma independente (FRUHWIRTH-SCHNATTER, 2006).

A distribuição condicional completa (4.13), da qual são gerados os valores de  $\mathbf{p} = (p_1, \dots, p_m)$ , é definida combinando a verossimilhança  $f(\mathbf{z}|\mathbf{p})$  com a distribuição a priori  $\pi(\mathbf{p})$ , ou seja,

$$f(\mathbf{p}|\mathbf{z}) \propto f(\mathbf{z}|\mathbf{p})\pi(\mathbf{p}).$$

Da expressão (4.13), nota-se que  $f(\mathbf{z}|\mathbf{p})$  é um núcleo da densidade da família de

distribuições Multinomial com parâmetro  $\mathbf{p}$ . Nesse caso, a distribuição a priori Dirichlet é uma escolha natural para o vetor  $\mathbf{p}$  (DIEBOLT e ROBERT, 1994; MCLACHLAN e PEEL, 2000), i.e,

$$\pi(\mathbf{p}) \propto \prod_{k=1}^m p_k^{\dot{x}_k - 1}.$$

Logo,

$$\begin{aligned} f(\mathbf{p}|\mathbf{z}) &\propto f(\mathbf{z}|\mathbf{p})\pi(\mathbf{p}) \\ &\propto \prod_{k=1}^m p_k^{x_k + \dot{x}_k - 1} \end{aligned} \quad (4.14)$$

é a densidade da distribuição Dirichlet,  $\mathbf{p}|\mathbf{z} \sim D(\ddot{x})$ , onde  $\ddot{x} = (\ddot{x}_1, \dots, \ddot{x}_m)$ ,  $\ddot{x}_k = x_k + \dot{x}_k$ ,  $k = 1, \dots, m$ .

O algoritmo de dados aumentados permite simular iterativamente os parâmetros  $\mathbf{z}$  e  $\boldsymbol{\theta}$  de acordo com as distribuições (4.8) e (4.11), respectivamente. Especificando um valor inicial  $\boldsymbol{\theta}^{(0)}$ , o algoritmo executa a  $i$ -ésima iteração como se segue(DIEBOLT e ROBERT, 1994):

1. gere  $\mathbf{z}^{(i)} \sim f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}^{(i)})$ ;
2. gere  $\psi_k^{(i+1)} \sim f(\psi_k|\mathbf{y}, \mathbf{z}^{(i)})$ ;
3. gere  $\mathbf{p}^{(i+1)} \sim f(\mathbf{p}|\mathbf{z}^{(i)})$ .

No primeiro passo, as variáveis latentes  $z_j$  são geradas de acordo com a distribuição multinomial, com probabilidades definidas em (4.9). No segundo passo, os parâmetros do componente  $k$  são gerados de acordo com a condicional completa correspondente, dada por (4.12). Finalmente, o vetor dos pesos da mistura é gerado da condicional completa (4.13), ou seja,  $\mathbf{p}^{(i+1)} \sim Dirchlet(x_1 + \dot{x}_1, \dots, x_m + \dot{x}_m)$ .

Vale destacar que, no processo de estimação (direta ou baseada em dados aumentados) envolvendo mistura de modelos separados, a distribuição a posteriori dos parâmetros não possui forma analítica fechada. Pode-se, no entanto, de forma aproximada, gerar amostras da distribuição a posteriori dos parâmetros de interesse através de métodos de simulação estocástica (FRUHWIRTH-SCHNATTER, 2006). Entre estes, destacam-se os métodos adaptativos de MCMC (FIRMINO, 2009; ROBERTS e ROSENTHAL, 2009).

## 4.5 O FBST para mistura de modelos separados

Suponha que a função densidade definida em (4.5) seja uma mistura de  $m$  modelos cujas famílias de distribuições são separadas, i.e, cada componente  $k$  possui uma função densidade  $f(\cdot|\psi_k)$  de família distinta dos demais componentes. O objetivo é medir a evidência que os dados carregam em favor de cada modelo.

Neste trabalho, as funções densidades de probabilidade dos modelos candidatos são reparametrizadas em termos da média  $\mu$  e da variância  $\sigma^2$  da população. A opção pela reparametrização dos componentes da mistura em função de parâmetros comuns, deveu-se ao fato de a comparação entre as distribuições ser baseada no mesmo conjunto de observações, isto é, na mesma população. Por isso, entende-se que seria inconveniente considerar, nesses modelos, médias e variâncias diferentes e, ainda por cima, com prioris diferentes, como acontece, frequentemente, na análise Bayesiana de um modelo de mistura tradicional. Dessa forma, a função densidade da mistura dos modelos a serem confrontados é dada por

$$f(y_j|\mu, \sigma, \mathbf{p}) = p_1 f_1(y_j|\mu, \sigma) + \dots + p_m f_m(y_j|\mu, \sigma), \quad p_k \geq 0, \quad \sum_{k=1}^m p_k = 1,$$

e a verossimilhança será

$$f(\mathbf{y}|\mu, \sigma, \mathbf{p}) = \prod_{j=1}^n \sum_{k=1}^m p_k f_k(y_j|\mu, \sigma). \quad (4.15)$$

Uma vantagem de estabelecer a relação entre os parâmetros dos componentes da mistura através de  $\mu$  e  $\sigma^2$ , do ponto de vista de inferência, é a de reduzir a quantidade de parâmetros a serem estimados.

A seguir são apresentadas as equações dos parâmetros dos modelos citados na seção 4.2.2, expressas em função de  $\mu$  e  $\sigma^2$ .

i) Seja  $y \sim \text{Gama}(\gamma_1, \gamma_2)$ ,  $\gamma_1, \gamma_2 > 0$ . Então tem-se

$$\begin{cases} \mu = E(y|\gamma_1, \gamma_2) = \gamma_1 \gamma_2 \\ \sigma^2 = \text{Var}(y|\gamma_1, \gamma_2) = \gamma_1 \gamma_2^2 \end{cases} \Rightarrow \begin{cases} \gamma_1 = \frac{\mu^2}{\sigma^2} \\ \gamma_2 = \frac{\sigma^2}{\mu}, \end{cases} \quad (4.16)$$

ii) Seja  $y \sim \text{Lognormal}(\alpha_1, \alpha_2)$ ,  $-\infty < \alpha_1 < \infty, \alpha_2 > 0$ . Então tem-se

$$\begin{cases} \mu = E(y|\alpha_1, \alpha_2) = e^{\alpha_1 + \alpha_2/2} \\ \sigma^2 = \text{Var}(y|\alpha_1, \alpha_2) = (e^{\alpha_2} - 1)e^{2\alpha_1 + \alpha_2} \end{cases} \Rightarrow \begin{cases} \alpha_1 = \log \frac{\mu^2}{\sqrt{\mu^2 + \sigma^2}} \\ \alpha_2 = \sqrt{\log \frac{\mu^2 + \sigma^2}{\mu^2}}, \end{cases} \quad (4.17)$$

iii) Seja  $y \sim \text{Weibull}(\beta_1, \beta_2)$ ,  $\beta_1, \beta_2 > 0$ . Então tem-se

$$\begin{cases} \mu = E(y|\beta_1, \beta_2) = \beta_2 \Gamma(1 + 1/\beta_1) \\ \sigma^2 = \text{Var}(y|\beta_1, \beta_2) = \beta_2^2 \Gamma(1 + 2/\beta_1) - \beta_2^2 \Gamma^2(1 + 1/\beta_1) \end{cases}$$

$$\Rightarrow \begin{cases} 2 \log \Gamma(1 + 1/\beta_1) - \log \Gamma(1 + 2/\beta_1) + \log \frac{\mu^2 + \sigma^2}{\mu^2} = 0 \\ \beta_2 = \frac{\mu}{\Gamma(1 + 1/\beta_1)}. \end{cases} \quad (4.18)$$

O valor de  $\beta_1$  é calculado por um método numérico. Aqui, utilizou-se o método de Newton-Raphson, via função `nleqslv` do pacote com mesmo nome, do *software* R, versão 3.2.2.

Assumindo independência entre os parâmetros da mistura, a função densidade conjunta a priori para  $\boldsymbol{\theta} = (\mathbf{p}, \mu, \sigma^2)$  é definida da seguinte forma:

$$\pi(\boldsymbol{\theta}) = \pi_1(\mathbf{p})\pi_2(\mu)\pi_3(\sigma^2). \quad (4.19)$$

A função da verossimilhança (4.15) é combinada com a densidade conjunta a priori (4.19) para a obtenção da distribuição a posteriori  $f(\boldsymbol{\theta}|\mathbf{y})$ , usando a regra de Bayes, ou seja,

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \quad (4.20)$$

Em qualquer modelo de mistura considerado neste trabalho, assumiu-se uma priori vaga para os parâmetros  $\mu$  e  $\sigma^2$ , modelada pela distribuição gama de média 1 e variância 100, ou seja,  $\mu, \sigma^2 \sim \text{gamma}(0.01, 100)$  (GELMAN, 2006; PEREIRA e PEREIRA, 2017). Em relação aos pesos da mistura, para modelos com três componentes ( $m = 3$ ), considerou-se uma priori Dirichlet,  $\mathbf{p} \sim \text{Dirichlet}(1, 1, 1)$ ; e para  $m = 2$ ,  $p \sim \text{Beta}(1, 1)$  (uniforme(0, 1)).

Para medir as evidências em favor de cada modelo, as hipóteses sobre os pesos da mistura ( $p_k$ ) são testadas (COX, 1961; KAMARY *et al.*, 2014). Neste trabalho, o procedimento utilizado é o do FBST.

A hipótese de que  $y$  tem função densidade de probabilidade  $f_k(y|\psi_k)$  é equivalente à hipótese

$$H_k : p_k = 1 \wedge p_i = 0, i \neq k. \quad (4.21)$$

Por outro lado, a hipótese de que  $y$  não tem densidade  $f_k(y|\psi_k)$  é equivalente à

$$H : p_k = 0 \wedge \sum_{i \neq k} p_i = 1. \quad (4.22)$$

As hipóteses alternativas para (4.21) e (4.22) são, respectivamente,  $A_k : p_k < 1$

e  $A : p_k > 0$ , que não são precisas.

O cálculo da evidência a favor da hipótese  $H_k$ , com base no procedimento do FBST, consiste em dois passos:

1. **Otimização:** estimar o valor da máxima densidade a posteriori conjunta (4.20) sob  $H_k$ ,  $f_k^* = \max_{\boldsymbol{\theta} \in H_k} f(\boldsymbol{\theta}|\mathbf{y})$ ;
2. **Integração:** estimar o valor da evidência a favor da hipótese  $H_k$ , integrando a função densidade a posteriori conjunta nos pontos do espaço paramétrico com densidade inferior à probabilidade máxima obtida no passo de otimização, ou seja,

$$\text{ev}(H_k) = \Pr(\boldsymbol{\theta} \in T_k(\mathbf{y})|\mathbf{y}) = \int_{T_k(\mathbf{y})} f(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad \text{onde}$$

$$T_k(\mathbf{y}) = \{\boldsymbol{\theta} \in \Theta : f(\boldsymbol{\theta}|\mathbf{y}) \leq f_k^*\}.$$

Usando um procedimento de simulação Monte Carlo, obtém-se a aproximação do e-valor, conforme descrito a seguir:

- simula-se uma amostra de  $N$  pontos com distribuição igual a  $f(\boldsymbol{\theta}|\mathbf{y})$ ;
- contam-se os ( $q$ ) pontos da amostra simulada que possuem densidade a posteriori menor que  $f_k^*$ , ou seja,

$$q = \sum_{i=1}^N \mathbb{I}\{f(\boldsymbol{\theta}_i|\mathbf{y}) \leq f_k^*\}; \quad \text{e}$$

- calcula-se a proporção  $\frac{q}{N}$ , que corresponde à aproximação de  $\text{ev}$ .

Para o passo de otimização, foi utilizado o método de gradiente conjugado (FLETCHER e REEVES, 1964; NAZARETH, 2009), via função `optim`, que pertence ao pacote `stats` do software R, versão 3.2.2.

No passo de integração, foi adotado o algoritmo Metropolis adaptativo (AM, do inglês *Adaptive Metropolis*), proposto por HAARIO *et al.* (2001), para simular as amostras da distribuição a posteriori dos parâmetros de interesse. Este método adaptativo do MCMC (ROBERTS e ROSENTHAL, 2009) permite que a cadeia seja iterada mesmo diante de uma distribuição a posteriori com forma analítica não-padrão (cuja amostragem direta é inviável), através da distribuição proposta alternativa.

Foi simulada, para cada parâmetro, uma cadeia com 80 mil valores da distribuição a posteriori. Após descartar cinco mil valores iniciais (*burn-in*) e tomando um salto de tamanho cinquenta, foi obtida uma amostra final de 1500 observações utilizadas para calcular as estimativas e a medida de evidência do FBST.

Aqui, a implementação do modelo Bayesiano foi desenvolvida usando o pacote `LaplacesDemon`, do software R. O `LaplacesDemon` é um pacote estatístico gratuito, de código-fonte aberto e livremente distribuído que proporciona um ambiente completo para simulação em inferência Bayesiana (STATISTICAT, LCC, 2016). .

# Capítulo 5

## Simulações e Exemplos de aplicações

### 5.1 Simulações

Esta seção apresenta alguns resultados numéricos baseados na simulação de pontos amostrais, com o objetivo de avaliar o comportamento do FBST na discriminação entre pares de distribuições de famílias separadas para diferentes tamanhos de amostras. Nessas simulações, foram comparados os resultados empíricos do FBST com os do teste de Cox e feita a descrição do padrão geral observado.

Os resultados dos testes foram obtidos usando o software R. Todos os programas computacionais foram rodados em um Intel(R) Core(TM) i7-5500U CPU@ 2.40GHz.

#### 5.1.1 Esquema para simulação de pontos amostrais

Sejam  $H_L$ ,  $H_W$  e  $H_G$  as hipóteses que especificam as funções densidades de probabilidade das famílias das distribuições lognormal, Weibull e gama, respectivamente,

$$H_L : f_L(y|\alpha) = \frac{1}{y\sqrt{2\pi\alpha_2}} \exp\left\{-\frac{(\log y - \alpha_1)^2}{2\alpha_2}\right\}, \quad \alpha = (\alpha_1, \alpha_2),$$
$$H_W : f_W(y|\beta) = \frac{\beta_1}{\beta_2^{\beta_1}} y^{\beta_1-1} \exp\left\{-\left(\frac{y}{\beta_2}\right)^{\beta_1}\right\}, \quad \beta = (\beta_1, \beta_2)$$
$$H_G : f_G(y|\gamma) = \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}} y^{\gamma_1-1} \exp\left\{-\frac{y}{\gamma_2}\right\}, \quad \gamma = (\gamma_1, \gamma_2).$$

Para cada par de hipóteses, foram simuladas 500 amostras de tamanhos  $n = 20, 40, 60, 80, 100, 150$  e  $200$  de cada distribuição e, para cada amostra, foram testadas as hipóteses através do FBST e do teste de Cox. O interesse nesses testes é verificar a convergência das taxas de decisões corretas, i.e, aceitação do modelo verdadeiro e

rejeição do modelo falso.

A medida de evidência do FBST é invariante com respeito a parametrizações da hipótese nula e com respeito a transformações regulares do sistema de coordenadas (MADRUGA *et al.*, 2003; PEREIRA *et al.*, 2008). Conforme discutido por DUMONCEAUX *et al.* (1973), as estatísticas do teste da razão de verossimilhança são independentes de alguns parâmetros dos modelos especificados nas hipóteses. Por exemplo, para comparação lognormal *vs.* Weibull, a estatística da razão de verossimilhança é independente dos parâmetros  $(\alpha_1, \alpha_2)$  e  $(\beta_1, \beta_2)$ , quando a amostra é gerada da lognormal e da Weibull, respectivamente. Devido essa propriedade da invariância, PEREIRA (1976) mostra que as estatísticas do teste de Cox também são independentes de alguns parâmetros dos modelos. Dessa forma, para gerar as amostras das distribuições, foram utilizados os valores dos parâmetros considerados em (PEREIRA, 1976), conforme consta na Tabela 5.1.

Tabela 5.1: Valores dos parâmetros das distribuições utilizadas para geração das amostras

Testes	Valores dos parâmetros
Lognormal $\times$ Weibull	Lognormal(0, 1) Weibull(1, 1)
Lognormal $\times$ Gama	Lognormal(0, 0.1); (0, 0.25); (0, 0.5); (0, 1); (0, 2) Gama(0.5, 1); (0.8, 1); (1, 1); (2, 1); (5, 1); (10, 1)
Gama $\times$ Weibull	Weibull(0.6, 1); (1.2, 1); (3.6, 1); (5, 1) Gama(0.6, 1); (0.8, 1); (1.2, 1); (2, 1); (5, 1); (10, 1)

O nível crítico  $c$  para aceitação/rejeição da hipótese testada através do FBST foi definido de acordo com o critério assintótico, apresentado na seção 3.5.2, considerando um nível de significância de 5%. A opção por este critério deveu-se ao simples fato de o teste de Cox, utilizado neste trabalho como *benchmark*, ser um procedimento assintótico. Observe que os modelos de mistura considerados nestas simulações contêm três parâmetros, nomeadamente o peso da mistura,  $p$ ; a média,  $\mu$ ; e a variância,  $\sigma^2$ , da população. O valor de  $c$  é o ponto de corte, a partir do qual a hipótese é rejeitada, isto é, se  $\bar{ev}(H) > c$  ou, equivalentemente, se  $ev(H) < 1 - c$ . Uma vez que, nas comparações entre pares de modelos, o modelo de mistura completo (reparametrizado) e o modelo restrito (uma componente) possuem, respectivamente, 3 e 2 graus de liberdade, tem-se, então,  $c = F_3[F_1^{-1}(0.95)] = 0.72$ . Portanto, a hipótese nula é rejeitada se  $\bar{ev} > 0.72$ , ou equivalentemente, se  $ev < 0.28$ .

Para o teste de Cox, considerando o nível de significância  $\gamma = 5\%$ , definiu-se a região de rejeição como  $R = \{y : |T^*| > 1.96\}$ , onde  $T^* \sim N(0, 1)$ . Os estimadores de máxima verossimilhança são calculados via função `fitdistr`, do pacote `MASS`, enquanto que as estimativas dos limites em probabilidade, quando necessárias, são obtidas com recurso ao método de Newton-Raphson, via função `nleqslv` do pacote

com mesmo nome, do *software* R, versão 3.2.2.

## 5.1.2 Resultados das simulações

### i) Testes para distribuições lognormal e Weibull

As evidências em favor das hipóteses  $H_L$  e  $H_W$  são calculadas a partir do modelo de mistura lognormal-Weibull,

$$f(y|p, \mu, \sigma) = pf_L(y|\mu, \sigma) + (1 - p)f_W(y|\mu, \sigma), \quad 0 \leq p \leq 1.$$

A hipótese de que  $y$  tem função densidade de probabilidade  $f_L$  é equivalente à hipótese

$$H_L : p = 1,$$

e a hipótese de que  $y$  provém da distribuição Weibull é equivalente à hipótese

$$H_W : p = 0.$$

As estatísticas do teste de Cox para as distribuições lognormal *vs.* Weibull são apresentadas a seguir. Suponha, primeiramente, que  $H_L$  é a hipótese nula e  $H_W$  a hipótese alternativa. Tem-se, então:

$$\beta_{1\alpha} = \alpha_2^{-1/2}, \quad \beta_{2\alpha} = \exp\{\alpha_1 + \sqrt{\alpha_2}/2\},$$

$$T_{LW} = n \left\{ \hat{\beta}_1 \ln \hat{\beta}_2 - \beta_{1\hat{\alpha}} \ln \beta_{2\hat{\alpha}} - \ln \hat{\beta}_1 + \ln \beta_{1\hat{\alpha}} - \hat{\alpha}_1 (\hat{\beta}_1 - \beta_{1\hat{\alpha}}) \right\},$$

$$V_L(T_{LW}) = 0.2183n.$$

Supondo, agora, que  $H_W$  é a hipótese nula e  $H_L$  a hipótese alternativa, tem-se:

$$\alpha_{1\beta} = \psi(1)/\beta_1 + \ln \beta_2; \quad \alpha_{2\beta} = \psi'(1)/\beta_1^2,$$

$$T_{WL} = n \left\{ \hat{\beta}_1 (\hat{\alpha}_1 - \alpha_{1\hat{\beta}}) + \frac{1}{2} \ln \frac{\hat{\alpha}_2}{\alpha_{2\hat{\beta}}} \right\}$$

$$V_W(T_{WL}) = 0.2834n.$$

Os resultados das simulações dos testes estão resumidos nas Tabelas abaixo. Conforme esperado, ambos os procedimentos apresentam altas taxas de aceitação das hipóteses que especificam as verdadeiras distribuições geradora dos dados (Tabelas 5.2 e 5.4). As taxas de erro tipo I (taxa de rejeição do modelo verdadeiro), obtidas pelo FBST, mantêm-se sempre abaixo do nível de significância estabelecido (5%), para todos os tamanhos das amostras. E o teste de Cox alcança níveis de significância muito próximos de 5%, que corresponde ao padrão normal estabelecido.

Isso é o que se espera numa aplicação específica (PEREIRA, 1978).

Quanto à rejeição das hipóteses que especificam os modelos falsos, observa-se que, à medida que o tamanho da amostra aumenta, as taxas convergem para 1 (Tabelas 5.3 e 5.5), o que mostra a consistência de ambos os testes. As taxas de rejeição fornecidas pelo FBST são, em geral, maiores do que as do teste de Cox, principalmente quando  $n$  tende a ser pequeno. Isso significa que o FBST apresenta maior poder discriminatório em comparação com o teste de Cox. Note que o teste de Cox (baseado na modificação da razão de máxima verossimilhança) é um procedimento formulado justamente para alcançar um alto poder contra a hipótese nula falsa (COX, 1961).

Tabela 5.2: Taxas de aceitação da hipótese nula verdadeira,  $H_L$ . Testes para lognormal *vs.* Weibull usando dados gerados da lognormal(0, 1).

$n$	20	40	60	80	100	150	200
FBST	0.992	0.978	0.984	0.980	0.984	0.982	0.994
Teste de Cox	0.988	0.980	0.968	0.972	0.964	0.974	0.974

Tabela 5.3: Taxas de rejeição da hipótese nula falsa,  $H_W$ . Testes para lognormal *vs.* Weibull usando dados gerados da lognormal(0, 1).

$n$	20	40	60	80	100	150	200
FBST	0.316	0.608	0.704	0.930	0.998	1.000	1.000
Teste de Cox	0.160	0.384	0.670	0.822	0.938	1.000	1.000

Tabela 5.4: Taxas de aceitação da hipótese nula verdadeira,  $H_W$ . Testes para lognormal *vs.* Weibull usando dados gerados da Weibull(1, 1).

$n$	20	40	60	80	100	150	200
FBST	0.998	0.966	0.972	0.982	0.976	0.982	0.984
Teste de Cox	0.994	0.972	0.964	0.978	0.950	0.944	0.936

Tabela 5.5: Taxas de Rejeição da hipótese nula falsa,  $H_L$ . Testes para lognormal *vs.* Weibull usando dados gerados da Weibull(1, 1).

$n$	20	40	60	80	100	150	200
FBST	0.410	0.784	0.860	0.928	0.956	0.990	1.000
Teste de Cox	0.304	0.580	0.774	0.896	0.942	0.994	1.000

## ii) Testes para distribuições lognormal e gama

O modelo de mistura considerado para o cálculo das evidências a favor das hipóteses  $H_L$  e  $H_G$  é representado na forma

$$f(y|p, \mu, \sigma) = pf_L(y|\mu, \sigma) + (1 - p)f_G(y|\mu, \sigma), \quad 0 \leq p \leq 1.$$

Para o cálculo das estatísticas do teste de Cox, suponha, primeiramente, que  $H_L$  seja a hipótese nula e  $H_G$  a hipótese alternativa. Então, tem-se:

$$\begin{aligned} \ln \gamma_{1\alpha} - \psi(\gamma_{1\alpha}) &= \ln \gamma_{2\alpha} - \alpha_1 = \alpha_2/2, \quad \gamma_{2\alpha} = \exp \{ \alpha_1 + \alpha_2/2 \} \\ T_{LG} &= n \{ \ln \Gamma(\hat{\gamma}_1) - \hat{\gamma}_1 \psi(\hat{\gamma}_1) + \hat{\gamma}_1 - \ln \Gamma(\gamma_{1\hat{\alpha}}) + \gamma_{1\hat{\alpha}} \psi(\gamma_{1\hat{\alpha}}) - \gamma_{1\hat{\alpha}} \}, \\ V_L(T_{LG}) &= n \gamma_{1\alpha}^2 \left[ \exp(\alpha_2) - 1 - \alpha_2 - \frac{\alpha_2^2}{2} \right], \end{aligned}$$

onde a solução  $\gamma_{1\hat{\alpha}}$  é única.

Quando  $H_G$  representa a hipótese nula e  $H_L$  a hipótese alternativa, tem-se

$$\begin{aligned} \alpha_{1\gamma} &= \psi(\gamma_1) - \ln(\gamma_1/\gamma_2), \quad \alpha_{2\gamma} = \psi'(\gamma_1), \\ T_{GL} &= \frac{n}{2} \ln(\hat{\alpha}_2/\alpha_2 \hat{\gamma}), \\ V_G(T_{GL}) &= n \left\{ \frac{\psi'''(\gamma_1)}{4\{\psi'(\gamma_1)\}^2} - \frac{\gamma_1 \{\psi''(\gamma_1)\}^2}{4\{\psi'(\gamma_1)\}^2 \{\gamma_1 \psi'(\gamma_1) - 1\}} + \frac{1}{2} \right\}. \end{aligned}$$

Os resultados das Tabelas 5.6 e 5.8 são similares aos encontrados nas simulações anteriores (comparação lognormal *vs.* Weibull). As taxas de aceitação das hipóteses que especificam os modelos verdadeiros, calculadas pelo FBST, mostram que os níveis de significância mantêm-se sempre abaixo do estipulado (5%). E, para o teste de Cox, os níveis de significância obtidos alcançam valores próximos de 5%, conforme o esperado.

Quanto à rejeição das hipóteses que especificam os modelos falsos, observa-se que as taxas calculadas pelo FBST são, em geral, maiores do que as do teste de Cox (Tabelas 5.7 e 5.9). Mais uma vez, isto é uma indicação de que o poder discriminatório do FBST é maior do que o do teste de Cox, na comparação entre essas distribuições. Observa-se também que, para valores pequenos de  $\alpha_2$  e valores grande de  $\gamma_1$ , ambos os testes apresentam taxas de rejeição do modelo falso muito baixas. Isto acontece porque, para esses valores de  $\alpha_2$  e  $\gamma_1$ , as distribuições lognormal e gama são muito parecidas. Para valores de  $\gamma_1$  próximos de 1, os testes apresentam bom poder discriminatório, uma vez que as duas distribuições tem formas bem distintas. Este fato também é constatado em PEREIRA (1976).

Tabela 5.6: Taxas de aceitação da hipótese nula  $H_L$  com base no FBST e no teste de Cox, usando dados gerados da lognormal.

$n$	$\alpha_2 = 0.1$		$\alpha_2 = 0.25$		$\alpha_2 = 0.5$		$\alpha_2 = 1.0$		$\alpha_2 = 2.0$	
	FBST	Cox	FBST	Cox	FBST	Cox	FBST	Cox	FBST	Cox
20	1.000	0.970	1.000	0.976	0.998	0.986	0.998	0.992	0.998	1.000
40	1.000	0.976	1.000	0.974	0.996	0.966	0.980	0.980	0.992	0.996
60	0.996	0.972	0.990	0.974	0.980	0.968	0.992	0.994	0.980	0.988
80	0.982	0.934	0.984	0.956	0.970	0.962	0.972	0.982	0.992	0.988
100	0.986	0.968	0.977	0.970	0.970	0.972	0.984	0.976	0.988	0.986
150	0.976	0.964	0.960	0.958	0.974	0.944	0.986	0.978	0.998	0.978
200	0.970	0.956	0.978	0.956	0.980	0.952	0.994	0.962	0.996	0.972

Tabela 5.7: Taxas de rejeição da hipótese nula  $H_G$  com base no FBST e no teste de Cox, usando dados gerados da lognormal.

$n$	$\alpha_2 = 0.1$		$\alpha_2 = 0.25$		$\alpha_2 = 0.5$		$\alpha_2 = 1.0$		$\alpha_2 = 2.0$	
	FBST	Cox	FBST	Cox	FBST	Cox	FBST	Cox	FBST	Cox
20	0.006	0.034	0.014	0.054	0.040	0.076	0.066	0.142	0.016	0.264
40	0.056	0.082	0.136	0.144	0.262	0.246	0.444	0.400	0.618	0.640
60	0.088	0.120	0.286	0.256	0.458	0.418	0.664	0.640	0.774	0.832
80	0.152	0.166	0.354	0.318	0.590	0.540	0.816	0.820	0.908	0.956
100	0.220	0.214	0.510	0.424	0.714	0.654	0.862	0.882	0.936	0.978
150	0.374	0.328	0.726	0.610	0.860	0.856	0.944	0.978	0.996	1.000
200	0.512	0.410	0.808	0.742	0.920	0.930	0.994	0.998	0.998	1.000

Tabela 5.8: Taxas de aceitação da hipótese nula  $H_G$  com base no FBST e no teste de Cox, usando dados gerados da gama.

$n$	$\gamma_1 = 0.5$		$\gamma_1 = 0.8$		$\gamma_1 = 1.0$		$\gamma_1 = 2.0$		$\gamma_1 = 5.0$		$\gamma_1 = 10.0$	
	FBST	Cox	FBST	Cox								
20	1.000	0.990	1.000	0.994	0.998	0.978	1.000	0.984	1.000	0.974	1.000	0.964
40	0.988	0.972	0.982	0.972	0.984	0.976	0.992	0.972	0.994	0.962	0.996	0.976
60	0.986	0.968	0.984	0.962	0.982	0.970	0.990	0.992	0.982	0.954	1.000	0.970
80	0.980	0.966	0.986	0.964	0.976	0.948	0.980	0.972	0.986	0.958	0.952	0.960
100	0.980	0.970	0.998	0.970	0.980	0.972	0.968	0.968	0.966	0.964	0.982	0.962
150	0.996	0.958	0.978	0.962	0.974	0.958	0.970	0.948	0.972	0.970	0.968	0.940
200	0.988	0.964	0.980	0.962	0.980	0.968	0.970	0.958	0.968	0.958	0.968	0.936

Tabela 5.9: Taxas de rejeição da hipótese nula  $H_L$  com base no FBST e no teste de Cox, usando dados gerados da gama.

$n$	$\gamma_1 = 0.5$		$\gamma_1 = 0.8$		$\gamma_1 = 1.0$		$\gamma_1 = 2.0$		$\gamma_1 = 5.0$		$\gamma_1 = 10.0$	
	FBST	Cox	FBST	Cox								
20	0.116	0.000	0.140	0.002	0.150	0.024	0.088	0.072	0.004	0.058	0.000	0.052
40	0.644	0.002	0.606	0.054	0.520	0.186	0.358	0.244	0.12	0.146	0.018	0.066
60	0.882	0.028	0.794	0.284	0.766	0.452	0.526	0.364	0.241	0.206	0.054	0.106
80	0.942	0.054	0.878	0.612	0.850	0.710	0.680	0.550	0.350	0.312	0.096	0.108
100	0.976	0.256	0.964	0.876	0.914	0.866	0.768	0.682	0.406	0.324	0.132	0.190
150	0.994	0.602	0.986	0.988	0.978	0.982	0.876	0.858	0.612	0.534	0.334	0.318
200	0.998	0.824	0.994	0.998	0.994	1.000	0.958	0.962	0.732	0.680	0.432	0.390

### iii) Testes para distribuições gama e Weibull

O procedimento do FBST é utilizado para calcular as evidências em favor das hipóteses  $H_G$  e  $H_W$ , a partir do modelo de mistura gama-Weibull:

$$f(y|p, \mu, \sigma) = p_G(y|\mu, \sigma) + (1 - p)f_W(y|\mu, \sigma), \quad 0 \leq p \leq 1.$$

Para o cálculo das estatísticas do teste de Cox, suponha, primeiramente, que  $H_G$  seja a hipótese nula e que  $H_W$  seja a hipótese alternativa. Note que  $H_G$  e  $H_W$  são parcialmente não separadas quando  $\beta_1 = \gamma_1 = 1$ , uma vez que especificam a distribuição exponencial. Tem-se

$$\begin{aligned} \ln \beta_{2\gamma} &= \ln \left( \frac{\gamma_2}{\gamma_1} \right) + \beta_{1\gamma}^{-1} \ln \frac{\Gamma(\beta_{1\gamma} + \gamma_1)}{\Gamma(\gamma_1)}, \quad \psi(\beta_{1\hat{\gamma}} + \gamma_1) - \frac{1}{\beta_{1\hat{\gamma}}} = \psi(\gamma_1) \\ T_{GW} &= n \left[ \ln \left( \frac{\beta_{1\hat{\gamma}}}{\hat{\beta}_1} \right) - (\beta_{1\hat{\gamma}} \ln \beta_{2\hat{\gamma}} - \hat{\beta}_1 \ln \hat{\beta}_2) \right. \\ &\quad \left. + \{\beta_{1\hat{\gamma}} - \hat{\beta}_1\} \left\{ \psi(\hat{\gamma}_1) - \ln \left( \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \right) \right\} \right], \\ V_G(T_{GW}) &= n \left[ \frac{\Gamma(2\beta_{1\gamma} + \gamma_1)\Gamma(\gamma_1)}{\{\Gamma(\beta_{1\gamma} + \gamma_1)\}^2} \right. \\ &\quad \left. + \frac{1}{\{\gamma_1\psi'(\gamma_1) - 1\} \beta_{1\gamma}^2} \left\{ 3\beta_{1\gamma}^2 - \gamma_1 - \beta_{1\gamma}^4\psi'(\gamma_1) - \gamma_1\psi'(\gamma_1)\beta_{1\gamma}^2 \right\} \right], \end{aligned}$$

onde as soluções  $\beta_{1\gamma}$  e  $\beta_{2\gamma}$  são únicas.

Suponha, agora, que  $H_W$  seja a hipótese nula e  $H_G$  a hipótese alternativa, tem-se:

$$\ln \gamma_{1\beta} - \psi(\gamma_{1\beta}) = \ln \Gamma \left( 1 + \frac{1}{\beta_1} \right) - \frac{\psi(1)}{\beta_1}, \quad \gamma_{2\beta} = \beta_2 \Gamma \left( 1 + \frac{1}{\beta_1} \right),$$

$$T_{WG} = n \left[ \hat{\gamma}_{1\beta} \left\{ \psi(\gamma_{1\hat{\beta}}) - 1 \right\} - \ln \Gamma(\gamma_{1\hat{\beta}}) - \hat{\beta}_1 \left\{ \psi(\gamma_{1\hat{\beta}}) - \ln \left( \frac{\gamma_{1\hat{\beta}}}{\hat{\gamma}_2} \right) \right\} \right. \\ \left. - \left( \hat{\gamma}_1 \left\{ \psi(\hat{\gamma}_1) - 1 \right\} - \ln \Gamma(\hat{\gamma}_1) - \hat{\beta}_1 \left\{ \psi(\hat{\gamma}_1) - \ln \left( \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \right) \right\} \right) \right],$$

$$V_W(T_{WG}) = n \left[ \left( \frac{\beta_1 - \gamma_{1\beta}}{\beta_1} \right)^2 \psi'(1) + \gamma_{1\beta}^2 \frac{\Gamma \left( 1 + \frac{2}{\beta_1} \right)}{\left\{ \Gamma \left( 1 + \frac{1}{\beta_1} \right) \right\}^2} - \gamma_{1\beta}^2 - 1 \right. \\ \left. + 2 \left( \gamma_{1\beta} - \frac{\gamma_{1\beta}}{\beta_1} \right) \left\{ \psi \left( 1 + \frac{1}{\beta_1} \right) - \psi(1) \right\} \right. \\ \left. - \frac{1}{\psi'(1)} \left\{ 1 - \frac{\gamma_{1\beta}}{\beta_1} \left\{ \psi \left( 1 + \frac{1}{\beta_1} \right) - \psi(1) \right\} \right\}^2 \right].$$

Os resultados das simulações para esta comparação estão apresentados nas Tabelas 5.10 a 5.13. Para taxas de aceitação das hipóteses que especificam os modelos verdadeiros, observa-se que, ao contrário do teste de Cox que, em alguns casos, fornece níveis de significância acima de 5%, o FBST apresenta níveis abaixo deste valor, considerando todos os valores de  $\gamma_1$ , para todos os tamanhos das amostras.

Quanto à rejeição da hipótese nula falsa, como era esperado, para valores de  $\beta_1$  e  $\gamma_1$  próximos de 1, o poder dos dois testes é muito baixo. Isto porque, para esses valores, as distribuições gama e Weibull são muito parecidas. PEREIRA (1976) mostra que, para valores de  $\beta_1$  maiores que 3.6, assim como para valores de  $\gamma_1$  menores que 0.8 e maiores que 1.2, as duas distribuições tendem a ter formas distintas. Este fato pode ser observado nas Tabelas 5.11 e 5.13. Nesta comparação, a convergência das taxas de rejeição do modelo falso, calculadas pelo FBST, é mais lenta em comparação com a do teste de Cox.

Tabela 5.10: Taxas de aceitação da hipótese nula  $H_W$  com base no FBST e no teste de Cox, usando dados gerados da Weibull.

$n$	$\beta_1 = 0.6$		$\beta_1 = 0.8$		$\beta_1 = 1.2$		$\beta_1 = 3.6$		$\beta_1 = 5.0$	
	FBST	Cox								
50	0.998	0.950	1.000	0.930	0.998	0.940	0.984	0.970	0.982	0.970
100	0.986	0.980	0.998	0.970	1.000	0.920	0.980	0.950	0.986	0.950
200	0.976	0.970	0.990	0.910	0.998	0.880	0.988	0.930	1.000	0.930

Tabela 5.11: Taxas de rejeição da hipótese nula  $H_G$  com base no FBST e no teste de Cox, usando dados gerados da Weibull.

$n$	$\beta_1 = 0.6$		$\beta_1 = 0.8$		$\beta_1 = 1.2$		$\beta_1 = 3.6$		$\beta_1 = 5.0$	
	FBST	Cox								
50	0.118	0.340	0.000	0.120	0.002	0.130	0.188	0.580	0.308	0.670
100	0.270	0.460	0.006	0.170	0.004	0.130	0.222	0.800	0.436	0.890
200	0.646	0.730	0.054	0.340	0.012	0.180	0.224	0.990	0.560	1.000

Tabela 5.12: Taxas de aceitação da hipótese nula  $H_G$  com base no FBST e no teste de Cox, usando dados gerados da gama.

$n$	$\gamma_1 = 0.6$		$\gamma_1 = 0.8$		$\gamma_1 = 1.2$		$\gamma_1 = 2.0$		$\gamma_1 = 5.0$		$\gamma_1 = 10.0$	
	FBST	Cox	FBST	Cox								
50	1.000	0.920	1.000	0.890	1.000	0.920	0.994	0.960	0.976	0.940	0.944	0.950
100	0.998	0.970	0.998	0.940	1.000	0.930	0.986	0.930	0.9824	0.930	0.938	0.980
200	0.988	0.950	1.000	0.940	1.000	0.860	0.988	0.940	0.950	0.930	0.966	0.920

Tabela 5.13: Taxas de rejeição da hipótese nula  $H_W$  com base no FBST e no teste de Cox, usando dados gerados da Weibull.

$n$	$\gamma_1 = 0.6$		$\gamma_1 = 0.8$		$\gamma_1 = 1.2$		$\gamma_1 = 2.0$		$\gamma_1 = 5.0$		$\gamma_1 = 10.0$	
	FBST	Cox	FBST	Cox								
50	0.010	0.180	0.000	0.170	0.00	0.100	0.016	0.200	0.220	0.320	0.386	0.490
100	0.092	0.260	0.000	0.110	0.00	0.120	0.108	0.290	0.576	0.600	0.660	0.820
200	0.284	0.410	0.160	0.220	0.04	0.140	0.336	0.520	0.846	0.930	0.846	0.980

## 5.2 Exemplos de aplicações

Nesta seção, serão analisados três conjuntos de dados não censurados para ilustrar os procedimentos do FBST e do teste de Cox na discriminação entre os pares de distribuições de famílias separadas. O objetivo é decidir qual, dentre os modelos estatísticos alternativos, se ajusta melhor a esses conjuntos de dados. Nesses exemplos, o FBST também é utilizado para a escolha de distribuições, considerando o modelo de mistura com três componentes.

**Exemplo 1** QUESENBERRY e KENT (1982) apresentam dados referentes a tensão de fios de poliéster/viscose para estudar o problema de ruptura durante a tecelagem. O experimento consiste de 100 amostras de fios submetidos a 80 ciclos por minutos ao nível de 2.3% de tensão. O percentual de tensão é medido com respeito a resistência à ruptura do fio. Os dados são os ciclos que os fios suportam até a ocorrência da falha. Para discriminar entre as distribuições exponencial, gama, Weibull e lognormal, eles utilizam um procedimento de seleção que exige um conjunto de dados não censurados, e que consiste, essencialmente, em comparar estatísticas que são invariantes sob transformação na escala dos dados e selecionam aquela com mais alto valor.

Na procura de um modelo adequado para os dados de ciclos até a falha do fio de

poliéster, comparações binárias entre as distribuições lognormal, gama e Weibull, são efetuadas primeiro. A Tabela 5.14 apresenta as medidas de evidência a favor de cada modelo, calculadas através do FBST e do teste de Cox.

Tabela 5.14: Evidências fornecidas pelos dados de ciclos até a falha do fio de poliéster

Comparação	Hipótese nula	Evidência a favor da hipótese nula		
		e-valor (FBST)	Variável normal padrão, $T_{fg}^*$	p-valor (Teste de Cox)
$H_L \times H_W$	$H_L$	0.000	-3.048	0.002
	$H_W$	0.871	-0.549	0.583
$H_L \times H_G$	$H_L$	0.000	-3.033	0.002
	$H_G$	0.997	0.773	0.439
$H_G \times H_W$	$H_G$	0.697	-1.016	0.309
	$H_W$	0.725	0.967	0.333

Os testes para comparação entre as distribuições lognormal e Weibull forneceram os seguintes resultados: o valor da evidência Bayesiana a favor da hipótese  $H_L$  é igual a 0.000 e o da variável normal padrão para a estatística do teste de Cox é  $T_{LW}^* = -3.048$ , que corresponde ao p-valor de 0.002. Isto implica rejeitar o modelo lognormal. Por outro lado, a evidência em favor de  $H_W$  é igual a 0.871 e  $T_{WL}^* = -0.549$  (p-valor= 0.583), sugerindo a não rejeição da hipótese, ou seja, o modelo Weibull se ajusta melhor aos dados em relação ao modelo lognormal. ARAUJO e PEREIRA (2007) calcularam os fatores de Bayes intrínseco e fracionário para comparar as mesmas distribuições usando este conjunto de dados. Eles também encontraram uma decisiva evidência contra a distribuição lognormal.

Para comparação lognormal *vs.* gama, os testes mostram que o modelo gama se ajusta melhor aos dados em relação ao modelo lognormal, o que, novamente, confere com o encontrado por QUESENBERRY e KENT (1982).

Uma vez que o modelo Weibull já foi preferido em detrimento do modelo lognormal, comparou-se, então, a distribuição gama com a Weibull. Os testes para esta comparação forneceram os seguintes resultados: o valor da evidência a favor da hipótese  $H_G$  é igual a 0.697 e  $T_{GW}^* = -1.016$  (p-valor= 0.309). Para a hipótese  $H_W$ , a evidência obtida foi de 0.725 e  $T_{WG}^* = 0.967$  (p-valor= 0.333). Isto indica que nenhuma das hipóteses foi rejeitada e que qualquer um desses modelos pode ser utilizado para a modelagem dos dados, embora a evidência a favor do modelo Weibull seja ligeiramente maior do que a evidência a favor do modelo lognormal. Pelos fatores de Bayes intrínseco e fracionário, ARAUJO e PEREIRA (2007), usando este conjunto de dados, constataram que as distribuições gama e Weibull são muito similares. QUESENBERRY e KENT (1982) também não conseguiram discriminar

estes modelos, pois os valores de suas estatísticas de seleção são muito próximos.

### Aplicando o modelo de mistura lognormal-gama-Weibull (LGW)

Feitas as comparações entre os pares de distribuições, a seguir, o modelo de mistura com três componentes,

$$f(y|p, \mu, \sigma) = p_1 f_L(y|\mu, \sigma) + p_2 f_G(y|\mu, \sigma) + p_3 f_W(y|\mu, \sigma), \quad (5.1)$$

é aplicado aos dados analisados neste exemplo, com o objetivo de identificar a distribuição com melhor ajuste.

As estimativas dos parâmetros do modelo (5.1) são apresentadas na Tabela 5.15. Aqui, D.P., 2.5% e 97.5% denotam o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente. Na Tabela 5.16 constam as evidências a favor de cada modelo, calculadas através do FBST, com base nos testes de hipóteses sobre os pesos da mistura. O p-valor com asterisco foi calculado segundo DINIZ *et al.* (2012), com base na propriedade de normalidade assintótica, descrita na seção 3.5.2. As evidências confirmam que o modelo lognormal pode ser desconsiderado entre as distribuições a serem usadas para a modelagem destes dados, uma vez que a hipótese  $H : p_1 = 0$  não é rejeitada.

Observe que a preferência pelo modelo Weibull é mais evidente quando se analisa o modelo de mistura com três componentes (Tabela 5.16) do que na comparação gama *vs.* Weibull (modelo com dois componentes), onde a dificuldade em distinguir entre os modelos é maior, conforme visto anteriormente. Esta pode ser uma vantagem do teste FBST formulado para discriminar hipóteses separadas, considerando os modelos sob comparação como componentes do modelo de mistura. Ou seja, a possibilidade de comparar simultaneamente todos os modelos candidatos, evitando-se os problemas que surgem quando comparações entre pares de modelos são efetuadas (SAWYER, 1984).

Tabela 5.15: Sumário da distribuição a posteriori dos parâmetros do modelo LGW, aplicado aos dados de ciclos até a falha do fio de poliéster

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$p_1$ -lognormal	0.170	0.127	0.007	0.143	0.469
$p_2$ -gama	0.381	0.249	0.018	0.355	0.869
$p_3$ -Weibull	0.449	0.237	0.032	0.461	0.869
$\mu$	220.423	14.239	193.679	219.966	249.759
$\sigma^2$	20665.944	4001.274	14369.620	20248.042	30166.612

Na Figura 5.1 estão representadas as curvas de sobrevivência estimadas pela mistura LGW (estimativas da Tabela 5.15), pelo modelo Weibull (estimativas da Tabela 5.17) e pelo estimador PEXE (do inglês, *Piecewise Exponential Estimator*), discutido em KIM e PROSCHAN (1991), que representa os dados observados. O

Tabela 5.16: Testes de hipóteses para os pesos da mistura LGW usando dados de ciclos até a falha do fio de poliéster

Hipótese	e-valor	p-valor*
$p_1 = 0$	0.652	0.116
$p_2 = 0$	0.206	0.015
$p_3 = 0$	0.073	0.003

\*p-valor calculado segundo DINIZ *et al.* (2012)

PEXE suaviza o estimador de Kaplan-Meier utilizando a função acumulada da distribuição Exponencial, ajustada entre os intervalos dos tempos de falha. Parece que a função de sobrevivência estimada pelo modelo Weibull se ajusta melhor aos dados.

Tabela 5.17: Sumário da distribuição a posteriori dos parâmetros do modelo Weibull, aplicado aos dados de ciclos até a falha do fio de poliéster

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$\mu$	220.409	13.675	194.595	219.938	248.975
$\sigma^2$	19862.278	3170.874	14708.324	19523.516	27038.00

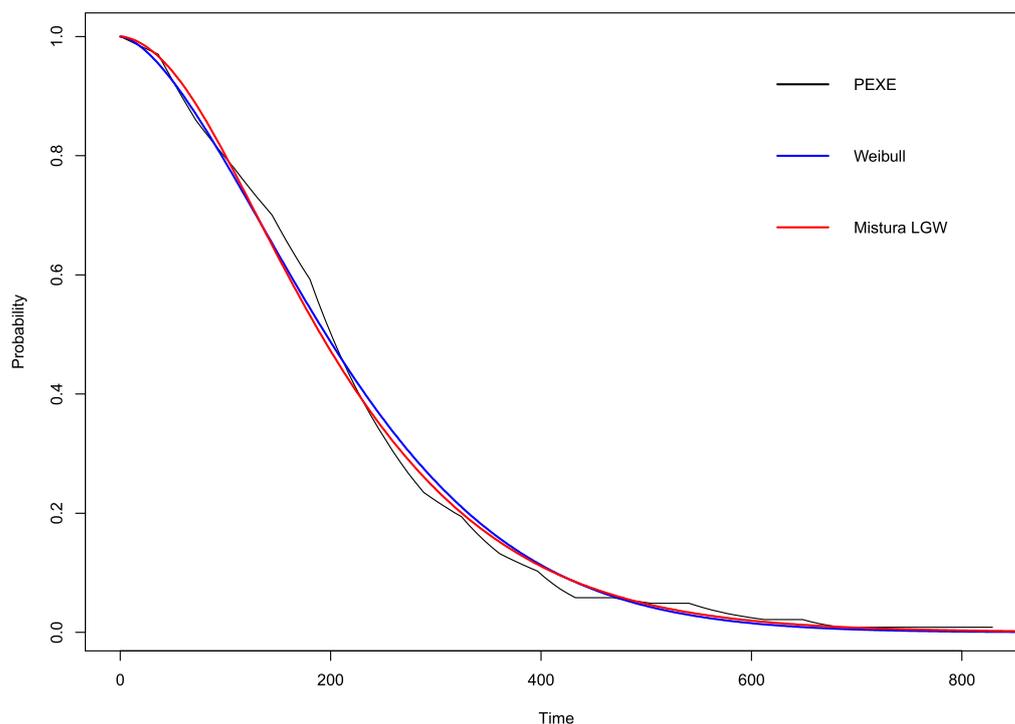


Figura 5.1: Curvas de sobrevivência estimadas pelo estimador PEKE e pelos modelos Weibull e de mistura LGW para dados de ciclos até a falha do fio de poliéster

**Exemplo 2** LINDSEY (1974), considerando respostas observadas como medidas discretas, desenvolveu um método de comparação de distribuições de probabilidades usando a distribuição multinomial como base e função de verossimilhança para fazer inferências. Intervalos apropriados são determinados por um método gráfico. Com este método, testou a distribuição do tempo de sobrevivência, em minutos, de grupos de larvas do bicho-da-seda, quando sujeitos a dose de 0.10 mg de arsenato de sódio/g de peso corporal. Neste trabalho, pretende-se escolher um modelo conveniente, entre as distribuições lognormal, gama e Weibull, para os dados de larvas do grupo A, com  $n = 70$ .

A Tabela 5.18 apresenta as medidas de evidência a favor de cada modelo, calculadas através do FBST e do teste de Cox, nas comparações efetuadas entre pares de distribuições candidatas.

Tabela 5.18: Evidências fornecidas pelos dados de larvas do grupo A

Comparação	Hipótese nula	Evidência a favor da hipótese nula		
		e-valor (FBST)	Variável normal padrão, $T_{fg}^*$	p-valor (Teste de Cox)
$H_L \times H_W$	$H_L$	0.839	0.454	0.649
	$H_W$	0.000	-3.096	0.002
$H_L \times H_G$	$H_L$	0.986	0.607	0.544
	$H_G$	0.907	-0.928	0.353

Para comparação entre as distribuições lognormal e Weibull, o valor da evidência Bayesiana em favor de  $H_L$  é igual a 0.839 e o da variável normal padrão para a estatística do teste de Cox é  $T_{LW}^* = 0.454$ , que corresponde ao p-valor de 0.649. Por outro lado,  $ev(H_W) = 0.000$  e  $T_{WL}^* = -3.096$  (p-valor= 0.002). Isto é uma indicação de que o modelo lognormal se ajusta melhor aos dados em relação ao modelo Weibull.

Comparando a distribuição gama com a lognormal, os testes forneceram evidências  $ev(H_L) = 0.986$  e  $ev(H_G) = 0.907$  e estatísticas  $T_{LG}^* = 0.607$  (p-valor= 0.544) e  $T_{GL}^* = -0.928$  (p-valor= 0.353). Estes resultados indicam que ambas as distribuições fornecem um bom ajuste para os dados, embora a evidência a favor do modelo lognormal seja maior em relação ao modelo gama. Novamente, esta constatação confere com os resultados de ARAUJO e PEREIRA (2007), que utilizaram os fatores de Bayes parcial, intrínseco e fracionário para comparar as distribuições.

A Figura 5.2 apresenta as curvas de sobrevivência estimadas pelo estimador PEXE e pelos modelos lognormal e gama cujas estimativas constam nas Tabelas 5.19 e 5.20, respectivamente. Visualmente, parece que os dois modelos produzem ajustes similares para esses dados.

Tabela 5.19: Sumário da distribuição a posteriori dos parâmetros do modelo lognormal, aplicado aos dados de larvas do grupo A.

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$\mu$	269.497	3.482	262.879	269.447	276.551
$\sigma^2$	839.160	151.715	591.763	269.446	1188.593

Tabela 5.20: Sumário da distribuição a posteriori dos parâmetros do modelo gama, aplicado aos dados de larvas do grupo A

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$\mu$	269.487	3.494	262.642	269.533	276.364
$\sigma^2$	835.456	145.769	596.712	822.937	1161.577

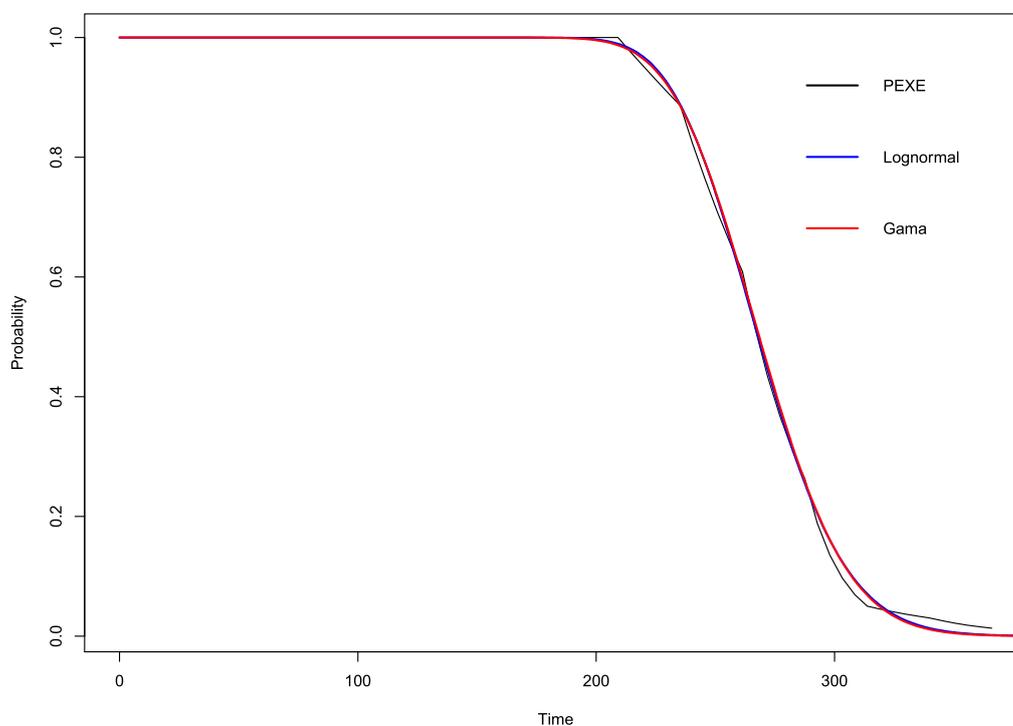


Figura 5.2: Curvas de sobrevivência estimadas por PEXE e pelos modelos lognormal e gama para dados de larvas do grupo A

**Exemplo 3** Período de latência do HIV. Dados relativos a infecção pelo vírus de AIDS através de transfusão sanguínea. Estes dados se referem ao período de latência do HIV para 258 adultos e 37 crianças (menores que 5 anos), infectados via transfusão sanguínea, até 30 de junho de 1986, fornecidos pelo US Center for Disease Control e estão disponíveis em LAGAKOS *et al.* (1988).

Aqui, a análise é restrita aos dados da população adulta ( $n = 258$ ). Na procura de um modelo conveniente para esses dados, os modelos lognormal, gama e Weibull

Tabela 5.21: Evidências fornecidas pelos dados de latência do HIV em adultos

Comparação	Hipótese nula	Evidência a favor da hipótese nula		
		e-valor (FBST)	Variável normal padrão, $T_{fg}^*$	p-valor (Teste de Cox)
$H_L \times H_W$	$H_L$	0.000	-5.349	0.000
	$H_W$	1.000	-0.484	0.628
$H_L \times H_G$	$H_L$	0.000	-5.142	0.000
	$H_G$	1.000	1.864	0.062
$H_G \times H_W$	$H_G$	0.090	-2.492	0.012
	$H_W$	1.000	-0.299	0.765

são testados. A Tabela 5.21 apresenta os resultados dos testes para comparação entre os pares de modelos.

Para discriminação entre as distribuições lognormal e Weibull, o valor da evidência em favor de  $H_L$  é 0.000 e  $T_{LW}^* = -5.349$  (p-valor=0.000). Por outro lado, a evidência obtida a favor da hipótese  $H_W$  foi de 1.000 e  $T_{WL}^* = -0.484$  (p-valor=0.628), sugerindo a não rejeição da hipótese, ou seja, o modelo Weibull se ajusta melhor aos dados em relação ao modelo lognormal.

Na comparação entre as distribuições lognormal e gama, os resultados apontam, novamente, que o modelo lognormal é o menos plausível e, portanto, pode ser descartado. Comparando a distribuição Weibull com a gama, obtiveram-se os seguintes resultados: para a hipótese  $H_G$ , o valor da evidência foi de 0.090 e  $T_{GW}^* = -2.492$  (p-valor=0.012). Para a hipótese  $H_W$ ,  $ev(H_W) = 1.000$  e  $T_{WG}^* = -0.299$  (p-valor=0.765), indicando a não rejeição da hipótese. Portanto, o modelo Weibull é o que melhor se ajusta aos dados. O resultado do fator de Bayes intrínseco calculado em ARAUJO e PEREIRA (2007), na comparação gamma *vs.* Weibull, também indica uma substancial evidência contra o modelo gama, sugerindo, assim, o modelo Weibull como o mais adequado.

### Aplicando o modelo de mistura LGW

As estimativas dos parâmetros do modelo de mistura LGW aplicado aos dados de adultos são apresentadas na Tabela 5.22. As evidências calculadas a favor das hipóteses sobre os pesos da mistura, através do FBST, indicam que tanto a distribuição gama quanto a lognormal devem ser descartadas para a modelagem dos dados de latência do vírus HIV em adultos, uma vez que as hipóteses  $H : p_1 = 0$  e  $H : p_2 = 0$  não são rejeitadas (Tabela 5.23).

Na Figura 5.3 estão representadas as curvas de sobrevivência estimadas pela mistura LGW (estimativas da Tabela 5.22), pelos modelo Weibull (estimativas da Tabela 5.24) e pelo estimador PEXE. Parece que as estimativas da função de sobrevivência do modelo Weibull se ajustam melhor aos dados.

Tabela 5.22: Sumário da distribuição a posteriori dos parâmetros do modelo LGW, aplicado aos dados de latência do HIV em adultos

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$p_1$ -lognormal	0.126	0.098	0.005	0.104	0.360
$p_2$ -gama	0.256	0.201	0.009	0.206	0.745
$p_3$ -Weibull	0.618	0.201	0.151	0.647	0.929
$\mu$	2.712	0.091	2.537	2.709	2.895
$\sigma^2$	2.298	0.260	1.861	2.277	2.875

Tabela 5.23: Testes de hipóteses para os pesos da mistura LGW aplicado aos dados de latência do HIV em adultos

Hipótese	e-valor	p-valor*
$p_1 = 0$	0.834	0.227
$p_2 = 0$	0.856	0.249
$p_3 = 0$	0.078	0.004

\*p-valor calculado segundo DINIZ *et al.* (2012)

Tabela 5.24: Sumário da distribuição a posteriori dos parâmetros do modelo Weibull, aplicado aos dados de latência do HIV em adultos

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$\mu$	2.713	0.093	2.538	2.710	2.901
$\sigma^2$	2.212	0.215	1.840	2.195	2.681

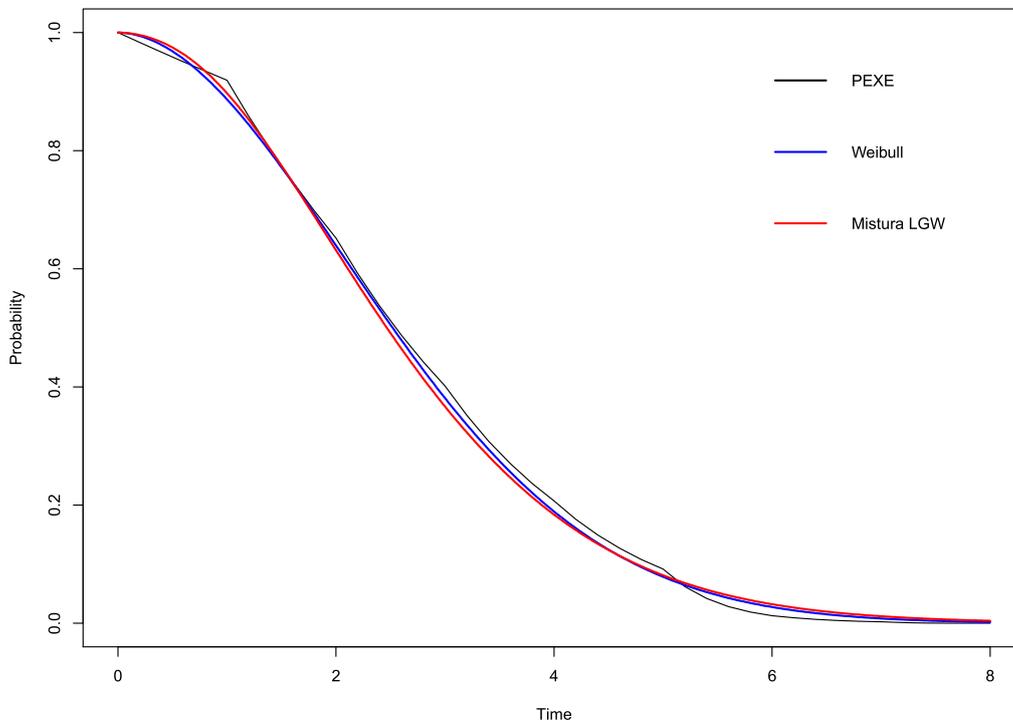


Figura 5.3: Curvas de sobrevivência estimadas por PEKE e pelos modelos Weibull e de mistura LGW para dados de latência do HIV em adultos

# Capítulo 6

## Análise de modelo de mistura finita com dados censurados

### 6.1 Introdução

Uma característica comum nos dados de sobrevivência é a presença de observações censuradas. Uma observação é censurada quando o instante exato da ocorrência da falha de um indivíduo não é conhecido, mas sabe-se que a falha ocorreu dentro de um intervalo ou após um certo instante. Em estudos médicos e áreas afins, censuras ditas à direita são as mais comuns. Discussão mais aprofundada sobre os tipos de censura pode ser encontrada em KLEIN e MOESCHBERGER (2003).

Neste capítulo será analisada a mistura de modelos separados para dados de sobrevivência sujeitos à censura. O objetivo é estimar os parâmetros de mistura e utilizar o procedimento do FBST para comparar os modelos.

A função de verossimilhança 4.6 para esses dados é definida incorporando componentes que representam cada tipo de censura considerada (KLEIN e MOESCHBERGER, 2003). No caso de dados censurados à direita, que é o foco deste trabalho, será considerada a seguinte notação:

- $T_j$  - tempo de sobrevivência (falha);
- $C_j$  - tempo de censura à direita;
- $y_j = \min\{T_j, C_j\}$  - tempo observado
- $\delta_j = \mathbb{I}(T_j \leq C_j)$  - variável indicadora de falha que assume valor 1 se o tempo de sobrevivência foi observado ou valor 0 se tempo foi censurado.

Assumindo mecanismo de censura não-informativa e independência entre os tempos de sobrevivência e de censura (GIJBELS, 2010; KLEIN e MOESCHBERGER, 2003),

a função de verossimilhança, condicionada aos dados observados, é definida como

$$\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{y}) &= \prod_{j=1}^n f(y_j, \delta_j|\boldsymbol{\theta}) \\
&\propto \prod_{j=1}^n [f(y_j|\boldsymbol{\theta})]^{\delta_j} [S(y_j|\boldsymbol{\theta})]^{1-\delta_j} \\
&\propto \prod_{j=1}^n \left[ \sum_{k=1}^m p_k f_k(y_j|\mu, \sigma) \right]^{\delta_j} \left[ \sum_{k=1}^m p_k S_k(y_j|\mu, \sigma) \right]^{1-\delta_j}, \tag{6.1}
\end{aligned}$$

onde  $S_k$  representa a função de sobrevivência da distribuição  $F_k$  que compõe a mistura.

Aqui, a incerteza a priori para os parâmetros  $\mu$  e  $\sigma^2$  também foi modelada pela distribuição gama de média 1 e variância 100, ou seja,  $\mu, \sigma^2 \sim \text{gamma}(0.01, 100)$ . Em relação aos pesos da mistura, para modelos com dois componentes ( $m = 2$ ), considerou-se a priori  $p \sim \text{Beta}(1, 1)$ ; e para  $m = 3$ ,  $\mathbf{p} \sim \text{Dirichlet}(1, 1, 1)$  (PEREIRA e PEREIRA, 2017).

## 6.2 Simulações

O estudo de simulação apresentado nesta seção tem por objetivo aferir o desempenho do FBST na discriminação entre os modelos separados a partir da mistura lognormal-gama-Weibull (LGW), diante de observações censuradas à direita.

### 6.2.1 Esquema para simulação de pontos amostrais

Sejam  $H_L$ ,  $H_G$  e  $H_W$  as hipóteses apresentadas na seção 5.1.1. A partir de cada uma das distribuições especificadas nas hipóteses, foram simuladas 200 amostras de tamanhos  $n = 100, 200, 300$  e  $500$ , de tal forma que uma parte (predefinida) das observações de cada amostra fosse censurada à direita.

Por exemplo, para simular uma amostra,  $\mathbf{y}$ , de tamanho  $n$ , a partir da distribuição lognormal, com parte das observações censurada à direita, observam-se os seguintes passos:

- Definidos os valores dos parâmetros  $\mu$  e  $\sigma^2$ , calculam-se os parâmetros  $(\alpha_1, \alpha_2)$  da lognormal usando as expressões (4.17);
- Para  $j = 1, \dots, n$ 
  - Gera-se o tempo de sobrevivência,  $T_j \sim \text{Lognormal}(\alpha_1, \alpha_2)$ ;
  - Gera-se o tempo de censura à direita a partir de uma distribuição exponencial, i.e,  $C_j \sim \text{Exp}(\lambda)$ , com o valor do parâmetro escolhido de tal modo que uma parte das observações da amostra simulada seja censurada.

- Obtém-se o tempo observado,  $y_j = \min\{T_j, C_j\}$
- Cria-se uma variável indicadora de falha,  $\delta_j = \mathbb{I}(T_j \leq C_j)$

Com a amostra gerada, estimam-se os parâmetros do modelo de mistura LGW, considerando a priori  $\mu, \sigma^2 \sim \text{gamma}(0.01, 100)$  e  $\mathbf{p} \sim \text{Dirichlet}(1, 1, 1)$ . Utilizando o algoritmo *Adaptive Metropolis* (HAARIO *et al.*, 2001), simula-se, para cada parâmetro, uma cadeia com 80 mil valores da distribuição a posteriori. Após serem descartados cinco mil valores iniciais (*burn-in*) e tomando um salto de tamanho cinquenta, obtém-se uma amostra final de 1500 observações, que são utilizadas para calcular as estimativas e a medida de evidência do FBST.

O valor do parâmetro da distribuição dos tempos de censura,  $\lambda$ , é calculado através de um método numérico (WAN, 2017). Formalmente, define-se a probabilidade de censura à direita,  $p_c$ , para a amostra a ser simulada; considera-se que  $C$  tem distribuição exponencial com função de densidade  $g(c|\lambda)$ ; e assume-se independência entre as distribuições dos tempos  $T$  e  $C$ . Para determinar o valor de  $\lambda$  de tal modo que, aproximadamente,  $p_c\%$  das observações simuladas sejam censuradas, resolve-se a seguinte equação:

$$\begin{aligned}
p_c &= \Pr(\delta = 0 | \lambda, \mu, \sigma^2) \\
&= \Pr(C \leq T \leq \infty, 0 \leq C \leq \infty) \\
&= 1 - \Pr(0 \leq T \leq C, 0 \leq C \leq \infty) \\
&= 1 - \int_0^\infty g(c|\lambda) \int_0^c f_L(t|\mu, \sigma) dt dc \\
&= 1 - \int_0^\infty g(c|\lambda) F_L(c|\mu, \sigma) dc,
\end{aligned} \tag{6.2}$$

sendo  $f_L$  a função densidade da distribuição lognormal utilizada para a geração dos tempos de falha e  $F_L$  a respectiva função de distribuição. A equação (6.2) foi resolvida utilizando a função `uniroot`, que pertence ao pacote `stats` do software R.

Para a geração de amostras a partir das distribuições gama e Weibull, um procedimento análogo ao utilizado para o caso da lognormal é empregado.

## 6.2.2 Avaliação do desempenho do FBST

Para avaliar o desempenho do FBST na escolha da verdadeira distribuição utilizada para a geração dos tempos  $T_j$ , adotou-se o critério baseado nos valores das evidências calculadas em favor das hipóteses  $H : p_k = 0$  e  $H : p_k = 1$ ,  $k = L, G, W$ , onde os parâmetros  $p_k$  são os pesos da mistura relacionados a componentes lognormal, gama e Weibull, respectivamente, no modelo LGW.

Por exemplo, para tempos de falha gerados a partir da distribuição lognormal, considera-se que o FBST fez uma escolha correta no modelo LGW, quando a

evidência a favor de  $H : p_L = 0$  é menor que as evidências em favor de  $H : p_G = 0$  e  $H : p_W = 0$ , e a evidência a favor de  $H : p_L = 1$  é maior que as evidências em favor de  $H : p_G = 1$  e  $H : p_W = 1$ . Para tempos de falha gerados a partir das distribuições gama e Weibull, o critério de avaliação do teste é análogo.

As taxas de acerto do FBST no modelo LGW são calculadas com base nas duzentas réplicas. A precisão dos estimadores para os parâmetros do modelo é avaliada através das médias das estimativas das duzentas réplicas.

Nestas simulações, considerou-se  $\mu = 20$  e  $\sigma^2 = 50$ . O procedimento do FBST foi avaliado considerando amostras com diferentes porcentagens de censura: 0%, 10%, 30% e 50%.

### 6.2.3 Resultados das simulações

A Tabela 6.1 apresenta as médias das estimativas dos parâmetros do modelo LGW ajustado aos dados censurados à direita e as taxas de acerto do FBST na escolha da verdadeira distribuição geradora dos dados. Observa-se que, independentemente da distribuição utilizada para gerar os dados e do tamanho da amostra considerado, as estimativas para a média  $\mu$  são muito próximas entre si e do verdadeiro valor do parâmetro. Em relação às estimativas da variância  $\sigma^2$ , constata-se uma variação entre as estimativas, mas que, no geral, se aproximam do verdadeiro valor do parâmetro à medida que o tamanho da amostra cresce.

Observe que o FBST apresenta um alto desempenho na identificação da distribuição Weibull como verdadeira geradora dos dados e baixo desempenho na identificação da distribuição gama. Isto acontece porque, conforme os resultados discutidos na seção 5.1.2, com os parâmetros escolhidos para estas simulações,  $\mu = 20$  e  $\sigma^2 = 50$  ( $\alpha_1 = 2.937$  e  $\alpha_2 = 0.118$ ;  $\gamma_1 = 8.000$  e  $\gamma_2 = 2.500$ ;  $\beta_1 = 3.094$  e  $\beta_2 = 22.366$ ), as distribuições gama e lognormal são parecidas, já que  $\alpha_2$  é pequeno e  $\gamma_1$  razoavelmente grande. Este fato também pode ser visualizado na Figura 6.1, que mostra as curvas de densidades de probabilidade das distribuições para esses valores dos parâmetros. Para amostras com porcentagem de censura igual a 50%, a taxa de acerto, quando a distribuição geradora dos dados é Weibull, atinge pelo menos 95% para  $n \geq 300$ .

Tabela 6.1: Estimativas do modelo LGW e taxas de acerto do FBST na escolha da distribuição geradora dos dados, para amostras com diferentes porcentagens de censura

Censura (%)	Modelo	$n$	$\mu$	$\sigma^2$	$p_L$	$p_G$	$p_W$	Acerto (%)
			20	50	-	-	-	
0	Lognormal	100	19.98	48.87	0.49	0.34	0.17	73
		200	19.98	50.38	0.60	0.30	0.10	91
		300	19.89	46.94	0.63	0.29	0.08	93
		400	19.97	48.05	0.72	0.22	0.06	97
	Gama	100	19.98	51.64	0.36	0.37	0.27	44
		200	20.01	50.49	0.35	0.42	0.23	51
		300	20.00	51.29	0.42	0.42	0.16	54
		500	19.97	50.31	0.36	0.47	0.17	63
	Weibull	100	20.11	51.32	0.18	0.25	0.57	93
		200	20.04	50.65	0.13	0.21	0.66	95
		300	20.03	49.12	0.14	0.20	0.66	100
		500	19.99	49.80	0.08	0.13	0.79	100
10	Lognormal	100	19.92	48.82	0.48	0.36	0.16	84
		200	20.02	48.87	0.59	0.30	0.11	84
		300	19.96	48.43	0.63	0.28	0.09	84
		500	19.97	48.14	0.69	0.25	0.07	93
	Gama	100	20.07	52.95	0.38	0.36	0.26	45
		200	20.01	50.60	0.38	0.41	0.21	53
		300	20.06	50.90	0.36	0.44	0.20	57
		500	20.05	50.91	0.34	0.48	0.18	69
	Weibull	100	20.17	52.06	0.19	0.26	0.55	86
		200	20.04	51.27	0.15	0.22	0.63	94
		300	19.76	50.24	0.12	0.18	0.70	100
		500	19.97	51.26	0.09	0.13	0.78	100
30	Lognormal	100	20.03	49.20	0.47	0.35	0.19	48
		200	20.01	48.71	0.55	0.33	0.13	63
		300	20.02	48.20	0.59	0.31	0.10	70
		500	19.97	47.06	0.64	0.28	0.08	86
	Gama	100	20.13	53.03	0.38	0.36	0.26	35
		200	19.96	50.67	0.40	0.38	0.22	47
		300	20.12	55.09	0.41	0.41	0.18	51
		500	20.00	50.77	0.35	0.47	0.18	70
	Weibull	100	20.07	54.25	0.21	0.28	0.51	81
		200	20.11	52.35	0.16	0.23	0.61	92
		300	19.97	50.77	0.15	0.21	0.64	100
		500	19.97	50.78	0.10	0.15	0.75	100
50	Lognormal	100	19.98	48.14	0.45	0.35	0.20	40
		200	19.91	45.79	0.50	0.35	0.15	50
		300	19.97	47.14	0.56	0.33	0.11	65
		500	19.92	46.82	0.63	0.29	0.08	78
	Gama	100	20.06	53.77	0.36	0.36	0.28	31
		200	19.99	51.17	0.37	0.38	0.25	43
		300	20.10	52.55	0.37	0.41	0.22	47
		500	20.08	51.65	0.40	0.43	0.17	57
	Weibull	100	20.27	58.79	0.24	0.30	0.46	74
		200	20.09	54.65	0.18	0.26	0.56	90
		300	20.05	53.08	0.15	0.23	0.62	95
		500	19.92	51.21	0.13	0.19	0.68	95

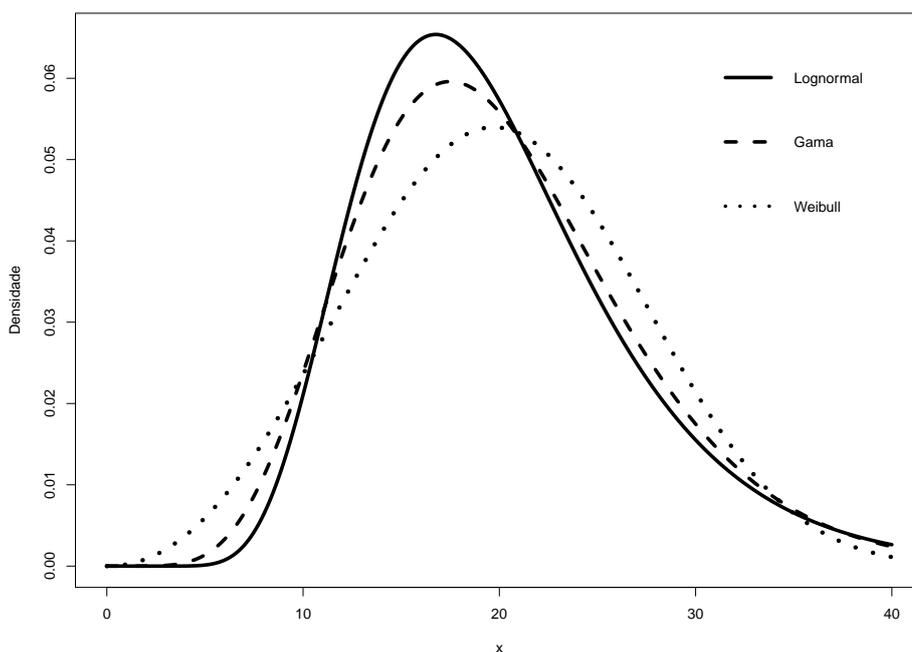


Figura 6.1: Densidades das distribuições Gama, Weibull e Lognormal para  $\mu = 20$  e  $\sigma^2 = 50$

## 6.3 Aplicação: Escolha de um modelo de sobrevivência para pacientes com insuficiência renal crônica terminal

### 6.3.1 Base de dados

A base de dados utilizada neste trabalho refere-se a uma coorte de 473 pacientes com insuficiência renal crônica terminal submetidos à hemodiálise em quatro centros de hemodiálise do Estado do Rio de Janeiro, acompanhados a partir de julho de 1997 a dezembro de 2008. O tempo de seguimento de cada paciente foi o número de meses a partir da data de início do tratamento hemodialítico até a ocorrência de óbito ou fim da observação (transplante renal ou término do estudo). A descrição completa dos dados encontra-se em ALVES *et al.* (2014).

Aqui, o interesse é aplicar o modelo de mistura LGW a esta base de dados e, através do procedimento do FBST, examinar os parâmetros de mistura para escolher o modelo com melhor ajuste aos dados. Mas, antes disso, faz-se a comparação entre todos os pares de modelos, ajustando as misturas lognormal-Weibull, lognormal-gama e gama-Weibull. As variáveis de interesse consideradas para análise são: tempo de sobrevivência em hemodiálise e indicador de desfecho (1 = ocorrência de óbito, 0 = censura à direita). Esta base de dados apresenta 50.7% de observações censuradas

e o tempo mediano de sobrevida (via Kaplan-Meier) é de 15.5 anos.

### 6.3.2 Resultados

Na procura de um modelo mais adequado para os dados de hemodiálise, modelos de mistura com dois componentes foram examinados primeiro. As medidas de evidência a favor de cada modelo, calculadas pelo FBST nas comparações entre os pares de modelos, são apresentadas na Tabela 6.2.

Tabela 6.2: Evidências fornecidas pelos dados de hemodiálise

Comparação	Hipótese nula	Evidência a favor da hipótese nula	
		e-valor	p-valor*
$H_L \times H_W$	$H_L$	0.874	0.404
	$H_W$	0.043	0.004
$H_L \times H_G$	$H_L$	0.901	0.446
	$H_G$	0.757	0.277
$H_G \times H_W$	$H_G$	1.000	1.000
	$H_W$	0.113	0.015

\*p-valor calculado segundo DINIZ *et al.* (2012)

Os resultados dos testes para comparação entre as distribuições lognormal e Weibull indicam a preferência pelo modelo lognormal em detrimento do modelo Weibull cuja evidência a favor é 0.043. Para comparação entre as distribuições lognormal e gama, as evidências em favor de  $H_L$  e  $H_G$  sugerem a não rejeição das hipóteses, indicando que ambos os modelos podem ser utilizados para a modelagem dos dados. Não obstante, o modelo lognormal mostra-se como o mais plausível, pois a evidência a seu favor é maior em relação a do modelo gama.

Comparando a distribuição gama com a Weibull, os testes forneceram os seguintes resultados: a evidência a favor da hipótese  $H_G$  é igual a 1.000, favorecendo, fortemente, o modelo gama. Para a hipótese  $H_W$  evidência obtida foi de 0.113, indicando a rejeição da hipótese. Portanto, nesta comparação, as evidências reforçam o que já tinha sido constatado anteriormente, de que o modelo Weibull não é adequado para a modelagem desses dados.

#### Aplicando o modelo de mistura LGW

Conforme mencionado inicialmente, o objetivo principal desta aplicação é ajustar o modelo de mistura com três componentes (LGW) à base de dados de hemodiálise, e utilizar o procedimento do FBST para escolher o modelo mais adequado.

As estimativas dos parâmetros do modelo LGW aplicado aos dados são apresentadas na Tabela 6.3. As evidências calculadas a favor das hipóteses sobre os pesos da mistura indicam que os modelos gama e Weibull devem ser descartados,

uma vez que as hipóteses  $H : p_2 = 0$  e  $H : p_3 = 0$  não são rejeitadas (Tabela 6.4). Consequentemente, entre os três modelos, apenas o modelo lognormal pode ser considerado para a modelagem dos dados de hemodiálise.

Tabela 6.3: Sumário da distribuição a posteriori dos parâmetros do modelo LGW aplicado aos dados de hemodiálise

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$p_1$ -lognormal	0.574	0.299	0.028	0.656	0.957
$p_2$ -gama	0.283	0.256	0.005	0.187	0.845
$p_3$ -Weibull	0.143	0.157	0.003	0.087	0.606
$\mu$	18.537	1.441	15.819	18.561	21.079
$\sigma^2$	204.744	78.788	84.547	197.103	339.416

Tabela 6.4: Testes de hipóteses para os pesos da mistura do modelo LGW aplicado aos dados de hemodiálise

Hipótese	e-valor	p-valor*
$p_1 = 0$	0.009	0.000
$p_2 = 0$	0.656	0.119
$p_3 = 0$	0.878	0.273

\*p-valor calculado segundo DINIZ *et al.* (2012)

Na Figura 6.2, parece que o modelo lognormal (Tabela 6.5), por si só, produz uma boa estimativa da função de sobrevivência. Note que, assim como acontece com o estimador Kaplan-Meier, a curva do estimador PEXE termina na última observação do conjunto de dados em análise.

Tabela 6.5: Sumário da distribuição a posteriori dos parâmetros do modelo lognormal aplicado aos dados de hemodiálise

Parâmetro	Média	D.P.	2.5%	Mediana	97.5%
$\mu$	20.298	1.543	17.493	20.251	23.285
$\sigma^2$	355.367	101.927	191.546	343.222	589.779

Observe que a preferência pelo modelo lognormal é mais evidente analisando o modelo de mistura LGW do que na comparação lognormal *vs.* gama, onde as evidências em favor dos modelos são muito próximas. Mais uma vez, isto mostra que o poder discriminatório do procedimento é maior quando todos os modelos candidatos são comparados simultaneamente, a partir de um modelo de mistura, do que quando comparações binárias são efetuadas, tal como discutido em SAWYER (1984).

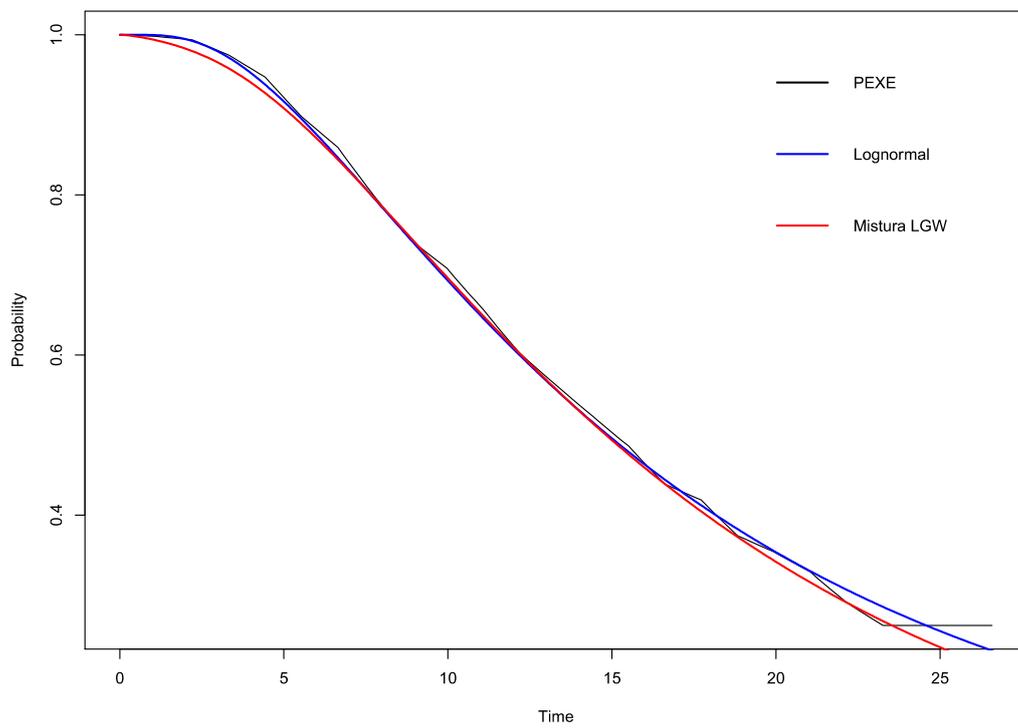


Figura 6.2: Curvas de sobrevivência estimadas pelo estimador PEXE e pelos modelos lognormal e de mistura LGW para dados de hemodiálise

# Capítulo 7

## Conclusões

A escolha de um modelo paramétrico conveniente, entre vários modelos alternativos, para ajuste de dados é uma tarefa importante em análise de sobrevivência e de confiabilidade. Mesmo diante de amostras de tamanho menor ou moderado, onde os modelos candidatos tendem a fornecer ajustes similares para os dados, a escolha do modelo mais plausível ainda torna-se necessária para que seja tomada a melhor decisão possível com base nas observações disponíveis (KUNDU e MANGLICK, 2004).

Neste trabalho considerou-se o FBST como procedimento de teste para escolha entre as distribuições de famílias separadas. O teste foi formulado num contexto de mistura linear dos modelos candidatos. A abordagem de mistura permite comparar, simultaneamente, todos os modelos alternativos através de testes de hipóteses sobre os pesos da mistura COX (1961); KAMARY *et al.* (2014), evitando-se, dessa forma, os problemas que surgem quando comparações binárias são efetuadas, conforme discutido em SAWYER (1984).

A novidade neste trabalho consistiu na reparametrização das funções densidades dos componentes da mistura em função da média  $\mu$  e da variância  $\sigma^2$ , da população. Isso fez com que todos os modelos candidatos compartilhassem dos mesmos parâmetros, uma vez que a comparação entre eles é baseada no mesmo conjunto de dados, isto é, na mesma população. Uma vantagem desta reparametrização, do ponto de vista de inferência, foi a de reduzir a quantidade de parâmetros do modelo de mistura.

O estudo de simulação de amostras com observações não censuradas foi desenvolvido com o objetivo de avaliar o desempenho do FBST, comparando-o com o do teste de Cox, na discriminação entre pares de modelos separados. O padrão geral dos resultados das simulações indica um comportamento do FBST similar ao do teste de Cox. As taxas de erro tipo I (taxa de rejeição do modelo verdadeiro), obtidas pelo FBST, mantêm-se sempre abaixo do nível de significância estipulado (5%), assim como os níveis de significância obtidos pelo teste de Cox, que, em geral, alcançam

valores muito próximos de 5%. Ao calcular as taxas de rejeição dos modelos falsos, ambos os testes mostraram-se consistentes, ou seja, as taxas convergiram para 1 com o aumento do tamanho de amostra. Não obstante, as taxas calculadas pelo FBST são, em geral, maiores do que as do teste de Cox, principalmente nas comparações lognormal *vs* Weibull e lognormal *vs* gama. Isto mostra que o poder discriminatório do FBST é maior do que o do teste de Cox. Observe que o teste de Cox é um procedimento formulado justamente para alcançar um alto poder contra a hipótese nula falsa (COX, 1961).

Nos exemplos de aplicações usados para ilustrar o procedimento proposto, os resultados levaram às mesmas conclusões anunciadas em outros trabalhos, que abordaram o mesmo problema, utilizando outros procedimentos de seleção de modelos, como por exemplo, os fatores de Bayes Fracionários, Intrínseco e Parcial. Portanto, considerando que o FBST permite o uso de prioris impróprias, a abordagem proposta neste trabalho pode ser utilizada para discriminar entre pares de modelos separados, mesmo para tamanhos de amostras pequenos.

Na simulação de amostras com dados censurados à direita, o FBST foi utilizado para discriminar entre os modelos lognormal, gama e Weibull, a partir da mistura LGW. Os resultados mostram uma boa eficiência do procedimento na identificação da verdadeira distribuição geradora dos tempos de falha, mesmo para amostras com taxa de censura alta.

A avaliação dos pesos da mistura do modelo LGW, através do FBST, permitiu identificar o modelo lognormal como o mais adequado para descrever os dados de pacientes com insuficiência renal crônica terminal submetidos à hemodiálise. Assim, para ajustar um modelo de regressão aos dados de hemodiálise, o modelo lognormal deverá ser considerado como a distribuição da variável resposta (o tempo de sobrevivência em hemodiálise).

Os resultados das aplicações mostram que a preferência do FBST pelo modelo com melhor ajuste é mais evidente quando se analisa o modelo de mistura com três componentes, do que os modelos com dois componentes. Em algumas comparações entre pares de modelos, é difícil escolher quando ambos os modelos fornecem um bom ajuste para os dados. Assim, para a escolha entre as distribuições consideradas neste trabalho, recomenda-se que o procedimento de teste seja aplicado sobre um modelo de mistura envolvendo todos os modelos candidatos cujas densidades devem ser reparametrizadas em termos da média e da variância da população.

## 7.1 Propostas para trabalhos futuros

Em sua tese de doutorado, ARAÚJO (1998) desenvolveu um estudo de simulação para comparar os fatores de Bayes alternativos (Intrínseco, a Posteriori e Fracionário) na discriminação entre os modelos separados, usando distribuições a priori impróprias para os parâmetros dos modelos. É de interesse comparar os resultados dessa simulação com os do procedimento FBST (formulado no contexto da mistura e da reparametrização dos componentes) usando essas priors e, principalmente, comparar os testes em estudos de casos aplicados como, por exemplo, modelagem de curvas de sobrevivência populacionais.

A identificação do modelo paramétrico com melhor ajuste aos tempos de falhas constitui o primeiro passo num estudo de dados de sobrevivência. O passo seguinte seria a construção de um modelo de regressão para avaliar os fatores prognósticos, considerando o modelo escolhido anteriormente como a distribuição da variável resposta (tempo de sobrevivência). E, além disso, obter uma distribuição preditiva e usar a abordagem *leave-one-out cross validation* para avaliar o desempenho do modelo de regressão nas previsões sobre novas observações. Este tópico foi objeto de estudo da tese de doutorado de Brian Alvarez Ribeiro de Melo, apresentado recentemente no Instituto de Matemática e Estatística da USP.

## 7.2 Trabalhos resultantes da tese

Os artigos listados a seguir abordam os tópicos desenvolvidos nesta tese. O primeiro foi aceito para publicação na Revista Brasileira de Biometria, e os demais estão em processo de submissão.

- The FBST procedure for discriminating between the lognormal and the Weibull distributions.
- Model Choice in separate families: A comparison between the FBST and the Cox test
- Bayesian significance test for discriminating between survival distribuitons

# Referências Bibliográficas

- AITKIN, M., 1991, “Posterior Bayes factors (with discussion)”, *Journal of the Royal Statistical Society*, v. 53, pp. 111–142.
- AITKIN, M., 1992, “Evidence and the posterior factor”, *The Mathematical Scientist*, v. 17, pp. 15–25.
- AITKIN, M., BOYS, R., CHADWICK, T., 2005, “Bayesian point null hypothesis testing via the posterior likelihood ratio”, *Statistics and Computing*, v. 15, pp. 215–230.
- ALLENBY, G. M., 1993, “Hypothesis testing with scanner data: the advantage of Bayesian methods”, *J. Marketing Res.*, v. 27, pp. 379–389.
- ALVES, M. AND SOUZA E SILVA, N. A., SALIS, L. H. A., PEREIRA, B. B., et al., 2014, “Sobrevida e Fatores Preditores de Letalidade em Hemodiálise: Polimorfismos D/I do Gene da Enzima Conversora da Angiotensina I e o M235T do Angiotensinogênio”, *Arq Bras Cardiol.*, v. 103, pp. 209–218.
- ARAÚJO, M. I., PEREIRA, B. B., 2007, “A Comparison of Bayes Factors for Separated Models: Some Simulation Results”, *Communications in Statistics–Simulation and Computation*, v. 36, pp. 297–309.
- ARAÚJO, M. I., PEREIRA, B. B., 2007b, “Fatores de Bayes para discriminação de modelos separados de regressão multivariada usando priores impróprios”, *Revista Brasileira de Estatística*, v. 68, pp. 33–50.
- ARAÚJO, M. I., PEREIRA, B. B., CLEROUX, R., et al., 2005, “Separate families of models: Sir David Cox contributions and recent developments”, *Student*, v. 5, pp. 251–258.
- ARAÚJO, M. I., 1998, *Comparação de modelos separados – uma abordagem Bayesiana usando fatores de Bayes com distribuição a priori impróprias*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ATKINSON, A. C., 1970, “A method for discriminating between models (with discussion)”, *Journal of the Royal Statistical Society*, v. 32, pp. 323–353.

- AURIKKO, E., 1985, “Testing the functional form of Finnish aggregate imports”, *Econ. Lett.*, v. 18, pp. 223–228.
- BASU, S., 1996, “A New Look at Bayesian Point Null Hypothesis Testing”, *Sankhya*, v. 58, pp. 292–310.
- BERGER, J. O., PERICCHI, L. R., 2001, “Objective Bayesian Methods for Model Selection: Introduction and Comparison”, *Lecture Notes-Monograph Series*, v. 38, pp. 135–207.
- BERGER, J. O., PERICCHI, L. R., 1996, “The Intrinsic Bayes Factor for Model Selection and Prediction”, *Journal of the American Statistical Association*, v. 91, pp. 109–122.
- BERGER, J., DELAMPADY, M., 1987, “Testing precise hypothesis”, *Statistical Science*, v. 2, pp. 317–352.
- BÖHNING, D., SEIDEL, W., ALFÓ, M., et al., 2007, “Advances in Mixture Models”, *Computational Statistics and Data Analysis*, v. 51, pp. 5205–5210.
- BORGES, W., STERN, J. M., 2007, “The Rules of Logic Composition for the Bayesian Epistemic e-Values”, *Logic Journal of the IGPL*, v. 15, pp. 401–420.
- BURKE, S. P., GODFREY, L. G., TREMAYNE, A. R., 1990, “Testing AR(1) against MA(1) disturbances in the linear regression models: an alternative procedure”, *Rev. Econ. Studies*, v. 57, pp. 135–145.
- BUSE, A., 1982, “The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note”, *The American Statistician*, v. 36, pp. 153–157.
- CABRAS, S., RACUGNO, W., VENTURA, L., 2015, “Higher order asymptotic computation of Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters”, *Journal of Statistical Computation and Simulation*, v. 85, pp. 2989–3001.
- CHAMBERS, M. J., 1993, “A non nested approach to testing continuous time models against discrete alternatives”, *J. Econometrics*, v. 57, pp. 319–343.
- COX, D. R., 2013, “A return to an old paper: Tests of separate families of hypotheses (with discussion)”, *Journal of the Royal Statistical Society*, v. 75, pp. 207–215.

- COX, D. R., 1961, “Tests of separate families of hypotheses”, *Proceedings 4th Berkeley Symposium in Mathematical Statistics and Probability*, v. 1, pp. 105–123.
- COX, D. R., 1962, “Further results on test of separate families of hypotheses”, *Journal of the Royal Statistical Society*, v. B, pp. 406–424.
- COX, D. R., 1977, “The role of significance tests”, *Scand. J. Statist*, v. 4, pp. 49–70.
- DASTOOR, N. K., 1983, “Some aspects of testing non-nested hypotheses”, *Journal of Econometrics*, v. 21, pp. 213–228.
- D’CUNHA, J. G., RAO, K. A., 2016, “Full and Marginal Bayesian Significance Test: A Frequentist Comparison”, *IOSR Journal of Mathematics*, v. 12, pp. 71–78.
- DE ARRUDA, M. L., 2012, *FBST sequencial*. Tese de D.Sc., IME/USP, São Paulo, SP, Brasil.
- DE JONG, D. N., 1993, “Bayesian inference in limited dependent variable models: an application to measuring strike duration”, *Journal of Applied Econometrics*, v. 8, pp. 115–128.
- DIEBOLT, J., ROBERT, C. P., 1994, “Estimation of Finite Mixture Distributions Through Bayesian Sampling”, *Journal of the Royal Statistical Society*, v. 56, n. 2, pp. 363–375.
- DINIZ, M., PEREIRA, C. A. B., POLPO, A., et al., 2012, “Relationship between Bayesian and Frequentist significance indices”, *International Journal for Uncertainty Quantification*, v. 2, pp. 161–172.
- DUMONCEAUX, R., ANTLE, C. E., HAAS, G., 1973, “Likelihood ratio test for discriminating between two models with unknown location and scale parameters”, *Technometrics*, v. 15, pp. 19–27.
- FIRMINO, P. R. A., 2009, *Método Adaptativo de Markov Chain Monte Carlo para manipulação de modelos Bayesianos*. Tese de D.Sc., PPGEP/UFPE, Recife, PE, Brasil.
- FLETCHER, R., REEVES, C. M., 1964, “Function minimization by conjugate gradients”, *Computer Journal*, v. 7, pp. 148–154.
- FRUHWIRTH-SCHNATTER, S., 2006, *Finite Mixture and Markov Switching Models*. Springer.

- GELFAND, A. E., DEY, D. K., 1994, “Bayesian model choice: asymptotic and exact calculations”, *Journal of the Royal Statistical Society*, v. 56, pp. 501–504.
- GELMAN, A., 2006, “Prior distributions for variance parameters in hierarchical models”, *Bayesian Analysis*, v. 1, pp. 515–533.
- GHOSH, J. K., SUBRAMANYAM, K., 1975, “Inference about separated families in large samples”, *Sankhyā: The Indian Journal of Statistics*, v. 37, pp. 502–513.
- GHOSH, J. K., SAMANTA, T., 2002, “Nonsubjective Bayes testing – an overview”, *Journal of Statistical Planning and Inference*, v. 103, pp. 205–223.
- GHOSH, J. K., DELAMPADY, M., SAMANTA, T., 2006, *An Introduction to Bayesian Analysis: Theory and Methods*. Springer.
- GIJBELS, I., 2010, “Censored data”, *Computational Statistics*, v. 2, pp. 178–188.
- GODFREY, L. G., TREMAYNE, A. R., 1988, “Checks for model adequacy for univariate time series models and their application to econometric relationship”, *Econometric Rev.*, v. 7, pp. 1–64.
- GOURIEROUX, C., MONFORT, A., 1994, “Testing Non-Nested Hypotheses”, *In Handbook of Econometrics, Ed. Engle, R.F and McFadden, D.L.*, v. 4, pp. 2585–2637.
- HAARIO, H., SAKSMAN, E., TAMMINEN, J., 2001, “An adaptive Metropolis algorithm”, *Bernoulli*, v. 7, pp. 223–242.
- HARVEY, A. C., 1985, “Discrimination between CES and VES production functions”, *Ann. Econ. Soc. Measurement*, v. 4, pp. 463–471.
- KAMARY, K., MENGERSEN, K., ROBERT, C., et al., 2014, “Testing hypotheses via a mixture estimation model”, *arXiv:1412.2044v2*.
- KEMPTHORNE, O., 1976, “Of what use are tests of significance and tests of hypothesis”, *Communications in Statistics -Theory and Methods*, v. 8, pp. 763–777.
- KIM, J. S., PROSCHAN, F., 1991, “Piecewise exponential estimator of the survivor function.” *IEEE Transactions on Reliability*, v. 40, pp. 134–139.

- KLEIN, J., MOESCHBERGER, M. L., 2003, *Survival analysis: techniques for censored and truncated data*. Springer.
- KUNDU, D., MANGLICK, A., 2004, “Discriminating between the Weibull and Log-Normal Distributions”, *Naval Research Logistics*, v. 51, pp. 893–905.
- KUNDU, D., MANGLICK, A., 2005, “Discriminating between the Log-Normal and gamma distributions”, *Journal of the Applied Statistical Sciences*, v. 14, pp. 175–187.
- LAGAKOS, S. W., BARRAJ, L. M., DE GRUTTOLA, V., 1988, “Nonparametrics analysis of truncated survival data, with application to AIDS”, *Biometrika*, v. 75, pp. 515–523.
- LAURETTO, M., PEREIRA, C. A. B., STERN, J. M., et al., 2003, “Comparing parameters of two bivariate normal distributions using the invariant full Bayesian significance test”, *Brazilian Journal of Probability and Statistics*, v. 17, pp. 147–168.
- LAURETTO, M. S., 2007, *Seleção de Modelos Através de um Teste de Hipótese Genuinamente Bayesiano: Mistura de Normais Multivariadas e Hipóteses Separadas*. Tese de D.Sc., IME/USP, São Paulo, SP, Brasil.
- LAURETTO, M. S., STERN, J. M., 2005, “FBST for mixture model selection”, *AIP Conference Proceedings*, v. 803, pp. 121–128.
- LAURETTO, M. S., FARIA JR, S. R., PEREIRA, C. A. B. AND PEREIRA, B. B., et al., 2007, “The problem of separate hypotheses via mixture models”, *AIP Conference Proceedings*, v. 954, pp. 268–275.
- LAURETTO, M. S., NAKANO, F., FARIA JR, S. R., et al., 2009, “A straightforward multiallelic significance test for the Hardy-Weinberg equilibrium law”, *Genetics and Molecular Biology*, v. 32, pp. 619–625.
- LAWLESS, J. F., 2002, *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons.
- LEE, E. T., WANG, J. W., 2003, *Statistical Methods for Survival Data analysis*. Wiley.
- LEHMANN, E., 1999, *Elements of Large-Sample Theory*. Springer.
- LINDLEY, D. V., 1993, “On the presentation of evidence”, *Mathematical Scientist*, v. 18, pp. 60–63.

- LINDSEY, J. K., 1974, “Comparison of probability distribution”, *Journal of the Royal Statistical Society*, v. Series B 36, pp. 38–47.
- MACKINNON, J. G., 1983, “Model specification tests against non-nested alternatives”, *Econometric Rev.*, v. 2, pp. 85–110.
- MADRUGA, M., ESTEVES, L., WECHSLER, S., 2001, “On the Bayesianity of Pereira-Stern tests”, *Test*, v. 10, pp. 291–299.
- MADRUGA, M., PEREIRA, C. A. B., STERN, J. M., 2003, “Bayesian evidence test for precise hypotheses”, *Journal of Statistical Planning and Inference*, v. 117, pp. 185–198.
- MCALEER, M., 1995, “The significance of testing empirical non-nested models”, *Journal of Econometrics*, v. 67, pp. 149–171.
- MCALEER, M., PESARAN, M. H., 1987, “Specification tests for separate models: a survey”, *Specification Analysis in the Linear Model*, pp. 146–196.
- MCALEER, M., PESARAN, M. H., 1986, “Statistical inference in non nested econometric models”, *Appl. Math. Computat.*, v. 20, pp. 271–311.
- MCLACHLAN, G. L., PEEL, D., 2000, *Finite Mixture Models*. Wiley.
- MOHAMMED, Y. A., YATIM, B., ISMAIL, S., 2013, “A Simulation Study of a Parametric Mixture Model of Three Different Distributions to Analyze Heterogeneous Survival Data”, *Modern Applied Science*, v. 7, pp. 1844–1852.
- NAZARETH, J. L., 2009, “Conjugate gradient method”, *WIREs: Computational Statistics*, v. 1, pp. 348–353.
- NG, K. W., TIAN, G. L., TANG, M. L., 2011, *Dirichlet and Related Distributions: Theory, Methods and Applications*. Wiley.
- O’HAGAN, A., 1995, “Fractional Bayes Factors for Model Comparison (with discussion)”, *Journal of the Royal Statistical Society*, v. 57, pp. 99–138.
- O’HAGAN, A., 1997, “Properties of intrinsic and fractional Bayes factors”, *Test*, v. 6, pp. 101–108.
- PAULINO, C. D., TURKMAN, M., MURTEIRA, B., 2003, *Estatística Bayesiana*. Fundação Calouste Gulbenkian.
- PEREIRA, B. B., 2005, “Separate Families of Hypotheses”, *In Encyclopedia of Biostatistics*, Ed. Armitage, P. and Colton, T., v. 7, pp. 4881–4886.

- PEREIRA, B. B., 2010, “Tests for discriminating separate or non-nested models”, *In International Encyclopedia of Statistical Science, Ed. Lovic, M.*, v. 3, pp. 1592–1595.
- PEREIRA, B. B., 1977a, “A note on the consistency and on finite sample comparisons of some tests of separate families of hypotheses”, *Biometrika*, v. 64, pp. 109–113.
- PEREIRA, B. B., 1978, “Empirical comparisons of some tests of separate families of hypotheses”, *Metrika*, v. 25, pp. 219–234.
- PEREIRA, B. B., 1979, “Testes para discriminar entre as distribuições lognormal, gama e Weibull”, *Estadística-Journal of the Inter-American Statistical Institute*, v. 33, pp. 41–46.
- PEREIRA, B. B., 1981, “Choice of a survival model for patients with a brain tumour”, *Metrika*, v. 28, pp. 53–61.
- PEREIRA, B. B., PEREIRA, C. A. B., 2017, *Model Choice in Nonnested Families*. Springer.
- PEREIRA, B. B., 1976, *Some Results on Tests of Separate Families of Hypotheses*. Ph.D dissertation, Imperial college/University of London, London, UK.
- PEREIRA, C. A., WECHSLER, S., 1993, “On the concept of P-value”, *Revista Brasileira de Probabilidade e Estatística*, v. 7, pp. 159–177.
- PEREIRA, C. A. B., STERN, J., 1999, “Evidence and Credibility: full Bayesian significance test for precise hypotheses”, *Entropy*, v. 1, pp. 69–80.
- PEREIRA, C. A. B., STERN, J., WECHSLER, S., 2008, “Can a significance test be genuinely Bayesian”, *Bayesian Analysis*, v. 3, pp. 79–100.
- PERICCHI, L. R., 2005, “Model selection and hypothesis testing based on objective probabilities and Bayes factors”, *In Dey, D. K. and Rao, C. R., Handbook of Statistics: Bayesian Thinking. Modeling and Computation*, v. 25, pp. 115–149.
- PESARAN, B., PESARAN, M. H., 1995, “A non nested test of level-differenced versus logdifferenced stationary models”, *Econometric Rev.*, v. 14, pp. 213–227.
- PESARAN, M. H., 1974, “On the general problem of model selection”, *Rev. Econ. Stud.*, v. 41, pp. 153–171.

- PESARAN, M. H., 1982, “On the comprehensive method of testing non-nested regression models”, *Journal of Econometrics*, v. 18, pp. 263–274.
- PESARAN, M. H., 1994, “Global and partial nonnested hypothesis and asymptotic local power”, *Econometric Theory*, v. 3, pp. 69–97.
- QUANDT, R. E., 1974, “A comparison of methods for testing nonnested hypotheses”, *The Review of Economics and Statistics*, v. 56, pp. 92–99.
- QUESENBERRY, C. P., KENT, J., 1982, “Selecting among Probability Distributions Used in Reliability”, *Technometrics*, v. 24, pp. 59–65.
- ROBERTS, G. O., ROSENTHAL, J. S., 2009, “Exemples of Adaptive MCMC”, *Journal of Computational and Graphical Statistics*, v. 18, pp. 349–367.
- ROJAS, F. A. R., 2001, *Avaliação de testes de hipóteses separadas*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ROSSI, P., 1985, “Comparison of alternative functional forms in production”, *J. Econometrics*, v. 30, pp. 345–361.
- RUST, R. T., SCHMITTLEIN, D. S., 1985, “A Bayesian cross-validated likelihood method for comparing alternative specifications of quantitative models”, *Marketing Sci.*, v. 4, pp. 419–424.
- SAWYER, K. R., 1984, “Multiple hypotheses testing”, *Journal of the Royal Statistical Society-B* 46, pp. 419–424.
- SILVAPULLE, P., KING, M. L., 1993, “Non nested testing for autocorrelation in the linear regression model”, *J. Econometrics*, v. 58, pp. 295–314.
- STATISTICAT, LCC, 2016, *LaplacesDemon: A Complete Environment for Bayesian Inference within R*. R Package version 17.07.2016. <https://cran.r-project.org/web/packages/LaplacesDemon/LaplacesDemon.pdf>.
- STERN, J., PEREIRA, C. A. B., 2014, “Bayesian epistemic values: focus on surprise, measure probability”, *Logic Journal of The IGPL*, v. 22, pp. 236–254.
- STERN, J. M., ZACKS, S., 2002, “Testing independence of poisson variates under the Holgate bivariate distribution: The power of a new evidence test”, *Statistical and Probability Letters*, v. 60, pp. 313–320.
- STERN, J., 2005, “Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses”, *Cybernetics and Human Knowing*, v. 14, pp. 9–36.

- STERN, R. B., PEREIRA, C. A. B., 2012, “Statistical Information: A Bayesian Perspective”, *Entropy*, v. 14, pp. 2254–2264.
- TIAN, Y., LIU, Y., CHEN, P., 2010, “Bayesian analysis for mixture of log-normal distributions”, *Scientia Magna*, v. 6, pp. 65–71.
- WAN, F., 2017, “Simulating survival data with predefined censoring rates for proportional hazards models”, *Statistics in Medicine*, v. 36, pp. 838–854.
- WECHSLER, S., PEREIRA, C. A. B., MARQUES, P. C., 2008, “Birnbaum’s theorem redux”, *AIP Conference Proceedings*, v. 1073, pp. 96–100.
- WEST, M., HARRISON, J., 1997, *Bayesian Forecasting and Dynamic Models*. Springer.
- WIGINTON, J. C., 1993, “A Bayesian approach to discrimination among economic models”, *Decisions Sci.*, v. 5, pp. 182–193.
- ZELLNER, A., 1984, “Posterior odds ratios for regression hypotheses: general considerations and some specific results”, *In Zellner, A., ed. Basic Issues in Econometrics*, v. 57, pp. 275–305.
- ZHANG, H., HUANG, Y., 2015, “Finite Mixture Models and Their Applications: A Review”, *Austin Biometrics and Biostatistics*, v. 2, pp. id1013.