

# **Modelo de Mistura Padrão com Tempo de Falha Exponencial e Censura Informativa**

Luiz Antonio de Freitas

# Modelo de Mistura Padrão com Tempo de Falha Exponencial e Censura Informativa

Luiz Antonio de Freitas

Orientador: Prof. Dr. Josemar Rodrigues

Tese apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs - UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística.

UFSCar - São Carlos - SP

Junho de 2010

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária/UFSCar**

F866mm

Freitas, Luiz Antonio de.

Modelo de mistura padrão com tempo de falha  
exponencial e censura informativa / Luiz Antonio de Freitas.  
-- São Carlos : UFSCar, 2010.  
113 f.

Tese (Doutorado) -- Universidade Federal de São Carlos,  
2010.

1. Análise de sobrevivência. 2. Mistura de distribuições. 3.  
Inferência bayesiana. 4. Simulação estocástica. I. Título.

CDD: 519.9 (20<sup>a</sup>)

**Luiz Antonio de Freitas**

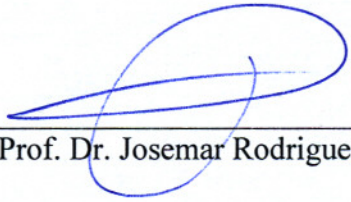
**Modelo de Mistura Padrão com Tempo de Falha Exponencial e Censura Informativa**

Tese apresentada à Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Doutor em Estatística.

Aprovada em 25 de junho de 2010.

**BANCA EXAMINADORA**


Presidente

  
\_\_\_\_\_  
Prof. Dr. Josemar Rodrigues (DEs-UFSCar/Orientador)

1º Examinador

  
\_\_\_\_\_  
Prof. Dr. Heleno Bolfini (IME-USP)

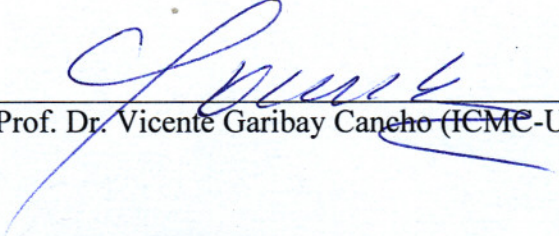
2º Examinador

  
\_\_\_\_\_  
Prof. Dr. Jorge Luis Bazán Guzman (PUC-Perú)

3º Examinador

  
\_\_\_\_\_  
Prof. Dr. Reinaldo Boris Arellano Valle (PUC-Chile)

4º Examinador

  
\_\_\_\_\_  
Prof. Dr. Vicente Garibay Cancho (ICMC-USP)

À Lara.

Os grandes logros da vida são produtos  
de pequenos esforços do dia a dia.

Anônimo

## Agradecimentos

- Ao Professor Dr. Josemar Rodrigues pela orientação e pelas idéias durante todo este trabalho.
- À CAPES (Coordenação de Aperfeiçoamento Pessoal de Nível Superior) pela assistência financeira.
- Do fundo do coração à todos aqueles que de uma forma ou de outra estiveram presentes na minha vida nestes dias de doutorado.
- À Lara, que nestes últimos anos de doutorado esteve comigo e me ensinou a observar o lado humano da vida.
- Àquele pantanero de Aquidauana. Resta lembra-lo aqui. Sou grato eternamente por aquele gesto sutil e descompromissado.
- Aos velhos pantaneros, Lito e Jayme, que agora tão por outras banda.
- À Dona Nanhã, velha pantanera do Paiaguás.

Muito obrigado,  
Luiz Antonio de Freitas.

# Sumário

Lista de Figuras . . . . .	v
Lista de Tabelas . . . . .	vi
Resumo . . . . .	viii
Abstract . . . . .	ix
<b>1 Introdução . . . . .</b>	<b>1</b>
1.1 Modelo de Mistura Padrão Exponencial . . . . .	2
1.2 Motivação . . . . .	3
1.3 Objetivos . . . . .	5
1.4 Organização do Trabalho . . . . .	6
<b>2 Modelo Unificado de Longa Duração . . . . .</b>	<b>10</b>
2.1 Função de Sobrevivência de Longa Duração . . . . .	11
2.2 Função Sub-Densidade e Função de Risco Populacional . . . . .	15



2.3	Heterogeneidade Entre os Indivíduos e Propriedade de Riscos Proporcionais . . . . .	21
2.4	Relações Paramétricas . . . . .	22
2.5	Função de Verossimilhança . . . . .	26
2.6	Resumo dos Principais resultados do Capítulo . . . . .	30
<b>3</b>	<b>Modelo de Mistura Padrão Exponencial . . . . .</b>	<b>32</b>
3.1	Modelo de Mistura Padrão Exponencial . . . . .	32
3.2	Relações Paramétricas . . . . .	35
3.3	Função de Verossimilhança . . . . .	38
3.4	Distribuição de $T$ . . . . .	39
3.5	Resumo dos Principais resultados do Capítulo . . . . .	40
<b>4</b>	<b>Inferência Clássica . . . . .</b>	<b>41</b>
4.1	Inferência Sob Suposição de Censura Não-Informativa . . . . .	42
4.1.1	Matriz de Informação . . . . .	42
4.2	Inferência Sob Suposição de Censura Informativa . . . . .	45
4.2.1	Matriz de Informação . . . . .	46
4.2.2	Uma Opção Alternativa de Inferência sob Censura Informativa . . . . .	47
4.3	Dados Com Variável Latente . . . . .	48
4.3.1	Sob Suposição de Censura Não-Informativa . . . . .	48

4.3.2	Sob Suposição de Censura Informativa . . . . .	52
4.4	Algoritmo de Maximização Iterativa . . . . .	53
<b>5</b>	<b>Inferência Bayesiana Objetiva . . . . .</b>	<b>57</b>
5.1	Modelo Com Censura Não-Informativa . . . . .	58
5.2	Modelo Com Censura Informativa . . . . .	60
5.3	Modelo Com Dados Ampliados . . . . .	61
5.3.1	Modelo Com Censura Não-Informativa . . . . .	61
5.3.2	Modelo Com Censura Informativa . . . . .	62
5.4	Uma Inferência Bayesiana Empírica . . . . .	63
<b>6</b>	<b>Estudo de Simulação e Aplicação com Dados Reais . . . . .</b>	<b>68</b>
6.1	Algoritmo de Simulação de Dados . . . . .	69
6.2	Metodologia Sob Abordagem Clássica . . . . .	70
6.3	Metodologia Sob Abordagem Bayesiana . . . . .	71
6.4	Apresentação Gráfica e em Tabela . . . . .	73
6.4.1	Notações dos Procedimentos . . . . .	73
6.5	Estudos com Dados Simulados . . . . .	74
6.5.1	Inferência Sobre $\lambda$ . . . . .	75
6.5.2	Inferência Sobre $p_0$ . . . . .	78
6.5.3	Algumas Considerações . . . . .	79

6.6	Dados Reais: Melanoma Maligno E1690 . . . . .	81
<b>7</b>	<b>Conclusões e Propostas de Trabalhos Futuros . . . . .</b>	<b>88</b>
7.1	Sobre censura Informativa . . . . .	90
7.2	Outros Resultados . . . . .	93
7.3	Propostas de Trabalhos Futuros . . . . .	93
<b>A</b>	<b>Demonstração do Teorema 3.4 . . . . .</b>	<b>95</b>
A.1	Demonstração de Teorema 3.4 . . . . .	96
<b>B</b>	<b>Funções Geradoras . . . . .</b>	<b>97</b>
<b>C</b>	<b>Programas em R . . . . .</b>	<b>98</b>
C.1	Geração de dados . . . . .	98
C.2	Inferência sob Censura Informativa, Função de Verossimilhança (4.16) . . . . .	99
C.3	Estudo de Cobertura dos Intervalos de Credibilidade Associados à Distribuição <i>A Posteriori</i> conjunta (5.28) . . . . .	103
C.4	Sobre as Inferências em Código R . . . . .	109
	<b>Referências . . . . .</b>	<b>110</b>

# Lista de Figuras

2.1	<i>Representação da região de censura dos indivíduos em risco. No eixo das ordenadas temos a variável <math>Y M \geq 1</math>.</i>	25
3.1	<i>Representação da região de censura dos indivíduos em risco. No eixo das ordenadas temos a variável <math>Y M = 1</math>.</i>	36
6.1	<i>Legendas utilizadas.</i>	73
6.2	<i>Inferências sobre <math>\lambda</math>: Na primeira linha temos dados simulados de tamanho <math>n = 40</math>, e na segunda, de tamanho <math>n = 400</math>; nas colunas, da esquerda para a direita, plot da porcentagem de coberturas, do comprimento médio dos intervalos e da média frequentista das estimativas pontuais.</i>	75
6.3	<i>Inferências sobre <math>p_0</math>: Na primeira linha temos dados simulados de tamanho <math>n = 40</math>, e na segunda, de tamanho <math>n = 400</math>; nas colunas, da esquerda para a direita, plot da porcentagem de coberturas, do comprimento médio dos intervalos e da média frequentista das estimativas pontuais.</i>	78
6.4	<i>Média das estimativas de máxima verossimilhança de <math>\eta</math> (linha cheia) e valor esperado de <math>\eta</math> (linha pontilhada), para amostras de tamanho <math>n = 40</math> e <math>n = 400</math>, obtida destas simulações, sob o procedimento <b>Ina</b>.</i>	79

6.5	<i>Estimativas de Kaplan-Meier sob os dois grupos dos dados E1690, em que o símbolo + indica a ocorrência de censura. . . . .</i>	81
6.6	<i>Ajuste das estimativas de Kaplan-Meier ao modelo de mistura padrão exponencial (3.4), modelo de Barkson-Gage, sob os emv's de <math>\lambda</math> e <math>p_0</math> com procedimento <b>Ina</b>. . . . .</i>	82
6.7	<i>Ajuste ao modelo exponencial, sob dois métodos alternativos; na primeira linha, método 1: plot de Kaplan-Meier versus <math>S(t)</math>: exponencial; na segunda linha, método 2: plot de tempos versus <math>-\log\left(\frac{ekm-p_0}{1-p_0}\right)</math>, em que <math>ekm</math> é a estimativa de Kaplan-Meier. Na primeira coluna, ajustes para o grupo observação; na segunda, ajustes para o grupo interferon. O ajuste ao modelo exponencial é feito sob os emv's de <math>\lambda</math> e <math>p_0</math>, com o procedimento <b>Ina</b>. . . . .</i>	83
6.8	<i>Sob o conjunto de dados E1684, ajuste das estimativas de Kaplan-Meier ao modelo de mistura padrão exponencial (3.4), modelo de Barkson-Gage, sob os emv's de <math>\lambda</math> e <math>p_0</math> com procedimento <b>Ina</b>. . .</i>	86
6.9	<i>MCMC da distribuição conjunta a posteriori (5.11): Na primeira e segunda linhas, grupo observação; na terceira e quarta, grupo interferon. Para cada gráfico, na primeira coluna, históricos das quatro cadeias, na segunda, autocorrelação das cadeias. . . . .</i>	87

# Lista de Tabelas

- 6.1 *Comprimento  $I_\lambda$  e  $I_{p_0}$ , dos intervalos de  $\lambda$  e de  $p_0$ , respectivamente, e estimativas  $\tilde{\lambda}$  de  $\lambda$ , e  $\tilde{p}_0$  de  $p_0$ . Na penúltima coluna, o procedimento **Iibe** é implementado com o uso da emv de  $\eta$  a partir dos dados E1690. Na última coluna, o procedimento **Iibe\*** é implementado com o uso da emv de  $\eta$  a partir dos dados E1684.* 84

# Resumo

Neste trabalho consideramos o modelo de sobrevivência de longa duração introduzido por Berkson & Gage (1952), que serve para modelar dados de populações não homogêneas, em que parte da população não apresenta o evento de interesse mesmo após um longo período de observação. Os modelos com fração de cura apresentados na literatura são usualmente desenvolvidos sob a suposição de censura não informativa. Sob o modelo usual de sobrevivência, Lawless (1982) considera que a variável de censura é informativa se suas funções de densidade e de distribuição acumulada envolvem algum parâmetro de interesse. Neste trabalho enunciamos uma nova definição de censura informativa, que é similar à de Lawless (1982). Esta definição é estendida para o modelo unificado de longa duração proposto por (Rodrigues *et al.*, 2009). Também verificamos, com uso de dados simulados, o impacto da censura informativa na cobertura e no comprimento dos intervalos assintóticos dos parâmetros de interesse. Uma abordagem bayesiana com distribuições *a priori* de Jeffreys é proposta. Um exemplo com dados reais é analisado.

# Abstract

In this work we consider the long-term survival model introduced by Berkson & Gage (1952), for modeling survival data of nonhomogeneous populations, where a subpopulation does not present the event of interest, despite a long follow-up period. The cure rate models presented in the literature usually are developed under the assumption that censorship is noninformative. In the usual survival models Lawless (1982) considers that the variable of censoring is informative if its density function and its distribution function involve some parameter of interest. We propose a new definition of informative censoring in a similar way. This definition is extended for the unified long-term survival models (Rodrigues *et al.*, 2009). Moreover, we verify, with simulated data, the impact caused by informative censoring in the coverage probabilities and in the lengths of asymptotic confidence intervals of the parameters of interest. A Bayesian approach with Jeffreys prior is also proposed. An example with real data is analysed.



# Capítulo 1

## Introdução

Na análise de dados de sobrevivência é frequente a ocorrência de um grande número de indivíduos censurados à direita. Destes, parte não apresenta o evento de interesse mesmo após um longo período de acompanhamento. Estas situações podem ocorrer, por exemplo, em ensaios clínicos sobre reincidência de câncer, em estudos relacionados ao desenvolvimento da AIDS em pacientes HIV-positivos e testes de durabilidade de componentes eletrônicos (Mizoi, 2004). Estes casos não são previstos pela teoria de sobrevivência tradicional, que considera que todos os indivíduos irão apresentar o evento de interesse devido a uma única causa, mas é previsto pelos modelos de sobrevivência com fração de cura, também conhecidos como modelos de sobrevivência de longa duração.

As primeiras propostas de modelos com fração de cura foram apresentadas por Boag (1949) e Berkson & Gage (1952), que propuseram um modelo de mistura em que há uma proporção  $p_0$  de indivíduos curados e uma proporção  $1 - p_0$  de indivíduos não-curados. Este modelo ficou também conhecido como modelo de mistura padrão (*standard cure rate model*, modelo de cura padrão), e é o modelo de longa duração mais conhecido. Em Yakovlev & Tsodikov (1996) são discutidos modelos alternativos ao de mistura padrão, proposto para modelar latência de tumores, e são similares ao usado em Chen *et al.* (1999) para modelar o número de células susceptíveis a desenvolver metástase. Rodrigues *et al.* (2009) propõem um

modelo unificado, do qual o modelo de mistura e o modelo de Chen *et al.* (1999) são casos particulares, o qual chamaremos, por comodidade didática, de modelo unificado. Fazemos uma apresentação detalhada deste modelo no Capítulo 2.

Existe uma extensa literatura sobre os modelos de longa duração, que está em evolução. Dentre os livros citamos Maller & Zhou (1996), Yakovlev & Tsodikov (1996) e Ibrahim *et al.* (2001), os quais apontamos como referências fundamentais. Entretanto, devemos destacar que o modelo de sobrevivência tradicional é um caso particular dos modelos de longa duração, e portanto não podemos deixar de citar também os livros de Kalbfleisch & Prentice (2002), Lawless (1982) e Colosimo & Giolo (2006).

Dentre os trabalhos, citamos alguns mais recentes, que julgamos serem importantes como referência. Chen & Ibrahim (2001) discutem estimação por máxima verossimilhança em um modelo semiparamétrico. Em Chen *et al.* (2002) são considerados três modelos e uma abordagem bayesiana, com distribuições *a priori* não-informativas e uma classe de distribuições *a priori* informativas, elicítadas mediante conjunto de dados pré-existentes. Mizoi (2004) aborda um modelo com fração de cura e com erro de medida nas covariáveis. Citamos também Yakovlev (1994), Yau & Ng (2001), Yin & Ibrahim (2005), Tsodikov *et al.* (2003) e Zhou & Maller (1995).

## 1.1 Modelo de Mistura Padrão Exponencial

Primeiramente apresentamos o modelo de mistura padrão, na forma

$$\mathcal{S}_p(t) = p_0 + (1 - p_0) \mathcal{S}(t), \quad (1.1)$$

em que  $\mathcal{S}_p(t)$  é a função de sobrevivência dos indivíduos da população, dos quais existe uma proporção  $p_0$  de indivíduos curados e uma proporção  $1 - p_0$  de indivíduos não-curados, e  $\mathcal{S}(t)$  é a função de sobrevivência dos indivíduos não-curados. O modelo de mistura padrão foi proposto por Boag (1949) e Berkson & Gage (1952), do qual o modelo a ser considerado como objeto deste

trabalho é o proposto em Berkson & Gage (1952).

Em Berkson & Gage (1952) é dada ênfase ao caso clínico de pacientes com câncer de estômago e de mama, e é proposta a expressão

$$P_T = p_0 c_0 + (1 - p_0) c_0 e^{-\beta t}, \quad (1.2)$$

a qual chamamos de modelo de mistura padrão exponencial. Em (1.2) a quantidade  $c_0$  se refere ao tempo de sobrevivência da população que não tem câncer mas está em condições de idade e sexo similares das dos pacientes com câncer. Os parâmetros envolvidos neste modelo são  $p_0$ , que representa a proporção de pacientes curados, e  $\beta > 0$  a taxa de morte na proporção  $1 - p_0$  dos indivíduos que estão no experimento. A expressão  $e^{-\beta t}$  é a função de sobrevivência dos indivíduos não-curados e  $P_T$  é a função de sobrevivência dos indivíduos que estão no experimento.

## 1.2 Motivação

Embora o modelo de mistura padrão seja o mais estudado dos modelos com fração de cura, sua exploração não está esgotada, e apontamos alguns aspectos que analisamos neste trabalho. O primeiro caso é com relação à censura informativa (Lagakos, 1979). Na análise de sobrevivência usual, sem fração de cura, a função de verossimilhança dos parâmetros de interesse é proposta em Williams & Lagakos (1977), em Lagakos (1979) e em Lawless (1982). Mas estes três textos usam duas abordagens diferentes para determinar se a censura é informativa. Sob os resultados em Williams & Lagakos (1977) e em Lagakos (1979), existe uma estrutura de dependência, sob a qual a censura é informativa. Na abordagem de Lawless (1982), a variável de censura é informativa se sua função densidade e sua função de distribuição acumulada envolvem algum parâmetro de interesse. Também devemos evidenciar que existem outras propostas para identificar e estudar a dependência entre as variáveis de falha e de censura, das quais citamos Emoto & Matthews (1990), Zheng & Klein (1995), Link (1989) e Huang & Wolfe (2002). No entanto, é importante observar que estes textos citados estudam a

dependência da censura apenas no caso da análise de sobrevivência usual, sem fração de cura.

Por outro lado, no contexto de modelos com fração de cura, em Maller & Zhou (1996) é apresentada a matriz de informação de Fisher para o modelo de mistura padrão exponencial, em que os elementos da matriz são dados na forma de esperanças com relação à variável de censura. Um problema a destacar é que estas esperanças não têm expressões matemática fechadas, o que dificulta a implementação de algoritmos computacionais, e em consequência, no mesmo texto é proposto o uso da matriz de informação observada. Mas o detalhe que consideramos de importância central neste trabalho é que as expressões para estas esperanças dependem da distribuição da variável de censura.

Brown *et al.* (2001) apresentam resultados de estudo com dados simulados, em que é estimada a probabilidade de cobertura do intervalo da proporção de uma distribuição binomial, sob abordagens clássica e bayesiana objetiva (Berger, 2006). Esta estimativa é obtida com base em conceitos frequentistas, a partir de um grande número de réplicas (amostras) de tamanho  $n$ , e esse procedimento é repetido para um grid fino de valores da proporção, igualmente espaçados no intervalo  $(0, 1)$ . Neste trabalho, sob o modelo de mistura padrão exponencial, este procedimento é usado para verificar o impacto da censura informativa na inferência dos parâmetros de interesse.

Em sequência à terceira observação e em uma quarta perspectiva, com relação ao modelo de mistura padrão exponencial, destacamos a necessidade de estudos comparativos entre inferência bayesiana objetiva (Berger, 2006), com distribuições *a priori* de Jeffreys (Jeffreys, 1961), e inferência clássica, via análise de cobertura com dados simulados. A nossa proposta é a realização deste estudo num grid fino do espaço paramétrico da fração de cura.

Finalizando esta seção, para utilizar os conceitos frequentistas nos trabalhos com dados simulados, estabelecemos a definição de cobertura de um intervalo (de confiança, de credibilidade, etc.).

**Definição 1.1** Considere-se um modelo  $\mathcal{S}^*$  qualquer, com parâmetros  $\theta_1, \dots, \theta_p$ . Para  $j = 1, \dots, p$ , considere-se  $\Theta_j$  o espaço paramétrico marginal do parâmetro  $\theta_j$ . Dado um conjunto de dados  $\mathcal{D}^*$  sob o modelo  $\mathcal{S}^*$ , seja  $\theta_{0j} \in \Theta_j$  o valor verdadeiro de  $\theta_j$  a ser estimado. Definimos que  $\theta_{0j}$  é coberto pelo intervalo  $[a, b]$  se  $\theta_{0j} \in [a, b]$ , em que  $[a, b] \subset \Theta_j$ .

## 1.3 Objetivos

Este trabalho tem cinco objetivos, os quais listados abaixo:

- Enunciar uma definição de censura informativa, similar à de Lawless (1982), e extendê-la para o modelo unificado de longa duração proposto por Rodrigues *et al.* (2009), e em particular para o modelo de mistura padrão com tempo de falha exponencial.
- Para o modelo de mistura padrão com tempo de falha exponencial, obter uma expressão matemática para as componentes da matriz de informação de Fisher, sob as suposições de censura não-informativa e de censura informativa, a partir das esperanças apresentadas em Maller & Zhou (1996), e sob o Lema 2 de Ghitany *et al.* (1994), de tal forma a que a mesma matriz possa ser tratada computacionalmente e assim ser dispensado o uso da matriz observada.
- Sob abordagem clássica, verificar, com o uso de dados simulados e com base em conceitos frequentistas como em Brown *et al.* (2001), o impacto da censura informativa na cobertura dos intervalos de confiança. O estudo de simulação deve varrer todo o espaço paramétrico da fração de cura.
- Comparar o procedimento bayesiano objetivo, com distribuições *a priori* de Jeffreys (Jeffreys, 1961) com os procedimentos da abordagem clássica, citados no item anterior, também com base em conceitos frequentistas como em Brown *et al.* (2001).
- Aplicar os quatro itens anteriores em conjunto de dados reais.

Listamos a seguir, outros objetivos secundários do trabalho.

- Apresentar a teoria unificada dos modelos de longa duração, a partir do trabalho de Rodrigues *et al.* (2009), e caracterizar o modelo de mistura padrão como caso particular;
- Comparar, sob abordagem clássica de inferência, as probabilidades estimadas de cobertura dos intervalos,
  1. Obtidos com o uso da matriz de informação de Fisher, sob suposição de censura informativa,
  2. Obtidos com o uso da matriz de informação observada sob censura não-informativa, que é a sugestão em Maller & Zhou (1996).
- Introduzir demais propostas de trabalhos futuros.

## 1.4 Organização do Trabalho

Nos cinco primeiros capítulos desenvolvemos teorias e metodologias, e apresentamos resultados que serão analisados em estudos de simulação e aplicados a um conjunto de dados reais. Inicialmente, é importante observar que no Capítulo 2, a teorização e resultados apresentados ocorrem sob o contexto do modelo unificado proposto por Rodrigues *et al.* (2009). Adotamos este procedimento devido aos resultados importantes que não queríamos deixar de apresentar, desenvolvidos inicialmente sob o modelo unificado, e, nos capítulos seguintes, particularizados sob o modelo de mistura padrão exponencial. Ademais, numa forma geral, o desenvolvimento do texto ocorre sob o ambiente deste modelo de mistura padrão exponencial.

Neste Capítulo 2 desenvolvemos inicialmente a teoria unificada de modelos de longa duração, a partir dos trabalhos de Rodrigues *et al.* (2009), e na forma de teoremas apresentamos resultados que envolvem às funções de risco e de risco acumulado. Em seguida, já sob o contexto dos trabalhos inferenciais, apresentamos

resultados que estabelecem relações paramétricas entre as variáveis envolvidas, relações que nos motivam a definir o nosso conceito de censura informativa, que é bastante similar à abordagem apresentada em Lawless (1982). Como consequência, propomos a função de verossimilhança para o modelo unificado, que prevê as duas suposições de mecanismo de censura, não-informativa e informativa. Na forma de teoremas apresentamos resultados que envolvem às funções de risco e de risco acumulado.

No Capítulo 3 descrevemos o modelo de mistura padrão exponencial (Berkson & Gage, 1952) como caso particular do modelo unificado (Rodrigues *et al.*, 2009). Portanto, a teorização vista no Capítulo 2 é particularizada para o modelo de mistura padrão exponencial. Inicialmente escrevemos o modelo de mistura padrão, e apresentamos o modelo proposto por Berkson & Gage (1952), modelo de mistura padrão exponencial, acompanhado de uma breve revisão da literatura. Em seguida, sob o mesmo modelo de mistura padrão exponencial desenvolvemos as relações paramétricas e apresentamos a particularização do modelo geral de função de verossimilhança. No contexto da tese, ou seja, a partir deste Capítulo 3 é considerando modelo exponencial para a variável de censura.

No Capítulo 4 a inferência clássica é desenvolvida sob as duas suposições de censura, não-informativa e informativa, considerando dados observados e dados ampliados. Sob as funções de verossimilhança com dados observados (não ampliados) a matriz de informação de Fisher é obtida a partir dos resultados em Maller & Zhou (1996), e Lema 2 de Ghitany *et al.* (1994). Sob as funções com dados ampliados, apresentamos os estimadores de máxima verossimilhança via algoritmo EM (Dempster *et al.*, 1977). Finalizamos o capítulo com o desenvolvimento de um algoritmo de maximização iterativa, alternativo ao algoritmo EM e de fácil programação computacional.

No Capítulo 5 obtemos as distribuições *a priori* de Jeffreys (Jeffreys, 1961), sob as suposições de censura não-informativa e censura informativa, com dados observados e com dados ampliados. Dadas as funções de distribuição *a priori* apresentamos as respectivas funções de distribuição *a posteriori*, seguidas das

funções distribuição condicional marginal a posteriori para cada parâmetro de interesse.

Nos estudos de simulação apresentados no Capítulo 6, verificamos o impacto da censura informativa nas inferências dos parâmetros de interesse e avaliamos o procedimento bayesiano desenvolvido no Capítulo 5. Após estes estudos com dados simulados, finalizamos o capítulo com a aplicação dos resultados teóricos ao conjunto de dados de melanoma maligno E1690 (Chen *et al.*, 2002), que nos permite uma análise comparativa com os resultados dos estudos de simulação, e com resultados apresentados em Chen *et al.* (2002).

Concluimos a tese no Capítulo 7, com análises dos resultados e propostas de trabalhos futuros. Finalmente, temos três apêndices. No Apêndice A demonstramos um dos teoremas relativo à matriz de Informação de Fisher, apresentados no Capítulo 4. No Apêndice B apresentamos relações diferenciais das funções geradoras com a média da quantidade de causas, e no Apêndice C apresentamos três dos programas, desenvolvidos no código R (R Development Core Team, 2009), que foram usados no trabalho de tese.

Quanto à organização do trabalho devemos fazer ainda duas observações. A primeira é que, por uma questão de simplificação didática, consideramos o trabalho dentro do contexto de ensaios clínicos. Nestes casos, o evento de interesse pode ser, por exemplo, a morte do paciente com câncer, a reincidência do câncer neste mesmo paciente, a reincidência dos sintomas em paciente HIV-positivo ou a morte deste mesmo paciente. Portanto, as nossas unidades experimentais serão apenas seres humanos, a menos de menção em contrário. Mas, de forma similar, a teoria também poderia ser usada, por exemplo, em testes de durabilidade de componentes industriais, veja Rodrigues *et al.* (2009).

Também, devido à natureza da teoria, consideraremos que no conjunto de dados, as censuras ocorrem apenas à direita (veja primeiro parágrafo deste capítulo). Portanto, de novo por comodidade didática, quando falamos em observações censuradas estamos nos referindo somente à censuras à direita. Esta censura à direita ocorre no momento em que o paciente sai do experimento, devido à



causas conhecidas ou não (mecanismos de censura). Diz-se então que o tempo observado (o momento em que o paciente sai do experimento) é de censura. Se o evento de interesse é observado diz-se que ocorreu falha.

## Capítulo 2

# Modelo Unificado de Longa Duração

Uma teoria unificada de análise de sobrevivência de longa duração é proposta por Rodrigues *et al.* (2009), e é desenvolvida neste capítulo a partir dos resultados apresentados em Feller (1968). No Capítulo XI de Feller (1968) é definida a função geradora da seqüência  $\{a_n\}$ . O desenvolvimento da Teoria Unificada de Longa Duração é baseado nessa definição, que apresentamos na Seção 1, a seguir.

Na Seção 2.2, apresentamos na forma de teoremas, alguns resultados que envolvem as funções de risco, da população e dos indivíduos não curados. Estes resultados ocorrem dentro do contexto do modelo unificado mas fora do contexto da tese. No entanto, optamos por apresenta-los como contribuições secundárias.

Como contribuições importantes dentro do contexto desta tese, e extendidas ao modelo unificado (Rodrigues *et al.*, 2009), apresentamos três resultados relacionados à censura informativa. O primeiro é uma expressão matemática para a probabilidade de censura do indivíduo não-curado. Este primeiro resultado motiva uma nova definição de censura informativa, que apresentamos como a segunda contribuição. Estes dois primeiros resultados estão na Seção 2.4. Na Seção 2.5 enunciamos em teorema a função de verossimilhança sob censura informativa, como a terceira contribuição.

## 2.1 Função de Sobrevivência de Longa Duração

No desenvolvimento da proposta de Rodrigues *et al.* (2009) são de fundamental importância variáveis aleatórias discretas que assumem valores  $k = 0, 1, 2, \dots$ . Em Feller (1968) estas variáveis são estudadas pelo método das funções geradoras, que são um caso especial das funções características, largamente usadas na teoria assintótica. Na teoria das probabilidades as funções geradoras vêm sendo usadas desde deMoivre e Laplace, mas raramente foram exploradas em todo o seu potencial (Feller, 1968).

**Definição 2.1** (Feller, 1968). *Seja  $\{a_m\} = a_0, a_1, a_2, \dots$  uma sequência de números reais. Se*

$$A(s) = a_0 + a_1s + a_2s^2 + \dots \quad (2.1)$$

*converge com  $s$  no intervalo  $[-s_0, s_0]$ ,  $s_0 > 0$ , então  $A(s)$  é definida como a função geradora da sequência  $\{a_m\}$ .*

Observe-se que na definição acima não se faz qualquer restrição quanto aos possíveis valores que cada termo da sequência  $\{a_m\}$  pode assumir. Ou seja, cada termo de  $\{a_m\}$  pode ser qualquer valor real. Neste trabalho porém, estabelecemos na proposição a seguir que cada um destes termos será uma probabilidade, e portanto pertencerá ao intervalo fechado  $[0, 1]$ .

**Proposição 2.1** *Se  $0 \leq a_m \leq 1$  e  $|s| \leq 1$ , então  $A(s)$  é uma função geradora da sequência  $\{a_m\}$ .*

**Demonstração.** Basta considerar que a sequência  $\{a_m\}$  é limitada, e então a comparação com séries geométricas mostra que (2.1) converge quando  $|s| < 1$ . Para o caso em que  $s = 1$ , se  $0 \leq \{a_m\} \leq 1$  teremos então que  $A(s)$  é limitada e não decrescente e, portanto converge. ■

Neste trabalho teremos sempre  $0 \leq s \leq 1$ . Estaremos então trabalhando com um caso específico, uma considerável restrição da Definição 2.1, em que  $A(s)$

é sempre não decrescente e convergente se  $a_m \in [0, 1]$ . Esta última condição é essencial e nos torna possível a aplicação da Definição 2.1 no desenvolvimento deste trabalho. Na proposição a seguir consideramos uma variável aleatória discreta, que neste trabalho expressa a quantidade de causas que competem para promover a ocorrência do evento de interesse.

**Proposição 2.2** *Seja  $M$  variável aleatória discreta, com distribuição dada por*

$$P\{M = m; \boldsymbol{\theta}_M\} = p_m, \quad \sum_{m=0}^{\infty} p_m = 1, \quad p_0 > 0, \quad m = 0, 1, 2, \dots, \quad (2.2)$$

*com vetor de parâmetros  $\boldsymbol{\theta}_M$ . Se  $0 \leq s \leq 1$ , a função geradora da seqüência  $\{p_m\}$  é dada por*

$$A(s) = p_0 + p_1s + p_2s^2 + p_3s^3 + \dots \quad (2.3)$$

**Demonstração.** Pela Proposição 2.1,  $A(s)$  converge no intervalo  $0 \leq s \leq 1$ . ■

Por outro lado, como  $A(1) = 1$ , temos outra forma de verificar que a série  $A(s)$  converge absolutamente no intervalo  $0 \leq s \leq 1$ . Aqui no estudo de modelos de longa duração a variável  $s$  tem significado especial, pois é a própria função de sobrevivência do indivíduo associada à  $i$ -ésima causa.

Nesta seção introduzimos a função geradora de longa duração, que permitirá unificar os modelos de longa duração existentes em análise de sobrevivência. O evento de interesse em análise de sobrevivência é o tempo de falha do indivíduo, ou seja, o tempo decorrente entre o início do experimento ou da observação até a falha do indivíduo. Este *evento de interesse* pode ser o momento da morte de um paciente do grupo ou da reincidência de determinado tipo de câncer após a cirurgia de retirada do tumor. É natural portanto que este evento seja representado por uma v. a. contínua que representa o tempo, que aqui será denotada por  $Y$ .

É razoável supor que uma certa quantidade de causas ou riscos competem para promover a ocorrência do evento de interesse. Na área médica, podem ser consideradas como causas as células que sofrem divisão descontrolada por algum agente externo. De forma competitiva uma dessas células poderá causar o

tumor. Mais detalhes podem ser vistos em Yakovlev & Tsodikov (1996). Outros exemplos, na área econômica e industrial, são citados em Rodrigues *et al.* (2009).

Para a quantidade de causas, que é não observável, Rodrigues *et al.* (2009) atribuem, de forma geral, a distribuição (2.2), em que, com probabilidade  $p_0$  pode não haver nenhuma causa promovendo a ocorrência do evento de interesse. Neste caso, na área médica, o indivíduo é considerado curado.

Neste caso, considerando-se  $M$  com distribuição (2.2), dado  $M = m$ , sejam  $Z_j$ ,  $j = 1, \dots, m$ , variáveis aleatórias contínuas, não-negativas, independentes e identicamente distribuídas, e independentes de  $M$ , com função de distribuição acumulada e vetor de parâmetros dados por

$$F(\cdot) \text{ e } \boldsymbol{\theta}_Z, \quad (2.4)$$

respectivamente, cada uma representando o tempo de ocorrência do evento de interesse devido à  $j$ -ésima causa ou risco. O tempo de ocorrência do evento de interesse é definido pela variável aleatória  $Y$ , dada por

$$Y = \min \{Z_0, Z_1, Z_2, \dots, Z_M\}, \quad (2.5)$$

em que

$$P[Z_0 = \infty] = 1. \quad (2.6)$$

As variáveis aleatórias  $Z_0, Z_1, Z_2, \dots, Z_m$ , e  $M$  são não-observáveis enquanto que  $Y$  é uma variável observável desde que não censurada.

Se  $M = 0$  então o tempo de ocorrência do evento de interesse é afetado pelas causas da censura pois o indivíduo não sofre o estresse de nenhuma causa ou risco de falha. Na medicina considera-se que o indivíduo está curado. Esta última suposição pressupõe que uma proporção  $p_0$  da população não apresenta a ocorrência do evento de interesse. Esta é a principal característica dos modelos de sobrevivência de longa duração, veja Chen *et al.* (2002).

Dado  $M = m \geq 1$ , a função de sobrevivência de um indivíduo da população, em risco associado à  $j$ -ésima causa é dada por

$$S(y) = P[Z_j > y] = 1 - F(\cdot), \quad j = 1, \dots, m, \quad (2.7)$$

que é própria, enquanto que a função de sobrevivência de um indivíduo da população é dada por

$$S_p(y) = P[Y > y] = 1 - P[Y \leq y] = 1 - F_p(y). \quad (2.8)$$

**Teorema 2.1** (Rodrigues *et al.*, 2009). *Dada a função de sobrevivência própria em (2.7), a função de sobrevivência da variável  $Y$  em (2.8) é dada por*

$$S_p(y) = A(S(y)) = \sum_{m=0}^{\infty} p_m [S(y)]^m, \quad (2.9)$$

em que  $A(S(y))$  é a função geradora da sequência  $\{p_m\}$ , que converge se  $0 \leq S(y) \leq 1$ , e  $p_m$  é definida em (2.2).

Sob o Teorema 2.1,  $S_p(y)$  não é uma função de sobrevivência própria, como ocorre na análise de sobrevivência clássica, o que é verificado no teorema a seguir.

**Teorema 2.2** (Rodrigues *et al.*, 2009). *Dada uma função de sobrevivência própria  $S(\cdot)$ , temos que*

$$\lim_{y \rightarrow \infty} S_p(y) = p_0, \quad (2.10)$$

em que  $p_0$  é a probabilidade de não ocorrência do evento de interesse na população.

Quando  $p_0 = 0$ ,  $S_p(y)$  se torna um caso generalizado da função de sobrevivência clássica. O modelo de sobrevivência tradicional dos cursos iniciais de sobrevivência, em que é considerada uma única causa, é um caso particular deste.

**Teorema 2.3** (Rodrigues *et al.*, 2009). *A função de sobrevivência de longa duração da variável  $Y$  em (2.8) é dada por*

$$S_p(y) = p_0 + (1 - p_0) S_p^*(y), \quad (2.11)$$

em que

$$S_p^*(y) = \sum_{m=1}^{\infty} p_m^* [S(y)]^m \quad e \quad p_m^* = \frac{p_m}{1 - p_0}. \quad (2.12)$$

Sob o Teorema 2.3, o modelo (2.11) é a função de sobrevivência de longa duração. Note-se que  $S_p^*(y)$  é uma função de sobrevivência própria (Rodrigues *et al.*, 2009), associada aos elementos da população que estão em risco, com função geradora da sequência  $\{p_m^*\}$  truncada em  $p_0$  para  $m \geq 1$ . Ou seja,  $\lim_{y \rightarrow \infty} S_p^*(y) = 0$ . Sob este teorema, o modelo (2.11) reduz a função (2.9) ao modelo de mistura padrão, proposto por Boag (1949) e Berkson & Gage (1952).

Quando nos referimos ao modelo unificado de longa duração proposto por Rodrigues *et al.* (2009), sem perda de generalidade e a menos de menção em contrário, estamos nos referindo ao modelo na forma (2.9). Quanto à sua forma padronizada sob o Teorema 2.3, dado a similaridade com o modelo de mistura padrão, objeto deste trabalho, convencionamos chamar o modelo (2.11) de modelo de mistura padrão unificado. No entanto as funções (2.9) e (2.11) são versões equivalentes do modelo unificado de longa duração proposto em Rodrigues *et al.* (2009).

## 2.2 Função Sub-Densidade e Função de Risco Populacional

A função sub-densidade do tempo de ocorrência do evento de interesse é denotada por  $f_p(y)$  e é obtida derivando  $S_p(y)$  com relação a  $y$ . Temos que

$$f_p(y) = -\frac{\partial S_p(y)}{\partial y} = f(y) \left( \frac{\partial A(s)}{\partial s} \Big|_{s=S(y)} \right), \quad (2.13)$$

em que, de (2.7)

$$f(y) = -\frac{\partial S(y)}{\partial y} \quad (2.14)$$

é a função densidade (própria) do tempo de sobrevivência do indivíduo em risco devido à  $j$ -ésima causa.

Derivando (2.11) temos que a sub-densidade (2.13) pode também ser escrita na forma

$$f_p(y) = -(1 - p_0) \frac{\partial S_p^*(y)}{\partial y}. \quad (2.15)$$

A função de risco é uma ferramenta essencial à análise de dados de sobrevivência (Lawless, 1982; Colosimo & Giolo, 2006). Motivado pela definição tradicional de função de risco em análise de sobrevivência, a função de risco correspondente à função de sobrevivência de longa duração (2.9), que passamos a chamar de função de risco populacional, é dada por

$$\lambda_p(y) = \lim_{\Delta y \rightarrow 0} \frac{P[y \leq Y \leq y + \Delta y | Y \geq y]}{\Delta y}.$$

Desenvolvendo esta expressão temos que

$$\begin{aligned} \lambda_p(y) &= \lim_{\Delta y \rightarrow 0} \frac{P[y \leq Y \leq y + \Delta y]}{P[Y \geq y] \Delta y} \\ &= - \lim_{\Delta y \rightarrow 0} \frac{S_p(y + \Delta y) - S_p(y)}{S_p(y) \Delta y} = - \frac{S'_p(y)}{S_p(y)}, \end{aligned}$$

e, portanto

$$\lambda_p(y) = \frac{f_p(y)}{S_p(y)}. \quad (2.16)$$

De (2.16) temos que a expressão do risco populacional associado ao modelo (2.9) é dada por

$$\lambda_p(y) = \frac{f_p(y)}{S_p(y)} = f(y) \frac{\sum_{m=1}^{\infty} m p_m [S(y)]^{m-1}}{p_0 + \sum_{m=1}^{\infty} p_m [S(y)]^m}. \quad (2.17)$$

Também é importante em análise de sobrevivência de longa duração (Rodrigues *et al.*, 2009) as funções densidade e de risco relativas aos indivíduos não-curados, que são dadas respectivamente por

$$f^*(y) = \frac{\partial S_p^*(y)}{\partial y} \quad (2.18)$$

e

$$\lambda^*(y) = \frac{f^*(y)}{S_p^*(y)}. \quad (2.19)$$

Por outro lado, dado que o indivíduo é não-curado, ou seja, dado que  $m > 0$ , devemos considerar, inerente a este indivíduo, o risco associado à  $j$ -ésima causa,  $j = 1, \dots, m$ , que, de (2.7) e (2.14) é dado por

$$\lambda(y) = \frac{f(y)}{S(y)}. \quad (2.20)$$



De Rodrigues *et al.* (2009) temos que a função (2.17) pode ser dada também por

$$\lambda_p(y) = \lambda^*(y) \frac{(1 - p_0) S_p^*(y)}{S_p(y)}. \quad (2.21)$$

Esta expressão para o risco populacional é de fácil demonstração, e enseja uma interpretação dos mecanismos dos riscos da população e dos não-curados, que apresentamos nos dois teoremas a seguir.

**Teorema 2.4** *A função de risco da população em (2.17) pode ser dada por*

$$\lambda_p(y) = \lambda(y) E[M|Y > y], \quad (2.22)$$

em que  $\lambda(y)$ , o risco devido à  $j$ -ésima causa,  $j = 1, \dots, m$ , é dado por (2.20).

**Demonstração.** Reescrevendo a expressão (2.17) com algumas manipulações algébricas temos que

$$\begin{aligned} \lambda_p(y) &= \frac{f(y)}{S(y)} \sum_{m=0}^{\infty} m \frac{p_m [S(y)]^m}{S_p(y)} \\ &= \frac{f(y)}{S(y)} \sum_{m=0}^{\infty} m \frac{P[M = m] P[Y > y | M = m]}{P[Y > y]} \\ &= \frac{f(y)}{S(y)} \sum_{m=0}^{\infty} m \frac{P[M = m, Y > y]}{P[Y > y]} \\ &= \frac{f(y)}{S(y)} \sum_{m=0}^{\infty} m P[M = m | Y > y] \\ &= \lambda(y) E[M | Y > y]. \quad \blacksquare \end{aligned}$$

**Teorema 2.5** *A função de risco dos indivíduos não-curados em (2.19) pode ser dada por*

$$\lambda^*(y) = \lambda(y) E[M | M > 0, Y > y]. \quad (2.23)$$

**Demonstração.** De (2.19) e com manipulações algébricas similares às da demonstração anterior temos que

$$\begin{aligned}
\lambda^*(y) &= \frac{f(y)}{S(y)} \sum_{m=1}^{\infty} m \frac{p_m^* [S(y)]^m}{S_p^*(y)} \\
&= \frac{f(y)}{S(y)} \sum_{m=1}^{\infty} m \frac{P[M = m | M > 0] P[Y > y | M = m, M > 0]}{P[Y > y | M > 0]} \\
&= \frac{f(y)}{S(y)} \sum_{m=1}^{\infty} m \frac{P[M = m, Y > y | M > 0]}{P[Y > y | M > 0]} \\
&= \frac{f(y)}{S(y)} \sum_{m=0}^{\infty} m P[M = m | Y > y, M > 0] \\
&= \lambda(y) E[M | Y > y, M > 0]. \quad \blacksquare
\end{aligned}$$

Uma terceira classe de funções importantes em modelos de sobrevivência de longa duração é a do risco acumulado. Definimos então as funções relativas à população, aos indivíduos não-curados, e para estes últimos a função relativa à  $j$ -ésima causa.

**Definição 2.2** *A função de risco acumulado da população é dada por*

$$\Lambda_p(y) = \int_0^y \lambda_p(u) du. \quad (2.24)$$

**Definição 2.3** *A função de risco acumulado dos indivíduos não-curados é dada por*

$$\Lambda^*(y) = \int_0^y \lambda^*(u) du. \quad (2.25)$$

**Definição 2.4** *Dado que o indivíduo é não-curado,  $m > 0$ , a função de risco acumulado devida à  $j$ -ésima causa,  $j = 1, \dots, m$ , é dada por*

$$\Lambda(y) = \int_0^y \lambda(u) du. \quad (2.26)$$

A seguir dois teoremas importantes que caracterizam os modelos de longa duração.

**Teorema 2.6** *Se  $E[M] < \infty$  e  $p_0 > 0$ , então a função de risco acumulado  $\Lambda_p(y)$  da população é limitada, ou seja*

$$0 < \Lambda_p(y) \leq \int_0^\infty \lambda_p(y) dy < \infty. \quad (2.27)$$

**Demonstração.** Note-se que, em (2.17), se  $p_0 > 0$ , então para qualquer  $y > 0$ ,

$$0 < \frac{\sum_{m=1}^\infty mp_m [S(y)]^{m-1}}{p_0 + \sum_{m=1}^\infty p_m [S(y)]^m} \leq \frac{\sum_{m=1}^\infty mp_m [S(y)]^{m-1}}{p_0} < \frac{E[M]}{p_0}, \quad (2.28)$$

e portanto

$$\begin{aligned} 0 < \int_0^\infty \lambda_p(y) dy &= \int_0^\infty \frac{\sum_{m=1}^\infty mp_m [S(y)]^{m-1}}{p_0 + \sum_{m=1}^\infty p_m [S(y)]^m} f(y) dy \\ &= E \left[ \frac{\sum_{m=1}^\infty mp_m [S(Y)]^{m-1}}{p_0 + \sum_{m=1}^\infty p_m [S(Y)]^m} \right] < \frac{E[M]}{p_0}, \end{aligned}$$

em que esta esperança é dada sob a função densidade (2.14). Portanto, se  $E[M] < \infty$  e  $p_0 > 0$ , teremos também que

$$\lim_{y \rightarrow \infty} \Lambda_p(y) < \infty. \quad \blacksquare \quad (2.29)$$

A densidade populacional imprópria (2.15) leva em conta uma proporção  $p_0$  de curados e  $1 - p_0$  de indivíduos em risco. Seguindo a literatura de modelos de longa duração (Maller & Zhou, 1996; Rodrigues *et al.*, 2008; Chen *et al.*, 2002) chamamos  $p_0$  de fração de cura. O termo 'fração de cura' é uma expressão genérica, usada para caracterizar modelos de sobrevivência que têm função de risco acumulado limitada (Chen *et al.*, 2002). De fato, sob o modelo unificado, se  $E[M] < \infty$ , por (2.29) o risco acumulado da população é limitado.

**Teorema 2.7** *A função de risco acumulado dos indivíduos não-curados é não limitada, ou seja*

$$\lim_{y \rightarrow \infty} \Lambda^*(y) = \infty. \quad (2.30)$$

**Demonstração.** Temos que  $0 < S_p^*(y) \leq 1$ , o que implica  $1 \leq \frac{1}{S_p^*(y)} < \infty$ . Portanto,

$$\begin{aligned} \lim_{y \rightarrow \infty} \Lambda^*(y) &= \int_0^\infty \lambda^*(y) dy = \int_0^\infty \frac{f^*(y) dy}{S_p^*(y)} = \int_0^\infty \frac{-dS_p^*(y)}{S_p^*(y)} \\ &= - \int_0^\infty \frac{dS_p^*(y)}{S_p^*(y)} = - \log S_p^*(y) \Big|_0^\infty \\ &= - \lim_{y \rightarrow \infty} [\log S_p^*(y) - \log S_p^*(0)] = \infty, \end{aligned}$$

em que  $dS_p^*(y) = \frac{\partial S_p^*(y)}{\partial y} dy = -f^*(y) dy$ . ■

De forma similar prova-se que em um indivíduo não-curado, o risco acumulado devido à  $j$ -ésima causa (2.26) também é não limitado, ou seja,

$$\lim_{y \rightarrow \infty} \Lambda(y) = \infty. \quad (2.31)$$

O Teorema 2.6 caracteriza os modelos de sobrevivência de longa duração, veja Chen *et al.* (2002), enquanto que o Teorema 2.7 caracteriza os modelos de sobrevivência usual e os indivíduos não curados.

Devido ao Teorema 2.6 temos que

$$\lim_{y \rightarrow \infty} \lambda_p(y) = 0, \quad (2.32)$$

e isso implica em (2.22) que, devido à (2.31),

$$\lim_{y \rightarrow \infty} E[M|Y > y] = 0. \quad (2.33)$$

Também devido ao Teorema 2.6, em (2.23) temos que

$$\lim_{y \rightarrow \infty} E[M|Y > y, M > 0] = 1. \quad (2.34)$$

Estes dois resultados sugerem que, quanto menor o número de fatores de risco promovendo o evento de interesse maior é a sobrevivência do indivíduo, culminando com o caso em que  $m = 0$  e  $P[Y = \infty] = 1$ , e no caso dos não-curados, culmina com  $m = 1$ .

Os teoremas 2.4 e 2.5 permitem ver a função de risco, da população e dos indivíduos não-curados, como uma soma média, no momento  $y$ , dos riscos devidos à  $j$ -ésima causa,  $j = 1, \dots, m$ . Isso nos motiva a conjecturar sobre a possível

existência de classes de estruturas de dependência que potencializam os riscos, da população e dos indivíduos não curados.

Note-se também que se  $M = 0$ ,  $P[Y = \infty] = 1$ , e  $\lambda_p(y) = 0$  para qualquer  $y \geq 0$ . Em geral para  $M$  com distribuição (2.2) temos que

$$\lim_{y \rightarrow 0} \lambda_p(y) = f(0) E(M).$$

Portanto,  $\lambda_p(y)$  é monótona decrescente e

$$0 < \lambda_p(y) < f(0) E(M). \quad (2.35)$$

O modelo (2.9) implica que uma proporção  $p_0$  de indivíduos não apresentará o evento de interesse (indivíduos curados) e que a outra proporção  $1 - p_0$  continuam em risco. Teoricamente, nos indivíduos curados não se tem nenhuma causa promovendo o evento de interesse enquanto que nos não-curados permanece uma quantidade  $M \geq 1$  de causas promovendo o mesmo.

## 2.3 Heterogeneidade Entre os Indivíduos e Propriedade de Riscos Proporcionais

Nos cursos de sobrevivência iniciais costuma-se considerar a variação dos riscos em (2.16) e (2.19) apenas em função do tempo, veja por exemplo Chen *et al.* (1999), mas na literatura vemos formas de inclusão de covariáveis na sua função, que determinam a heterogeneidade entre os indivíduos. A forma mais tradicional de inclusão de covariáveis na função de risco é o modelo de regressão de Cox (1972). Nesta tese, devido aos nossos objetivos, não trabalharemos com covariáveis na função de risco. Consideraremos a homogeneidade entre os indivíduos.

Uma propriedade importante em análise de sobrevivência é a propriedade de riscos proporcionais (PRP), que definimos a seguir, e uma de suas vantagens é com relação à redução do custo computacional.

**Definição 2.5** (Rodrigues et al., 2008). Uma função de risco  $\lambda(y)$  tem a propriedade de riscos proporcionais (PRP) se para duas unidades observacionais distintas ( $i_1 \neq i_2$ ), temos que  $\lambda_{i_1}(y) / \lambda_{i_2}(y)$  é independente de  $y$ .

**Teorema 2.8** (Rodrigues et al., 2009) A função de risco populacional (2.16) tem a propriedade de riscos proporcionais (PRP) se, e somente se,  $p_m = \frac{\theta^m e^{-\theta}}{m!}$ , ou seja,  $M \sim \text{Poisson}(\theta)$ .

## 2.4 Relações Paramétricas

Nestas duas últimas seções deste capítulo trabalhamos já sob o contexto de inferência. Isso requer a apresentação prévia das cinco variáveis envolvidas, que passamos a apresentar. Sejam  $M_i$ ,  $i = 1, \dots, n$ , variáveis aleatórias discretas não observáveis e identicamente distribuídas sob o modelo (2.2), definidas como

$$M_i = \begin{cases} 0, & \text{se o } i\text{-ésimo indivíduo não está em risco,} \\ 1, 2, \dots, & \text{se o } i\text{-ésimo indivíduo está em risco.} \end{cases} \quad (2.36)$$

Se o  $i$ -ésimo indivíduo está em risco, ou seja, se  $M_i \geq 1$ , então, de (2.18), o seu tempo de falha  $Y_i | M_i \geq 1$  tem função densidade dada por

$$f^*(y_i) = f^*(y_i; \boldsymbol{\theta}_{Y|M \geq 1}), \quad (2.37)$$

com vetor de parâmetros  $\boldsymbol{\theta}_{Y|M \geq 1}$ , em que, de (2.5),

$$Y_i | m_i = \min \{Z_1, \dots, Z_{m_i}\}, \quad m_i \geq 1, \quad i = 1, \dots, n. \quad (2.38)$$

Por outro lado, se o  $i$ -ésimo indivíduo não está em risco, de (2.6), temos que

$$P[Y_i = \infty | M_i = 0] = 1. \quad (2.39)$$

As duas variáveis,  $M$  e  $Y$ , são não observáveis. Uma terceira, também não observável, é a variável de censura, que denotamos por  $X$ ,  $X_i$  independente de  $Y_i$ ,

$i = 1, \dots, n$ , à qual está sujeito o  $i$ -ésimo indivíduo, curado ou não, com função densidade

$$g(x_i) = g(x_i; \boldsymbol{\theta}_X), \quad (2.40)$$

com vetor de parâmetros  $\boldsymbol{\theta}_X$ , e com função de distribuição acumulada dada por

$$G_{X_i}(x_i) = \int_0^{x_i} g(x_i; \boldsymbol{\theta}_X) dx_i = 1 - S_{X_i}(x_i), \quad i = 1, \dots, n. \quad (2.41)$$

Os dois teoremas anteriores estabelecem as relações paramétricas entre as variáveis  $M$  e  $Y$ . O teorema a seguir estabelece as relações paramétricas entre  $Y$  e  $X$ , que se traduz no conceito de censura informativa usado neste trabalho, que é definida em seguida. Definimos então duas variáveis observáveis que compõem o conjunto de dados usado nas inferências dos parâmetros de interesse e no desenvolvimento da teoria sobre censura informativa.

Para cada indivíduo  $i$  da amostra, de forma exclusiva, apenas uma das variáveis,  $Y_i$  e  $X_i$ , é observada, a que tiver o menor valor, e portanto, definimos uma variável aleatória  $T_i$ , que nos dá o tempo de sobrevivência observado no momento  $t_i$ , de falha ou censura, que é dada por

$$T_i = \min \{Y_i, X_i\} \quad (2.42)$$

Por último, definimos a variável discreta  $\Delta_i$ , indicadora de censura, dada por

$$\Delta_i = \begin{cases} 1, & \text{se } Y_i \leq X_i, \\ 0, & \text{caso contrário,} \end{cases} \quad (2.43)$$

com distribuição dada por

$$P[\Delta_i = \delta_i] = (1 - p_c)^{\delta_i} p_c^{1-\delta_i}, \quad 0 \leq p_c \leq 1, \quad \delta_i = 0, 1, \quad (2.44)$$

em que  $p_c = P[\Delta_i = 0]$ , é a probabilidade do  $i$ -ésimo indivíduo ser censurado.

Como estabelecemos no final do Capítulo 1 (penúltimo parágrafo), no nosso contexto, o experimento estatístico consiste em observar o tempo durante de vida do paciente sob determinado tratamento, com censura ou não. A seguir, para facilitar a compreensão dos próximos resultados, apresentamos, de forma objetiva, as cinco variáveis definidas anteriormente.

- $M$ : Não observável, define se o indivíduo é curado ou continua em risco.
- $Y$ : Define o tempo de vida do paciente, que pode ser observada ou não.
- $X$ : Define o tempo de censura do paciente.
- $T$ : O tempo de vida do paciente quando este é não censurado ( $Y$ ) ou o seu tempo de censura ( $X$ ).
- $\Delta$ : Variável indicadora de censura.

Na análise de sobrevivência tradicional (Kalbfleisch & Prentice, 2002; Lawless, 1982; Colosimo & Giolo, 2006) é comum a ocorrência de dados censurados, que configuram a perda de informação com relação aos parâmetros de interesse. Nos modelos de sobrevivência de longa duração parte destes dados censurados pode ser decorrente do fato de ter uma proporção (desconhecida a ser estimada) de indivíduos curados.

Sob a distribuição (2.2), a probabilidade  $p_c$  do  $i$ -ésimo indivíduo ser censurado é dada por

$$\begin{aligned} p_c &= P[\Delta_i = 0] \\ &= P[\Delta_i = 0 | M_i = 0] P[M_i = 0] + P[\Delta_i = 0 | M_i \geq 1] P[M_i \geq 1], \end{aligned} \quad (2.45)$$

em que, por (2.39), (2.40) e (2.43),  $P[\Delta_i = 0 | M_i = 0] = 1$ . Portanto, de (2.2),

$$p_c = p_0 + (1 - p_0) k^*, \quad (2.46)$$

em que

$$P[\Delta_i = 0 | M_i \geq 1] = k^*. \quad (2.47)$$

De (2.44) e (2.46) temos que  $0 \leq p_0 \leq p_c \leq 1$ .



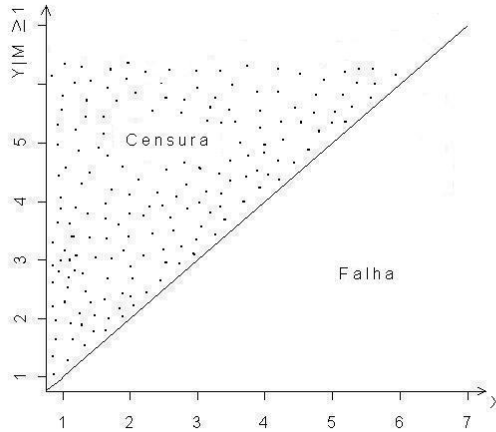


Figura 2.1. Representação da região de censura dos indivíduos em risco. No eixo das ordenadas temos a variável  $Y|M \geq 1$ .

**Teorema 2.9** Dado  $p_c$  fixo, desconhecido,

$$k^* = P[\Delta_i = 0 | M_i \geq 1] = E_X [S_p^*(X_i)]. \quad (2.48)$$

**Demonstração.** Observando a Figura 2.1, vemos que o valor de  $k^*$  é obtido integrando a função densidade conjunta de  $Y_i | M_i \geq 1$  e  $X_i$  na região hachurada. Temos, por (2.12), (2.37) e (2.40), que

$$\begin{aligned} k^* = P[\Delta_i = 0 | M_i \geq 1] &= \int_0^\infty \left[ \int_{x_i}^\infty f^*(y_i) dy_i \right] g(x_i) dx_i \\ &= \int_0^\infty P[Y_i > x_i | M_i \geq 1] g(x_i) dx_i \\ &= \int_0^\infty S_p^*(x_i) g(x_i) dx_i \\ &= E_X [S_p^*(X_i)]. \quad \blacksquare \end{aligned}$$

A expressão em (2.48) é uma esperança em  $X$  da função  $S_p^*(X_i)$ , que envolve o vetor de parâmetros da variável condicional  $Y_i | M_i \geq 1$ . Este último teorema motiva a definição de censura informativa que é utilizada neste trabalho. Existem várias abordagens sobre censura informativa, como por exemplo em Lagakos (1979), Lawless (1982) e Emoto & Matthews (1990). Nenhuma delas relaciona

em  $P[\Delta_i = 0 | M_i \geq 1]$ , os vetores de parâmetros das respectivas distribuições das variáveis de falha e de censura.

**Definição 2.6** *Sob o modelo unificado em (2.11), a censura é informativa se a probabilidade de censura de um indivíduo em risco depende funcionalmente do vetor de parâmetros da distribuição do tempo de ocorrência do evento de interesse e do vetor de parâmetros da distribuição da variável de censura.*

Uma consequência importante do Teorema 2.9 é que a censura não pode ser ignorada na função de verossimilhança, como normalmente ocorre na análise de sobrevivência de longa duração. Este, e o teorema seguinte, estão entre as principais contribuições teóricas desta tese. Baseado neste princípio de censura informativa é importante determinar a função de verossimilhança correspondente, que é introduzida no próximo teorema.

Para finalizar esta seção, devemos fazer uma última observação sobre o mecanismo de censura. De (2.36), se  $M_i = 0$ , ou seja, se o  $i$ -ésimo indivíduo não está em risco, das equações (2.5) e (2.6), e sob (2.40), será censurado com probabilidade 1. Caso contrário, se  $M_i \geq 1$ , pode ser censurado com probabilidade  $k^*$ ,  $0 \leq k^* \leq 1$ . Portanto, no modelo (2.11), dado  $p_c$ , existem dois mecanismos de censura atuando simultaneamente, representados pelas variáveis

$$\begin{cases} M_i, \text{ definida em (2.36), com distribuição em (2.2),} \\ C_i^*, \text{ com distribuição } P[C_i^* = c_i^*] = (1 - k^*)^{c_i^*} (k^*)^{1-c_i^*}, c_i^* = 0, 1. \end{cases} \quad (2.49)$$

Note-se que  $M_i$  indica se o indivíduo está ou não em risco, enquanto que  $C_i$  indica se o indivíduo em risco é censurado ou não. Se  $p_0 \downarrow 0$  então  $C_i^*$  tende em distribuição para  $\Delta_i$ , e se  $p_0 \uparrow p_c$  temos que  $M_i$  tende em distribuição para  $\Delta_i$ .

## 2.5 Função de Verossimilhança

Devemos desenvolver esta seção observando a Figura 2.1. No teorema a seguir apresentamos um modelo de função de verossimilhança para o modelo unificado

(2.11) de longa duração, que contempla as suposições de censura não-informativa e de censura informativa, cujo enunciado e demonstração requer a definição prévia das funções, vizinhanças e probabilidades listadas abaixo.

- Funções  $F_C(t_i)$  e  $F_D(t_i)$ , dadas respectivamente por

$$F_C(t_i) = P[T_i \leq t_i | \Delta_i = 0] \quad \text{e} \quad F_D(t_i) = P[T_i \leq t_i | \Delta_i = 1]. \quad (2.50)$$

- $N_{t_i} = (t_i - dt_i, t_i)$ , vizinhança  $N_{t_i}$  de  $t_i$ .
- $C_{t_i}$ : ocorrência de censura do  $i$ -ésimo indivíduo na vizinhança  $N_{t_i}$  de  $t_i$ .
- $D_{t_i}$ : ocorrência de falha do  $i$ -ésimo indivíduo na vizinhança  $N_{t_i}$  de  $t_i$ .
- De (2.50),

$$P[C_{t_i}] = P[T_i \in N_{t_i}, \Delta_i = 0] = p_c dF_C(t_i),$$

$$P[D_{t_i}] = P[T_i \in N_{t_i}, \Delta_i = 1] = (1 - p_c) dF_D(t_i),$$

em que  $dF_C(t_i) = F'_C(t_i) dt_i$  e  $dF_D(t_i) = F'_D(t_i) dt_i$ .

- Funções  $a$  e  $B$ , expressas por

$$dB(t_i) = p_c \frac{dF_C(t_i)}{[1 - F_p(t_i)]} \quad \text{e} \quad a(t_i) = (1 - p_c) \frac{dF_D(t_i)}{dF_p(t_i)}, \quad (2.51)$$

em que  $F_p(\cdot)$  é definida em (2.8),

$$dF_p(t_i) = F'_p(t_i) dt_i \quad \text{e} \quad B(t_i) = \int_0^{t_i} dB(u_i). \quad (2.52)$$

- Note-se que  $dB(t_i) = P[C_{t_i} | Y_i > t_i]$  e  $a(t_i) = P[\Delta_i = 1 | Y_i \in N_{t_i}]$ , e então

$$P[C_{t_i}] = [1 - F_p(t_i)] dB(t_i), \quad \text{e} \quad P[D_{t_i}] = a(t_i) dF_p(t_i). \quad (2.53)$$

O modelo  $(F_p, a, B)$  é objeto do teorema que enunciamos.

**Teorema 2.10** *Sob o modelo unificado (2.11), considere-se  $Y$  com função subdensidade (2.15), e  $X$ , variável de censura com função de distribuição acumulada  $G_X(\cdot)$ , definida em (2.41),  $Y$  independente de  $X$ . Sob uma amostra*

de  $n$  indivíduos, considere-se também  $T_i$  e  $\Delta_i$  definidas em (2.42) e (2.43), respectivamente. Dadas as realizações independentes  $(t_i, \delta_i)$  de  $(T_i, \Delta_i)$ , a função de verossimilhança dos vetores  $\boldsymbol{\theta}_Y$  e  $\boldsymbol{\theta}_X$  é dada por

$$\mathcal{L}(F_p, a, B) \propto \prod_{i=1}^n \{a(t_i) dF_p(t_i)\}^{\delta_i} \{[1 - F_p(t_i)] dB(t_i)\}^{1-\delta_i}, \quad (2.54)$$

em que  $a(t_i) = 1 - G_{X_i}(t_i)$ ,  $B(t_i) = G_{X_i}(t_i)$ , e os vetores  $\boldsymbol{\theta}_Y$  e  $\boldsymbol{\theta}_X$  são definidos (2.37) e em (2.40), respectivamente.

**Demonstração.** Dado que o  $i$ -ésimo indivíduo sofre falha ou censura, por (2.53) devemos verificar inicialmente que o modelo  $(F_p, a, B)$  satisfaz a restrição

$$\int_0^{\infty} \{a(t_i) dF_p(t_i) + [1 - F_p(t_i)] dB(t_i)\} = 1. \quad (2.55)$$

Pelo método de integração por partes temos que

$$\begin{aligned} \int_0^{\infty} \{a(t_i) dF_p(t_i) + [1 - F_p(t_i)] dB(t_i)\} &= \int_0^{\infty} a(t_i) dF_p(t_i) \\ &\quad + \int_0^{\infty} [1 - F_p(t_i)] dB(t_i) \\ &= \int_0^{\infty} a(t_i) dF_p(t_i) + \int_0^{\infty} dB(t_i) - \int_0^{\infty} F_p(t_i) dB(t_i) \\ &= \int_0^{\infty} a(t_i) dF_p(t_i) - \left[ [F_p(t_i) B(t_i)]_0^{\infty} - \int_0^{\infty} B(t_i) dF_p(t_i) \right] \\ &\quad + \int_0^{\infty} dB(t_i) \\ &= \int_0^{\infty} a(t_i) dF_p(t_i) + \int_0^{\infty} B(t_i) dF_p(t_i) - (1 - p_0) + \int_0^{\infty} dB(t_i) \\ &= \int_0^{\infty} [a(t_i) + B(t_i)] dF_p(t_i) + p_0. \end{aligned}$$

Portanto, a restrição em (2.55) é equivalente à

$$\int_0^{\infty} [a(t_i) + B(t_i)] dF_p(t_i) = 1 - p_0. \quad (2.56)$$

Para que (2.56) seja válida é suficiente que  $a(t_i) + B(t_i) = 1$ . Desta última igualdade, para obter  $a(t_i)$  é suficiente determinar  $B(t_i)$ . Calculando a função

de distribuição acumulada de  $T_i|\Delta_i = 0$ , temos

$$\begin{aligned}
 F_C(t_i) &= \frac{P[T_i \leq t_i, \Delta_i = 0]}{P[\Delta_i = 0]} \\
 &= \frac{P[T_i \leq t_i | \Delta_i = 0, M_i = 0] P[\Delta_i = 0 | M_i = 0] P[M_i = 0]}{P[\Delta_i = 0]} \\
 &\quad + \frac{P[T_i \leq t_i, \Delta_i = 0 | M_i \geq 1] P[M_i \geq 1]}{P[\Delta_i = 0]} \\
 &= \frac{p_0 G_X(t_i) + (1 - p_0) \int_0^{t_i} \left[ \int_{x_i}^{\infty} f^*(y_i) dy_i \right] g(x_i) dx_i}{P[\Delta_i = 0]} \\
 &= \frac{p_0 G_X(t_i) + (1 - p_0) \int_0^{t_i} S_p^*(x_i) g(x_i) dx_i}{P[\Delta_i = 0]},
 \end{aligned}$$

em que  $P[\Delta_i = 0]$  é dado por (2.46). A diferencial da expressão acima é dada por

$$\begin{aligned}
 dF_C(t_i) &= \frac{p_0 g(t_i) + (1 - p_0) S_p^*(t_i) g(t_i)}{P[\Delta_i = 0]} dt_i \\
 &= \frac{g(t_i) [p_0 + (1 - p_0) S_p^*(t_i)]}{P[\Delta_i = 0]} dt_i \\
 &= \frac{g(t_i) S_p(t_i)}{P[\Delta_i = 0]} dt_i. \tag{2.57}
 \end{aligned}$$

Portanto, de (2.51) e (2.52) segue que

$$dB(t_i) = g(t_i) dt_i, \quad B(t_i) = G_X(t_i), \quad a(t_i) = 1 - G_X(t_i). \quad \blacksquare \tag{2.58}$$

Este modelo de função de verossimilhança é a mesmo proposto em Williams & Lagakos (1977) e em Lagakos (1979), de cujos textos emprestamos as mesmas notações. Um detalhe que devemos evidenciar nestes dois textos é que a abordagem para censura informativa é diferente da vista em Lawless (1982) e em Huang & Wolfe (2002) (duas abordagens similares), e em outros textos como, por exemplo, Emoto & Matthews (1990) e Zheng & Klein (1995). Mas dentro dos objetivos desta tese, nos limitamos estabelecer a diferença entre a abordagem em (Williams & Lagakos, 1977; Lagakos, 1979) e a de Lawless (1982), esta última similar à nossa.

Na abordagem em Williams & Lagakos (1977) e em Lagakos (1979), a censura é informativa se, e somente se,  $a(t_i) + B(t_i)$  é uma variável aleatória não degenerada em 1, com  $E_{Y|M \geq 1} [a(t_i) + B(t_i)] = 1$ . Lawless (1982) considera que a censura

é informativa se a distribuição desta contém pelo menos um dos parâmetros da distribuição da variável de falha.

Outro detalhe que devemos considerar é que os textos citados nos dois parágrafos anteriores abordam a censura informativa sob o contexto da sobrevivência usual, sem fração de cura. No teorema acima extendemos um conceito de censura informativa, que é similar ao de Lawless (1982), para o modelo unificado de longa duração proposto em Rodrigues *et al.* (2009).

**Corolário 2.1** *Sob suposição de censura não-informativa a expressão (2.54) é dada por*

$$\mathcal{L}(F_p, a, B) \propto \prod_{i=1}^n \{dF_p(t_i)\}^{\delta_i} \{1 - F_p(t_i)\}^{1-\delta_i}. \quad (2.59)$$

Esta função é similar à função de verossimilhança apresentada em de Castro *et al.* (2010).

## 2.6 Resumo dos Principais resultados do Capítulo

Apresentamos a seguir os resultados que são usados como referência nos capítulos seguintes. As contribuições da Tese são citadas no item. Caso contrário, tratam-se de resultados já publicados, como é o caso dos Teoremas 2.1 e 2.3 (Rodrigues *et al.*, 2009). Alguns resultados são importantes como base em trabalhos futuros e, embora sejam contribuições da Tese, não são referência dentro do contexto deste trabalho. É o caso, por exemplo, do Teorema 2.4 e do Teorema 2.5. Reiteramos então que só listamos os resultados base do desenvolvimento dos outros capítulos. Entretanto, no texto da Tese são citados resultados não listados abaixo, como é caso, por exemplo, da expressão em (2.47).

- Distribuição de  $M$  em (2.2).
- Teorema 2.1.

- Teorema 2.3.
- Expressão em (2.46).
- Teorema 2.48, contribuição da Tese.
- Definição 2.6, contribuição da Tese.
- Mecanismo de censura em (2.49), contribuição da Tese.
- Função de verossimilhança, Teorema 2.10 e Corolário 2.1, contribuições da Tese.

# Capítulo 3

## Modelo de Mistura Padrão Exponencial

Numa forma ilustrativa do desenvolvimento teórico apresentado no capítulo anterior, apresentamos e descrevemos o modelo de mistura padrão exponencial acompanhado de breve revisão da literatura, e desenvolvemos as relações paramétricas, finalizando com a apresentação da função de verossimilhança e da distribuição da variável  $T$ . A primeira seção começa com a introdução do modelo de mistura padrão, introduzido por Boag (1949) e Berkson & Gage (1952). Neste, o tempo de falha dos indivíduos em risco tem um distribuição genérica, enquanto que no modelo de mistura padrão exponencial e tempo de falha segue distribuição exponencial.

### 3.1 Modelo de Mistura Padrão Exponencial

No modelo de mistura padrão é considerada uma só causa promovendo o evento de interesse, isto é, a variável  $M$  tem distribuição de Bernoulli. Note-se que a distribuição de Bernoulli vale para a variável aleatória  $M$  com distribuição (2.2), dada por

$$P(M = m; p_0) = (1 - p_0)^m p_0^{1-m}, \quad m = 0, 1, \quad (3.1)$$



em que  $0 \leq p_0 \leq 1$ .

Com  $M$  sob esta distribuição temos que  $Y$ , definida em (2.5), é dada por

$$Y = \min \{Z_0, Z_1\}, \text{ em que } M = 0, 1. \quad (3.2)$$

No modelo proposto por Berkson & Gage (1952), modelo de mistura padrão exponencial, descrito em (1.2), foi dada ênfase aos estudos clínicos de pacientes com câncer, em que é pressuposto que a morte dos indivíduos segue uma distribuição exponencial com taxa de falha  $\beta$ . Um detalhe importante no trabalho de Berkson & Gage (1952) é a comparação com a sobrevivência da população sem a doença, em condições de idade, sexo e outras variáveis similares às dos pacientes com câncer.

A literatura atual considera a versão

$$S_p(y) = p_0 + (1 - p_0) S(y), \quad (3.3)$$

em que  $S(y)$  provém de uma função densidade conhecida para  $Y > 0$ .

Observe-se que (3.1) é um caso particular de (2.2). De (2.9) segue-se que a função geradora de probabilidades da sequência (3.1) aplicada à função de sobrevivência (2.7) é dada por (3.3), o que permite considerar a restrição (3.3) do modelo de mistura padrão no contexto da teoria unificada de longa duração proposta por Rodrigues *et al.* (2009). Por outro lado, as expressões (2.11) e (2.12), do Teorema 2.3, mostram que também todo modelo da forma (2.9) pode ser reduzido ao modelo de mistura padrão na versão (1.1).

Definimos o modelo (1.1) como um modelo geral de longa duração, em que os riscos associados à  $\mathcal{S}(t)$  podem ser dependentes ou independentes, ou ainda, não identicamente distribuídos. Desta forma, as funções (2.9), (2.11) e (3.3) são casos particulares de (1.1).

Como caso particular de (3.3), o modelo objeto deste trabalho é dado por

$$S_p(y) = p_0 + (1 - p_0) e^{-\lambda y}, \quad \lambda > 0, \quad (3.4)$$

em que, de (2.11), (2.12) e (2.7), a função de sobrevivência do indivíduo em risco é dada por

$$S(y) = e^{-\lambda y} \Leftrightarrow S_p^*(y) = S(y). \quad (3.5)$$

Dado que (3.4) é o objeto deste trabalho, devemos definir as demais funções associadas. Por (2.15) e (3.5) a função sub-densidade da população, associada a este modelo é dada por

$$f_p(y) = (1 - p_0) \lambda e^{-\lambda y}, \quad (3.6)$$

em que, de (2.18) e (2.14),  $f(y)$  é dada por

$$f(y) = \lambda e^{-\lambda y} \Leftrightarrow f^*(y) = f(y), \quad (3.7)$$

e  $p_0$  é a proporção de curados da população, ou, dito de outra forma, é a probabilidade de um indivíduo da população estar curado.

Por (2.21) temos que a função de risco populacional  $\lambda_p(y)$  é dada por

$$\begin{aligned} \lambda_p(y) &= \frac{f(y)}{S(y)} \times \frac{(1 - p_0) S(y)}{p_0 + (1 - p_0) S(y)} \\ &= \lambda \frac{(1 - p_0) e^{-\lambda y}}{p_0 + (1 - p_0) e^{-\lambda y}}, \end{aligned} \quad (3.8)$$

em que

$$\lambda^*(y) = \lambda(y) = \lambda, \quad (3.9)$$

e, portanto  $\lambda_p(y)$  confirma o Teorema 2.8, ou seja, não possui a propriedade de riscos proporcionais (PRP).

Do Teorema 2.4 vê-se facilmente que, em (3.8)

$$E[M|Y > y] = \frac{(1 - p_0) e^{-\lambda y}}{p_0 + (1 - p_0) e^{-\lambda y}}, \quad (3.10)$$

e isso confirma os resultados (2.32) e (2.33). Do mesmo modo, do Teorema 2.5, em (3.9) temos obviamente que  $E[M|Y > y, M > 0] = 1$ , para todo  $y \geq 0$ , o que confirma (2.34).

As expressões (3.4), (3.6) e (3.8) dão informações a respeito de uma população de doentes que foram submetidos a tratamento. Ao longo do tratamento parte destes doentes foi curada. As expressões (3.5) e (3.7) dão informações a respeito dos pacientes não curados, e que portanto vão morrer com probabilidade 1.

O modelo proposto por Boag (1949) e Berkson & Gage (1952), conhecido na literatura como modelo de mistura padrão, é intensivamente discutido por vários autores, dos quais citamos Ewell & Ibrahim (1997), Farewell (1982), Farewell (1986), Goldman (1984), Gray & Tsiatis (1989), Greenhouse & Wolfe (1984), Halpern & Brown (1987a,b), Kuk & Chen (1992), Laska & Meisner (1992), Sposto *et al.* (1992), Stangl & Greenhouse (1998), Taylor (1995) e Yamaguchi (1992). Por último, recentemente, em Rodrigues *et al.* (2009) temos o mesmo modelo visto dentro de um contexto unificado (veja capítulo anterior).

## 3.2 Relações Paramétricas

Cosiderando então amostras de tamanho  $n$ , sob (3.4), e que as unidades amostrais sejam independentes, definimos para o  $i$ -ésimo indivíduo da amostra,  $i = 1, \dots, n$ , as cinco variáveis envolvidas. Sejam  $M_i$ , identicamente distribuídas sob (3.1), definidas como

$$M_i = \begin{cases} 0, & \text{se o } i\text{-ésimo indivíduo não está em risco após o tratamento,} \\ 1, & \text{se o } i\text{-ésimo indivíduo continua em risco após o tratamento.} \end{cases} \quad (3.11)$$

Portanto, neste trabalho teremos  $M \sim \text{Bernoulli}(1 - p_0)$ .

Se o  $i$ -ésimo indivíduo está em risco,  $M_i = 1$ , o seu tempo de falha  $Y_i$  em (3.2) tem função densidade dada por (3.7), isto é,

$$f(y_i; \lambda) = f(y_i; \lambda, M_i = 1) = \lambda e^{-\lambda y_i}, \quad \lambda > 0, \quad (3.12)$$

e se o mesmo não está mais em risco,

$$P[Y_i = \infty | M_i = 0] = 1. \quad (3.13)$$

Sob (3.12),  $\theta_{Y|M=1} = \lambda$ ,  $\theta_{Y|M=1}$  obtida de (2.37).

Atribuimos para a variável aleatória de censura  $X_i$  a função densidade

$$g(x_i|\lambda_c) = \lambda_c e^{-\lambda_c x_i}, \quad \lambda_c > 0, \quad (3.14)$$

com  $Y$  e  $X$  independentes. Sob (3.14),  $\theta_X = \lambda_c$ ,  $\theta_X$  obtida de (2.40). Supomos independência entre  $Y$  e  $X$ .

Sob  $M_i$ , definida em (3.11), a probabilidade  $p_c$  do  $i$ -ésimo indivíduo ser censurado é dada por

$$\begin{aligned} p_c &= P[\Delta_i = 0] \\ &= P[\Delta_i = 0|M_i = 0] P[M_i = 0] + P[\Delta_i = 0|M_i = 1] P[M_i = 1], \end{aligned} \quad (3.15)$$

em que, por (3.13), (3.14) e (2.43),  $P[\Delta_i = 0|M_i = 0] = 1$ . Portanto, de (3.1),

$$p_c = p_0 + (1 - p_0) k, \quad (3.16)$$

em que

$$P[\Delta_i = 0|M_i = 1] = k. \quad (3.17)$$

De (2.44) e (3.16) temos que  $0 \leq p_0 \leq p_c \leq 1$ . O teorema a seguir relaciona os parâmetros  $\lambda$  e  $\lambda_c$ .

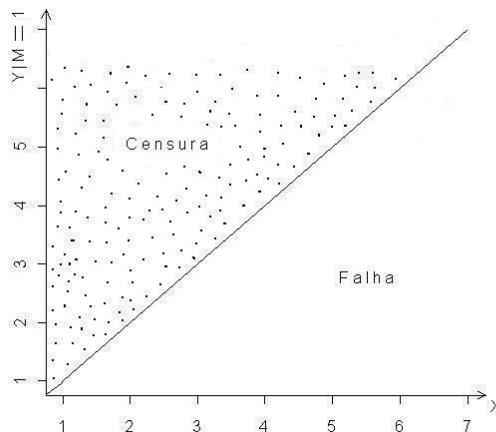


Figura 3.1. Representação da região de censura dos indivíduos em risco. No eixo das ordenadas temos a variável  $Y|M = 1$ .

**Teorema 3.1** Dado  $p_c$ , fixo, desconhecido,

$$k = P[\Delta_i = 0 | M_i = 1] = \frac{\lambda_c}{\lambda + \lambda_c}. \quad (3.18)$$

**Demonstração.** Temos que

$$\begin{aligned} k &= P[\Delta_i = 0 | M_i = 1] = E_X [P(Y_i > x_i) | X_i = x_i] \\ &= \int_0^\infty e^{-\lambda x_i} \lambda_c e^{-\lambda_c x_i} dx_i = \frac{\lambda_c}{\lambda + \lambda_c}. \quad \blacksquare \end{aligned}$$

Substituindo (3.18) em (3.16) temos que

$$\frac{\lambda_c}{\lambda + \lambda_c} = \frac{p_c - p_0}{1 - p_0}, \quad (3.19)$$

e então

$$\lambda_c = \lambda \frac{p_c - p_0}{1 - p_c}. \quad (3.20)$$

**Teorema 3.2** Dado  $p_c$ , fixo, desconhecido,

$$\lambda_c = \lambda \frac{k}{1 - k} = \lambda \eta, \quad (3.21)$$

em que

$$\eta = \frac{k}{1 - k}. \quad (3.22)$$

**Demonstração.** De (3.20) temos que provar que

$$\frac{k}{1 - k} = \frac{p_c - p_0}{1 - p_c}. \quad (3.23)$$

De (3.19) temos que

$$1 - k = \frac{1 - p_c}{1 - p_0},$$

o que completa a demonstração.  $\blacksquare$

Como uma particularização de (2.49), sob o Modelo (3.4), dado  $p_c$ , existem dois mecanismos de censura atuando simultaneamente:

$$M_i \text{ e } C_i, \text{ com } P[C_i = c_i] = (1 - k)^{c_i} (k)^{1 - c_i}, \quad c_i = 0, 1. \quad (3.24)$$

### 3.3 Função de Verossimilhança

Nesta seção, sob o Teorema 2.10, apresentamos a função de verossimilhança para o modelo de mistura padrão exponencial. Raciocinando sob a Figura 3.1, note-se que

$$\begin{aligned} P[\Delta_i = 1] &= P[\Delta_i = 1|M_i = 0] P[M_i = 0] + P[\Delta_i = 1|M_i = 1] P[M_i = 1] \\ &= P[\Delta_i = 1|M_i = 1] P[M_i = 1] = (1 - p_0)(1 - k). \end{aligned} \quad (3.25)$$

De (2.50) e (3.14),

$$\begin{aligned} F_C(t_i) &= \frac{P[T_i \leq t_i, \Delta_i = 0]}{P[\Delta_i = 0]} \\ &= \frac{P[T_i \leq t_i|\Delta_i = 0, M_i = 0] P[\Delta_i = 0|M_i = 0] P[M_i = 0]}{P[\Delta_i = 0]} \\ &\quad + \frac{P[T_i \leq t_i|\Delta_i = 0, M_i = 1] P[\Delta_i = 0|M_i = 1] P[M_i = 1]}{P[\Delta_i = 0]} \\ &= \frac{p_0 \int_0^{t_i} \lambda_c e^{-\lambda_c x_i} dx_i + (1 - p_0) \int_0^{t_i} \left[ \int_{x_i}^{\infty} \lambda e^{-\lambda y_i} dy_i \right] \lambda_c e^{-\lambda_c x_i} dx_i}{P[\Delta_i = 0]} \\ &= \frac{p_0 \int_0^{t_i} \lambda_c e^{-\lambda_c x_i} dx_i + (1 - p_0) \int_0^{t_i} \lambda_c e^{-(\lambda + \lambda_c)x_i} dx_i}{P[\Delta_i = 0]}. \end{aligned} \quad (3.26)$$

Por (2.50), (3.12) e (3.25),

$$\begin{aligned} F_D(t_i) &= \frac{P[T_i \leq t_i, \Delta_i = 1]}{P[\Delta_i = 1]} = \frac{P[T_i \leq t_i, \Delta_i = 1, M_i = 1]}{P[\Delta_i = 1, M_i = 1]} \\ &= \frac{P[T_i \leq t_i, \Delta_i = 1|M_i = 1] P[M_i = 1]}{P[\Delta_i = 1|M_i = 1] P[M_i = 1]} \\ &= \frac{P[T_i \leq t_i, \Delta_i = 1|M_i = 1]}{P[\Delta_i = 1|M_i = 1]} \\ &= \frac{\int_0^{t_i} \left[ \int_{y_i}^{\infty} \lambda_c e^{-\lambda_c x_i} dx_i \right] \lambda e^{-\lambda y_i} dy_i}{P[\Delta_i = 1|M_i = 1]} = \frac{\int_0^{t_i} \lambda e^{-(\lambda + \lambda_c)y_i} dy_i}{1 - k}. \end{aligned} \quad (3.27)$$

De e (2.51) e (2.53),

$$dB(t_i) = \lambda_c e^{-\lambda_c t_i} dt_i, \quad B(t_i) = 1 - e^{-\lambda_c t_i}, \quad e \quad a(t_i) = e^{-\lambda_c t_i}, \quad ,$$

o que confirma a restrição (2.55), e

$$\begin{aligned} P[C_{t_i}] &= [1 - F_p(t_i)] dB(t_i) = [p_0 + (1 - p_0) e^{-\lambda t_i}] \lambda_c e^{-\lambda_c t_i} dt_i, \\ P[D_{t_i}] &= a(t_i) dF_p(t_i) = (1 - p_0) \lambda e^{-(\lambda + \lambda_c)t_i} dt_i. \end{aligned}$$

Sob o Teorema 2.10, temos portanto que a função de verossimilhança de  $\lambda$ ,  $p_0$  e  $\lambda_c$  é dada por

$$L^*(\lambda, p_0, \lambda_c; \mathcal{D}) \propto \prod_{i=1}^n [(1 - p_0)\lambda e^{-(\lambda + \lambda_c)t_i}]^{\delta_i} \times \prod_{i=1}^n [\lambda_c e^{-\lambda_c t_i} (p_0 + (1 - p_0)e^{-\lambda t_i})]^{1 - \delta_i}, \quad (3.28)$$

com espaço paramétrico associado

$$\Theta^* = \{(\lambda, p_0, \lambda_c); \lambda > 0, 0 \leq p_0 \leq 1, \lambda_c > 0\}. \quad (3.29)$$

Por (3.21) e (3.22),  $a(t)$  e  $B(t)$  satisfazem a relação paramétrica (3.21) com

$$F_p(t_i) = (1 - p_0) [1 - e^{-\lambda t_i}]. \quad (3.30)$$

e esta relação é uma motivação que nos induz a considerar a suposição de censura informativa.

Uma segunda motivação é que, sob suposição de censura não-informativa, os elementos da matriz de informação de Fisher associada ao modelo (3.4), apresentada em Maller & Zhou (1996), são esperanças com relação à distribuição de  $X$ , e não têm uma forma fechada. Estas esperanças estão apresentadas no Capítulo 4 e são obtidas numericamente via R, e é uma das contribuições deste trabalho.

### 3.4 Distribuição de $T$

A função de distribuição acumulada da variável  $T_i$  é obtida marginalizando o vetor  $(T_i, \Delta_i)^T$ . Temos, por (2.50), (3.25), (3.26) e (3.27), que

$$\begin{aligned} F_{T_i}(t_i) &= F_C(t_i) P[\Delta_i = 0] + F_D(t_i) P[\Delta_i = 1] \\ &= p_0 \int_0^{t_i} \lambda_c e^{-\lambda_c x_i} dx_i + (1 - p_0) \int_0^{t_i} \lambda_c e^{-(\lambda + \lambda_c)x_i} dx_i \\ &\quad + (1 - p_0) \int_0^{t_i} \lambda e^{-(\lambda + \lambda_c)y_i} dy_i. \end{aligned}$$

Para obter a função densidade  $f(t_i)$ , da variável  $T_i$ , basta derivar esta função com relação à  $t_i$ , e temos que

$$f_{T_i}(t_i) = p_0 \lambda_c e^{-\lambda_c t_i} + (1 - p_0) (\lambda + \lambda_c) e^{-(\lambda + \lambda_c)t_i}. \quad (3.31)$$

A distribuição de  $T$  é útil como alternativa para o cálculo de um dos elementos da matriz de informação de Fisher de  $\lambda$ ,  $p_0$  e  $\eta$  (veja demonstração no Apêndice A).

### 3.5 Resumo dos Principais resultados do Capítulo

Valem aqui os mesmos procedimentos usados na Seção 2.6. Os resultados são listados a seguir.

- Modelo de interesse em (3.4).
- Distribuição da variável de censura em 3.14.
- Teorema 3.1, contribuição da Tese.
- Teorema 3.2, contribuição da Tese.



# Capítulo 4

## Inferência Clássica

Neste capítulo desenvolvemos a inferência clássica para o modelo de mistura padrão exponencial com censura exponencial, ou seja, para os modelos (3.4) e (3.14), para os quais, sob o Teorema 2.10,  $Y$  e  $X$  são independentes. Nas duas primeiras seções consideramos o conjunto de dados observados das variáveis  $T$  e  $\Delta$ . Na primeira, sob suposição de censura não-informativa, e na Seção 4.2 consideramos a suposição de censura informativa. Na Seção 4.3, sob as duas suposições de censura, a inferência é desenvolvida com a adição de uma variável latente ao conjunto de dados de  $T$  e  $\Delta$ , e implementação do algoritmo EM. Na Seção 4.4 desenvolvemos um algoritmo de maximização iterativa, alternativo ao algoritmo EM, que usamos nas estimativas de máxima verossimilhança dos parâmetros de interesse.

Estas duas últimas seções, Seção 4.3 e Seção 4.4, podem ser vistas como dois apêndices deste Capítulo 4, nos quais desenvolvemos dois algoritmos de maximização. O algoritmo de maximização desenvolvido na Seção 4.3, algoritmo EM (Dempster *et al.*, 1977), não é usado nos estudos com dados simulados e com dados reais, apresentados no Capítulo 6. Os motivos são discutidos nos capítulos 6 e 7, nos quais é proposto como trabalho futuro um estudo detalhado dos métodos de inferência clássica desenvolvidos nesta Seção 4.3, relacionados às funções de verossimilhança com variável latente e respectivas matriz de informação de Fisher.

Por outro lado, o algoritmo de maximização desenvolvido na Seção 4.4 é usado para maximizar as funções de verossimilhança dos parâmetros de interesse, sob as suposições de censura não-informativa e de censura informativa.

## 4.1 Inferência Sob Suposição de Censura Não-Informativa

Sob o modelo (3.4) e suposição de censura não-informativa os parâmetros de interesse são  $\lambda$  e  $p_0$ , taxa de falha e probabilidade de cura, respectivamente. A função de verossimilhança de  $\lambda$  e  $p_0$  segue a redução (2.59), e é dada por

$$L(\lambda, p_0; \mathcal{D}) \propto \prod_{i=1}^n [(1 - p_0) \lambda e^{-\lambda t_i}]^{\delta_i} [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1 - \delta_i}, \quad (4.1)$$

em que

$$\mathcal{D} = (\mathbf{t}, \boldsymbol{\delta}), \quad \mathbf{t} = (t_1, \dots, t_n)^T, \quad \boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T, \quad (4.2)$$

$\mathcal{D}$  matriz de dados observados. O espaço paramétrico associado é dado por

$$\Theta = \{(\lambda, p_0); \lambda > 0, 0 \leq p_0 < 1\}. \quad (4.3)$$

A função log-verossimilhança correspondente é dada por

$$\begin{aligned} l(\lambda, p_0; \mathcal{D}) \propto & \sum_{i=1}^n \delta_i [\log(1 - p_0) + \log \lambda - \lambda t_i] \\ & + \sum_{i=1}^n (1 - \delta_i) \log [p_0 + (1 - p_0) e^{-\lambda t_i}]. \end{aligned} \quad (4.4)$$

### 4.1.1 Matriz de Informação

A matriz de informação de Fisher dos parâmetros  $\lambda$  e  $p_0$  é apresentada em Maller & Zhou (1996), em que os elementos da matriz são esperanças com relação à distribuição da variável de censura  $X$ , de expressões em  $X$ ,  $\lambda$  e  $p_0$ . São elas:

$$I_{\lambda\lambda} = n(1 - p_0) E \left\{ \frac{1 - e^{\lambda X}}{\lambda^2} - \frac{p_0 X^2 e^{-\lambda X}}{p_0 + (1 - p_0) e^{-\lambda X}} \right\}, \quad (4.5)$$

$$I_{\lambda p_0} = n E \left\{ \frac{X e^{-\lambda X}}{p_0 + (1 - p_0) e^{-\lambda X}} \right\} \quad (4.6)$$

e

$$I_{p_0 p_0} = n E \left\{ \frac{1 - e^{-\lambda X}}{(1 - p_0) [p_0 + (1 - p_0) e^{-\lambda X}]} \right\}. \quad (4.7)$$

Em Maller & Zhou (1996), estes resultados são obtidos a partir do Teorema A.1, que apresentamos no Apêndice A.

Os elementos da matriz em (4.5)-(4.7) não têm uma forma fechada, e por isso aquele texto sugere o uso da matriz de informação observada. Um importante resultado, que apresentamos neste trabalho, é o cálculo destas esperanças. Para isso, faz-se necessária a definição da transformação  $\varphi$ -paramétrica

$$\varphi(\alpha, \nu, \beta) = \int_0^\infty \frac{x^{\nu-1} e^{-\alpha x}}{1 + \beta e^{-x}} dx, \quad (4.8)$$

com  $\alpha > 0$ ,  $\beta = \frac{1-p_0}{p_0} > 0$  e  $\nu = 1, 2, \dots$ , denominada de integral por transformação  $\varphi$  - paramétrica.

Esta transformação  $\varphi$ -paramétrica possui propriedades relacionadas com as funções Lerch, Lerch-Zeta, Zeta de Riemann, dilogaritmo, exponencial-integral, entre outras, que podem ser encontradas em Abramowitz & Stegun (1964) e Gradshteyn & Ryzhik (2007). Não tem uma forma fechada mas pode ser obtida de duas formas:

- Por meio de funções de integração existentes nos vários sistemas computacionais, como por exemplo, a função “integrate” do sistema R (R Development Core Team, 2009).
- Pelo método de Monte Carlo, dado que

$$I\varphi(\alpha, \nu, \beta) = \frac{(\nu - 1)!}{\alpha^\nu} E_X \left[ \frac{1}{1 + \beta e^{-x}} \right], \quad \text{se } X \sim \Gamma(\nu, \alpha). \quad (4.9)$$

Implementamos procedimentos analíticos e computacionais em busca de soluções com representação em série da integral  $\varphi$ , com o uso de métodos tradicionais como o método de Simpson, trapezóides, pontos médios e combinação destes.

Nenhum destes métodos de integração numérica produziram soluções eficientes, em relação ao tempo computacional e magnitude do erro, principalmente para valores de  $p_0$  próximos aos extremos do intervalo  $(0, 1)$ . Apenas as soluções por integração Monte Carlo produziram resultados uniformemente eficientes quando comparados às soluções obtidas com a função de integração “integrate” do código R (R Development Core Team, 2009).

No entanto o método de Monte Carlo requer um custo computacional maior e então optamos por usar a função “integrate” do sistema R, que é o sistema usado em nossos trabalhos computacionais. Apresentamos no teorema a seguir o cálculo das esperanças em (4.5)-(4.7), parametrizada em  $\lambda$ ,  $p_0$  e  $\lambda_c$ , a partir transformação  $\varphi$ -paramétrica (4.8), e sob o teorema subsequente, reparametrizada em  $\eta$ .

**Teorema 4.1** *Os elementos da matriz de informação de Fisher associada à função de verossimilhança (4.1) são dados por*

$$I_{\lambda\lambda} = n \frac{n(1-p_0)}{\lambda^2} \left[ \frac{\lambda}{\lambda + \lambda_c} - \frac{\lambda_c}{\lambda} \varphi \left( \frac{\lambda + \lambda_c}{\lambda}, 3, \frac{1-p_0}{p_0} \right) \right], \quad (4.10)$$

$$I_{\lambda p_0} = I_{p_0 \lambda} = \frac{n\lambda_c}{\lambda^2 p_0} \varphi \left( \frac{\lambda + \lambda_c}{\lambda}, 2, \frac{1-p_0}{p_0} \right), \quad (4.11)$$

$$I_{p_0 p_0} = \frac{n\lambda_c}{\lambda p_0 (1-p_0)} \left[ \varphi \left( \frac{\lambda_c}{\lambda}, 1, \frac{1-p_0}{p_0} \right) - \varphi \left( \frac{\lambda + \lambda_c}{\lambda}, 1, \frac{1-p_0}{p_0} \right) \right]. \quad (4.12)$$

**Demonstração.** Aplicando a transformação  $\varphi$ -paramétrica (4.8) em (4.5)-(4.7) obtemos os resultados (4.10)-(4.12). ■

**Teorema 4.2** *As expressões (4.10)-(4.12), reparametrizadas em  $\eta$  são dadas respectivamente por*

$$I_{\lambda\lambda} = \frac{n(1-p_0)}{\lambda^2} \left[ \frac{1}{1+\eta} - \eta \varphi \left( 1+\eta, 3, \frac{1-p_0}{p_0} \right) \right], \quad (4.13)$$

$$I_{\lambda p_0} = I_{p_0 \lambda} = \frac{n\eta}{\lambda p_0} \varphi \left( 1+\eta, 2, \frac{1-p_0}{p_0} \right), \quad (4.14)$$

$$I_{p_0 p_0} = \frac{n \eta}{p_0 (1 - p_0)} \left[ \varphi \left( \eta, 1, \frac{1 - p_0}{p_0} \right) - \varphi \left( 1 + \eta, 1, \frac{1 - p_0}{p_0} \right) \right]. \quad (4.15)$$

**Demonstração.** Considerando que  $X \sim exp(\lambda_c)$ , em que  $\lambda_c$  é dado por (3.21), Teorema 3.2, aplicar a transformação  $\varphi$ -paramétrica (4.8) em (4.5)-(4.7). ■

Como já foi comentado, note-se por (4.5)-(4.7) que a matriz de informação depende da variável de censura  $X$ . A variável  $X$  está relacionada parametricamente com  $Y$ , sob o modelo (3.4), e essa relação é explícita nas expressões em (3.18), (3.20) e (3.21). Devido às expressões em (4.10)-(4.12), em que se nota que os elementos da matriz são funções de  $k$ , somos motivados a considerar a reparametrização em  $k$ . No entanto optamos pela reparametrização em  $\eta$ , devido à praticidade de manipulação, do ponto de vista algébrico e computacional, isso melhor visto na próxima seção.

Na abordagem clássica, o uso da matriz de informação sob censura não informativa é possível em duas opções: a primeira é o uso da matriz observada na construção de intervalos assintóticos, sugerida em Maller & Zhou (1996). Na segunda opção devemos considerar  $\lambda_c$  conhecido nas expressões (4.10)-(4.12), ou  $\eta$  conhecido nas expressões (4.13)-(4.15). O problema que evidenciamos é que não é possível conhecer o valor de  $\lambda_c$ , parâmetro da variável de censura, ou de  $\eta$ , em que, de (3.21),  $\eta = \frac{\lambda_c}{\lambda}$ . A solução que vemos é o uso da função de verossimilhança sob suposição de censura informativa, visto na próxima seção, reparametrizada em  $\lambda$ ,  $p_0$  e  $\eta$ .

## 4.2 Inferência Sob Suposição de Censura Informativa

Optamos pela reparametrização em  $\eta$  através de (3.21) e (3.22), dado que esta, além de proporcionar uma forma simplificada para a matriz de informação de Fisher, nos permite obter o estimador de máxima verossimilhança (emv)  $\hat{\eta}$  de

$\eta$  como função de  $\lambda$ . Temos então três parâmetros desconhecidos,  $\lambda$ ,  $p_0$  e  $\eta$ , e a função (3.28) é reescrita da forma

$$L^{**}(\lambda, p_0, \eta; \mathcal{D}) \propto \prod_{i=1}^n [(1-p_0)\lambda e^{-\lambda(1+\eta)t_i}]^{\delta_i} \times \prod_{i=1}^n [\lambda\eta e^{-\lambda\eta t_i} (p_0 + (1-p_0)e^{-\lambda t_i})]^{1-\delta_i}, \quad (4.16)$$

com função log-verossimilhança associada

$$l^{**}(\lambda, p_0, \eta; \mathcal{D}) \propto \sum_{i=1}^n \delta_i [\log(1-p_0) - \lambda t_i] + n \log(\lambda) - \lambda\eta \sum_{i=1}^n t_i + \sum_{i=1}^n (1-\delta_i) \log(\eta) + \sum_{i=1}^n (1-\delta_i) \log[p_0 + (1-p_0)e^{-\lambda t_i}], \quad (4.17)$$

em que  $\mathcal{D}$  é dado por (4.2). O espaço paramétrico associado é dado por

$$\Theta^{**} = \{(\lambda, p_0, \eta); \lambda > 0, 0 \leq p_0 < 1, \eta > 0\}. \quad (4.18)$$

### 4.2.1 Matriz de Informação

**Teorema 4.3** *A matriz de informação de Fisher associada à função de verossimilhança (4.16) é dada por*

$$IF^{**}(\lambda, p_0, \eta) = \begin{bmatrix} I_{\lambda\lambda} & I_{\lambda p_0} & I_{\lambda\eta} \\ I_{p_0\lambda} & I_{p_0 p_0} & I_{p_0\eta} \\ I_{\eta\lambda} & I_{\eta p_0} & I_{\eta\eta} \end{bmatrix}, \quad (4.19)$$

em que

$$I_{\lambda\lambda}^{**} = n \frac{(1-p_0)}{\lambda^2} \left[ \frac{1}{1-p_0} - \eta\varphi\left(1+\eta, 3, \frac{1-p_0}{p_0}\right) \right], \quad (4.20)$$

$$I_{\lambda p_0}^{**} = I_{p_0\lambda}^{**} = n \frac{\eta}{\lambda p_0} \left[ \varphi\left(1+\eta, 2, \frac{1-p_0}{p_0}\right) \right], \quad (4.21)$$

$$I_{\lambda\eta}^{**} = I_{\eta\lambda}^{**} = n \frac{(p_0 + \eta)}{\lambda\eta(1+\eta)}, \quad (4.22)$$

$$I_{p_0 p_0}^{**} = n \frac{\eta}{(1-p_0)p_0} \left[ \varphi \left( \eta, 1, \frac{1-p_0}{p_0} \right) - \varphi \left( 1+\eta, 1, \frac{1-p_0}{p_0} \right) \right], \quad (4.23)$$

$$I_{p_0 \eta}^{**} = I_{\eta p_0}^{**} = 0, \quad (4.24)$$

$$I_{\eta \eta}^{**} = n \frac{(p_0 + \eta)}{\eta^2 (1 + \eta)}. \quad (4.25)$$

**Demonstração:** Vide Apêndice A. ■

A função de verossimilhança em (4.16), sob suposição de censura informativa, e sua respectiva matriz de informação em (4.19), cujos elementos são dados pelas expressões em (4.20)-(4.25), resolve o problema evidenciado no final da seção anterior, que é a impossibilidade do uso da matriz de informação de Fisher sob censura não-informativa, dado que  $\eta$  é não-conhecido.

## 4.2.2 Uma Opção Alternativa de Inferência sob Censura Informativa

**Proposição 4.1** Dado  $\lambda$ , o valor de  $\eta$  que maximiza a função (4.16) é dado por

$$\hat{\eta}(\lambda) = \frac{\sum_{i=1}^n (1 - \delta_i)}{\lambda \sum_{i=1}^n t_i}. \quad (4.26)$$

**Demonstração.** Fixando  $\lambda$ , maximizar a função (4.16) com relação a  $\eta$ . ■

Substituindo (4.26) em (4.16) obtemos a função de verossimilhança perfilada associada a (4.16), que é dada por

$$L_p^{**}(\lambda, p_0; \mathcal{D}) \propto \left[ \frac{\sum (1 - \delta_i)}{\sum t_i} \right]^{\sum (1 - \delta_i)} e^{\sum (1 - \delta_i)} \times \prod_{i=1}^n [(1 - p_0) \lambda e^{-\lambda t_i}]^{\delta_i} [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1 - \delta_i}, \quad (4.27)$$

e, a menos de uma constante, pode ser expressa na forma (4.1). Ou seja, os valores de  $\lambda$  e  $p_0$  que maximizam (4.27) também maximizam (4.1).

De (4.27) temos que as estimativas de máxima verossimilhança (emv) de  $\lambda$  e  $p_0$  são obtidas numericamente, maximizando a função log-verossimilhança perfilada

$$l_p^{**}(\lambda, p_0; \mathcal{D}) = \log [L_p^{**}(\lambda, p_0)] = l[\lambda, p_0, \hat{\eta}(\lambda)], \quad (4.28)$$

em que  $\hat{\lambda}$  e  $\hat{p}_0$  são estimadores de máxima verossimilhança emv. De (4.26) temos também que

$$\hat{k} = \frac{\hat{\eta}}{1 + \hat{\eta}}, \quad \text{em que } \hat{\eta} = \hat{\eta}(\hat{\lambda}). \quad (4.29)$$

### 4.3 Dados Com Variável Latente

Desenvolvemos nesta seção o algoritmo EM (Dempster *et al.*, 1977) como metodologia alternativa de maximização da função de verossimilhança dos parâmetros de interesse, sob as duas suposições, de censura não-informativa e de censura informativa. Note-se que nas funções (4.1), (3.28) e (4.16), o termo

$$\prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1-\delta_i}$$

não permite expressar a função como produto de componentes independentes em  $\lambda$  e  $p_0$ . Mas isso não é mais obstáculo para maximizar a função em  $\lambda$  e  $p_0$  dados os recursos computacionais existentes. Entretanto, acrescentamos aqui uma variável latente que permite fatorar a função de verossimilhança em  $\lambda$  e  $p_0$ . Isso permite uma simplificação na maximização, porque elimina a relação de dependência paramétrica.

#### 4.3.1 Sob Suposição de Censura Não-Informativa

A função de verossimilhança em (4.1) pode ser escrita na forma

$$L(\lambda, p_0; \mathcal{D}) \propto [(1 - p_0) \lambda]^{\sum \delta_i} e^{-\lambda \sum \delta_i t_i} \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1-\delta_i}. \quad (4.30)$$

Introduzimos então, para  $i = 1, 2, \dots, n$ , a seguinte variável aleatória:

$$W_i \sim \text{Bernoulli}(\rho_i), \quad (4.31)$$



em que

$$\rho_i = \frac{p_0}{p_0 + (1 - p_0) S(t_i|\lambda)}. \quad (4.32)$$

Definida a variável  $W_i$ , apresentamos então a função de verossimilhança ampliada, dada pela equação

$$\begin{aligned} L_a(\lambda, p_0; \mathbf{D}_a) &\propto [(1 - p_0) \lambda]^{\sum \delta_i} e^{-\lambda \sum \delta_i t_i} \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1-\delta_i} \\ &\times \prod_{i=1}^n \left\{ \rho_i^{w_i} (1 - \rho_i)^{(1-w_i)} \right\}^{(1-\delta_i)}, \end{aligned} \quad (4.33)$$

que pode ser escrita na forma

$$L_a(\lambda, p_0; \mathbf{D}_a) \propto p_0^{\sum (1-\delta_i) w_i} (1 - p_0)^{n - \sum (1-\delta_i) w_i} \lambda^{\sum \delta_i} e^{-\lambda \sum [1 - (1-\delta_i) w_i] t_i}, \quad (4.34)$$

em que

$$\mathbf{D}_a = [\mathbf{t}, \boldsymbol{\delta}, \mathbf{w}] = \mathbf{D} \cup [\mathbf{w}], \quad \mathbf{w} = (w_1, \dots, w_n), \quad (4.35)$$

e a matriz de dados  $\mathbf{D}$  é dada por (4.2). A matriz  $\mathbf{D}_a$  é conhecida como matriz de dados ampliados, ou também como matriz de dados completos.

O vetor de parâmetros e espaço paramétrico, respectivamente, associados à função (4.34) é o mesmo dado em (4.3). Tomando-se o logaritmo desta última expressão temos que

$$\begin{aligned} l_a(\lambda, p_0; \mathbf{D}_a) &= \log(p_0) \sum_{i=1}^n (1 - \delta_i) w_i + \log(1 - p_0) \left[ n - \sum_{i=1}^n (1 - \delta_i) w_i \right] \\ &+ \log(\lambda) \sum_{i=1}^n \delta_i - \lambda \sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i. \end{aligned} \quad (4.36)$$

Para a aplicação do algoritmo EM, necessitamos da distribuição condicional da variável latente  $W_i$ . A função densidade conjunta de  $(t_i, \delta_i, w_i)$  é dada por

$$f(t_i, \delta_i, w_i) = p_0^{(1-\delta_i)w_i} (1 - p_0)^{1-(1-\delta_i)w_i} \lambda^{\delta_i} e^{-\lambda[1-(1-\delta_i)w_i]t_i}. \quad (4.37)$$

Temos que

$$\begin{aligned} f(t_i, \delta_i) &= \sum_{w_i=0}^1 f(t_i, \delta_i, w_i) \\ &= (1 - p_0) \lambda^{\delta_i} e^{-\lambda t_i} + p_0^{(1-\delta_i)} (1 - p_0)^{\delta_i} \lambda^{\delta_i} e^{-\lambda \delta_i t_i}. \end{aligned} \quad (4.38)$$

Portanto, a função de probabilidade condicional de  $W_i$  é dada por

$$\begin{aligned} f(w_i|t_i, \delta_i) &= \frac{f(t_i, \delta_i, w_i)}{f(t_i, \delta_i)} \\ &= \frac{p_0^{(1-\delta_i)w_i} (1-p_0)^{1-(1-\delta_i)w_i} \lambda^{\delta_i} e^{-\lambda[1-(1-\delta_i)w_i]t_i}}{(1-p_0) \lambda^{\delta_i} e^{-\lambda t_i} + p_0^{(1-\delta_i)} (1-p_0)^{\delta_i} \lambda^{\delta_i} e^{-\lambda \delta_i t_i}}. \end{aligned}$$

A esperança condicional de  $W_i$  é dada então por

$$E(W_i|t_i, \delta_i) = \frac{p_0^{(1-\delta_i)} (1-p_0)^{\delta_i} \lambda^{\delta_i} e^{-\lambda \delta_i t_i}}{(1-p_0) \lambda^{\delta_i} e^{-\lambda t_i} + p_0^{(1-\delta_i)} (1-p_0)^{\delta_i} \lambda^{\delta_i} e^{-\lambda \delta_i t_i}}.$$

Denotando-se a esperança condicional de  $W_i$  por  $r_i^{(k+1)}$ , temos

$$\begin{aligned} r_i^{(k+1)} &= E[W_i|t_i, \delta_i, \lambda^{(k)}, p_0^{(k)}] \\ &= \frac{[p_0^{(k)}]^{(1-\delta_i)} [1-p_0^{(k)}]^{\delta_i} e^{-\lambda^{(k)} \delta_i t_i}}{[1-p_0^{(k)}] e^{-\lambda^{(k)} t_i} + [p_0^{(k)}]^{(1-\delta_i)} [1-p_0^{(k)}]^{\delta_i} e^{-\lambda^{(k)} \delta_i t_i}}. \end{aligned} \quad (4.39)$$

Segue que na  $(k+1)$ -ésima iteração o algoritmo é formado pelos seguintes passos:

- **Passo E** - Cálculo da esperança condicional da função log-verossimilhança dos dados completos: Considere-se as notações

$$\begin{aligned} \phi &= (\lambda, p_0)^T, \quad \mathcal{D} = [\mathbf{t}, \boldsymbol{\delta}], \quad \mathcal{D}_a = [\mathbf{t}, \boldsymbol{\delta}, \mathbf{w}], \quad \text{em que } \mathbf{t} = (t_1, \dots, t_n)^T, \\ \boldsymbol{\delta} &= (\delta_1, \dots, \delta_n)^T \quad \text{e} \quad \mathbf{w} = (w_1, \dots, w_n)^T. \end{aligned}$$

Desta forma,

$$\begin{aligned} E[l(\phi, \mathcal{D}_a) | \mathcal{D}, \lambda^{(k)}, p_0^{(k)}] &= \log(p_0) \sum_{i=1}^n (1-\delta_i) r_i^{(k+1)} \\ &\quad + \log(1-p_0) \left[ n - \sum_{i=1}^n (1-\delta_i) r_i^{(k+1)} \right] \\ &\quad + \log(\lambda) \sum_{i=1}^n \delta_i \\ &\quad - \lambda \sum_{i=1}^n t_i \left[ 1 - (1-\delta_i) r_i^{(k+1)} \right]. \end{aligned}$$

- **Passo M** - Maximização das expressões:

$$Q^{(1)}[\lambda | \lambda^{(k)}, p_0^{(k)}] = \log(\lambda) \sum_{i=1}^n \delta_i - \lambda \sum_{i=1}^n t_i \left[ 1 - (1-\delta_i) r_i^{(k+1)} \right],$$

$$Q^{(2)} \left[ p_0 | \lambda^{(k)}, p_0^{(k)} \right] = \log(p_0) \sum_{i=1}^n (1 - \delta_i) r_i^{(k+1)} \\ + \log(1 - p_0) \left[ 1 - \sum_{i=1}^n (1 - \delta_i) r_i^{(k+1)} \right].$$

Neste **Passo M** a maximização de  $Q^{(1)}$  gera  $\lambda^{(k+1)}$  e a maximização de  $Q^{(2)}$  gera  $p_0^{(k+1)}$ , e estes pontos maximizadores são dados respectivamente por

$$\lambda^{(k+1)} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i \left[ 1 - (1 - \delta_i) r_i^{(k+1)} \right]}$$

e

$$p_0^{(k+1)} = \frac{\sum_{i=1}^n (1 - \delta_i) r_i^{(k+1)}}{n}.$$

As estimativas de máxima verossimilhança em

$$\hat{\phi} = \left( \hat{\lambda}, \hat{p}_0 \right)^T \quad (4.40)$$

são obtidas com a convergência do algoritmo, segundo o critério de parada estabelecido. Consideramos aqui como critério a condição

$$\left| \phi^{(k+1)} - \phi^{(k)} \right| < \epsilon, \quad \epsilon > 0,$$

em que  $\phi^{(k)} = \left( \lambda^{(k)}, p_0^{(k)} \right)^T$ .

Para obter a matriz de informação relacionada à função (4.34) devemos considerar que, dado  $T_i = t_i$ , a variável aleatória  $\chi_i = (1 - \Delta_i) W_i$  tem distribuição

$$\chi_i \sim \text{Bernoulli}(p_0). \quad (4.41)$$

Note-se que, de (4.31),

$$P[\chi_i = 1] = P[1 - \Delta_i = 1, W_i = 1] = P[\Delta_i = 0, W_i = 1] \\ = P[\Delta_i = 0] P[W_i = 1 | \Delta_i = 0] = p_0.$$

Portanto, a matriz de informação de Fisher, dada pelo teorema abaixo é de fácil demonstração.

**Teorema 4.4** A matriz de informação relacionada à função (4.34) é dada por

$$IF_a(\lambda, p_0) = \begin{bmatrix} \frac{n(1-p_c)}{\lambda^2} & 0 \\ 0 & \frac{n}{p_0(1-p_0)} \end{bmatrix}, \quad (4.42)$$

e

$$Det [IF_a(\lambda, p_0)] = \frac{n^2(1-p_c)}{\lambda^2 p_0(1-p_0)}. \quad (4.43)$$

**Demonstração.** Basta considerar a distribuição (4.41) nas derivadas segundas da função (4.36). ■

### 4.3.2 Sob Suposição de Censura Informativa

A função de verossimilhança em (4.16) pode ser escrita na forma

$$L^{**}(\lambda, p_0, \eta; \mathbf{D}) = [1 - p_0]^{\sum \delta_i} \lambda^n e^{-\lambda \sum \delta_i t_i} \eta^{n - \sum \delta_i} e^{-\lambda \eta \sum t_i} \\ \times \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1 - \delta_i}. \quad (4.44)$$

Introduzindo, para  $i = 1, 2, \dots, n$ , a mesma variável aleatória  $W_i$ , com função de probabilidade (4.31), temos

$$L_a^{**}(\lambda, p_0, \eta; \mathbf{D}_a) = [1 - p_0]^{\sum \delta_i} \lambda^n e^{-\lambda \sum \delta_i t_i} \eta^{n - \sum \delta_i} e^{-\lambda \eta \sum t_i} \\ \times \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{1 - \delta_i} \prod_{i=1}^n \left\{ \rho_i^{w_i} (1 - \rho_i)^{(1 - w_i)} \right\}^{(1 - \delta_i)},$$

que pode ser escrita

$$L_a^{**}(\lambda, p_0, \eta; \mathbf{D}_a) \propto p_0^{\sum (1 - \delta_i) w_i} (1 - p_0)^{n - \sum (1 - \delta_i) w_i} \\ \times \lambda^n e^{-\lambda \sum [1 - (1 - \delta_i) w_i] t_i} \eta^{n - \sum \delta_i} e^{-\lambda \eta \sum t_i}. \quad (4.45)$$

O vetor de parâmetros e espaço paramétrico, respectivamente, associados à função (4.45) é o mesmo dado em (4.18). Tomando-se o logaritmo desta última expressão temos que

$$l_a^{**}(\lambda, p_0, \eta; \mathbf{D}_a) = \log(p_0) \sum_{i=1}^n (1 - \delta_i) w_i + \log(1 - p_0) \left[ n - \sum_{i=1}^n (1 - \delta_i) w_i \right] \\ + n \log(\lambda) - \lambda \sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i \\ + \log(\eta) \left( n - \sum \delta_i \right) - \lambda \eta \sum t_i. \quad (4.46)$$

Dado  $\lambda$ , valor de  $\eta$  que maximiza (4.16), dado por (4.26), é o mesmo que maximiza (4.45), e, a menos de uma constante, também reduz (4.45) e (4.46) ao caso não informativo em (4.34) e (4.36), respectivamente. Portanto, as estimativas de  $\lambda$  e  $p_0$  são as mesmas de (4.40). Quanto à estimativa de  $\eta$ , dado  $\hat{\lambda}$  em (4.40), temos por (4.26) que

$$\hat{\eta} = \frac{\sum_{i=1}^n (1 - \delta_i)}{\hat{\lambda} \sum_{i=1}^n t_i}. \quad (4.47)$$

As estimativas de máxima verossimilhança de  $\lambda$ ,  $p_0$  e  $\eta$ , associadas à função (4.45) são dadas por

$$\hat{\phi}^{**} = (\hat{\phi}, \hat{\eta}),$$

em que  $\hat{\phi}$  é dada por (4.40) e  $\hat{\eta}$  de (4.47).

Fácilmente se obtém a matriz de informação de Fisher associada à função (4.45), dada pelo teorema a seguir.

**Teorema 4.5** *A matriz de informação de Fisher relacionada à função de verossimilhança (4.45) é dada por*

$$IF_a^{**}(\lambda, p_0, \eta) = \begin{bmatrix} \frac{n}{\lambda^2} & 0 & \frac{n(p_0 + \eta)}{\lambda\eta(1 + \eta)} \\ 0 & \frac{n}{p_0(1 - p_0)} & 0 \\ \frac{n(p_0 + \eta)}{\lambda\eta(1 + \eta)} & 0 & \frac{np_c}{\eta^2} \end{bmatrix}. \quad (4.48)$$

**Demonstração.** De (3.31), com  $\lambda_c$  dada por (3.21), temos que  $-E \left[ \frac{\partial^2 l_a^{**}(\lambda, p_0, \eta; \mathcal{D}_a)}{\partial \lambda \partial \eta} \right] = \frac{n(p_0 + \eta)}{\lambda\eta(1 + \eta)}$ . Da mesma forma verificamos que  $-E \left[ \frac{\partial^2 l_a^{**}(\lambda, p_0, \eta; \mathcal{D}_a)}{\partial \eta^2} \right] = \frac{np_c}{\eta^2}$ . ■

## 4.4 Algoritmo de Maximização Iterativa

Em estudos preliminares com dados simulados desenvolvidos no início deste projeto de pesquisa, em amostras de tamanho pequeno ( $n = 20$  e  $n = 50$ ), o uso das funções de maximização em código R acusava erro com uma frequência não desprezível. Esta frequência tendia para zero com o aumento do tamanho

da amostra. Em face disso optamos por programar o algoritmo de Newton-Rapson em código R, o qual acusava os mesmos erros e comportamentos similares. Omitimos a investigação das causas, mas a nossa conjectura é de que a ocorrência de erros está relacionada à ocorrência de matriz de derivadas negativa-definida. A verificação desta conjectura e uma investigação detalhada deste fato deixamos como trabalho futuro.

Este problema nos motivou à desenvolver um algoritmo alternativo, que convençionalmente chamamos de algoritmo de maximização iterativa, que, aplicado conjuntamente com a função *optimize* (Brent, 1973) do código R (R Development Core Team, 2009), permite maximizar as funções (4.1) e (4.16) mesmo para amostras pequenas. Por este novo algoritmo a estimação de máxima verossimilhança dos parâmetros é realizada iterativamente, com maximização marginal sob cada parâmetro, a cada iteração, de forma que, na  $k$ -ésima iteração, a função de verossimilhança é maximizada com relação à um único parâmetro, dado o valor obtido na iteração anterior dos outros parâmetros envolvidos. Em código R, a função *optimize* maximiza em apenas uma dimensão, ideal para o uso deste novo algoritmo. A seguir o enunciado e demonstração do algoritmo de maximização iterativa.

**Teorema 4.6** *Soponha-se uma função de verossimilhança qualquer, dos parâmetros  $\theta_1, \dots, \theta_p$ , log-côncava com relação à cada um destes parâmetros, que denotamos por*

$$L(\theta_1, \dots, \theta_p; \mathcal{D}^*), \quad (4.49)$$

em que  $\mathcal{D}^*$  é um conjunto de dados. Se existe  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^T$  que a maximiza, este pode ser obtido com o algoritmo abaixo:

- Valores iniciais:  $\theta_1^{(1)} = \theta_{11}, \theta_2^{(1)} = \theta_{21}, \dots, \theta_p^{(1)} = \theta_{p1}$ .
- Na segunda iteração,

1. Obter  $\theta_1^{(2)} | \theta_2^{(1)}, \dots, \theta_p^{(1)}$ . Dados  $\theta_2^{(1)}, \dots, \theta_p^{(1)}$ , maximiza-se (4.49) com relação à  $\theta_1$ .

2. Obter  $\theta_2^{(2)} | \theta_1^{(2)}, \theta_3^{(1)}, \dots, \theta_p^{(1)}$ . Dados  $\theta_1^{(2)}, \theta_3^{(1)}, \dots, \theta_p^{(1)}$ , maximiza-se (4.49) com relação à  $\theta_2$ .
3. Para  $j = 3, \dots, p-1$ , obter  $\theta_j^{(2)} | \theta_1^{(2)}, \dots, \theta_{j-1}^{(2)}, \theta_{j+1}^{(1)}, \dots, \theta_p^{(1)}$ . Dados  $\theta_1^{(2)}, \dots, \theta_{j-1}^{(2)}, \theta_{j+1}^{(1)}, \dots, \theta_p^{(1)}$ , maximiza-se (4.49) com relação à  $\theta_j$ .
4. Obter  $\theta_p^{(2)} | \theta_1^{(2)}, \dots, \theta_{p-1}^{(2)}$ . Dados  $\theta_1^{(2)}, \dots, \theta_{p-1}^{(2)}$ , maximiza-se (4.49) com relação à  $\theta_p$ .

- Na  $k$ -ésima iteração, para  $k > 2$ ,

1. Obter  $\theta_1^{(k)} | \theta_2^{(k-1)}, \dots, \theta_p^{(k-1)}$ . Dados  $\theta_2^{(k-1)}, \dots, \theta_p^{(k-1)}$ , maximiza-se (4.49) com relação à  $\theta_1$ .
2. Obter  $\theta_2^{(k)} | \theta_1^{(k)}, \theta_3^{(k-1)}, \dots, \theta_p^{(k-1)}$ . Dados  $\theta_1^{(k)}, \theta_3^{(k-1)}, \dots, \theta_p^{(k-1)}$ , maximiza-se (4.49) com relação à  $\theta_2$ .
3. Para  $j = 3, \dots, p-1$ , obter  $\theta_j^{(k)} | \theta_1^{(k)}, \dots, \theta_{j-1}^{(k)}, \theta_{j+1}^{(k-1)}, \dots, \theta_p^{(k-1)}$ . Dados  $\theta_1^{(k)}, \dots, \theta_{j-1}^{(k)}, \theta_{j+1}^{(k-1)}, \dots, \theta_p^{(k-1)}$ , maximiza-se (4.49) com relação à  $\theta_j$ .
4. Obter  $\theta_p^{(k)} | \theta_1^{(k)}, \dots, \theta_{p-1}^{(k)}$ . Dados  $\theta_1^{(k)}, \dots, \theta_{p-1}^{(k)}$ , maximiza-se (4.49) com relação à  $\theta_p$ .

- Repetir o item anterior até que se satisfaça o critério de parada.

- As estimativas de máxima verossimilhança

$$\hat{\boldsymbol{\theta}} = \left( \hat{\theta}_1, \dots, \hat{\theta}_p \right)^T \quad (4.50)$$

são obtidas com a convergência do algoritmo, segundo o critério de parada estabelecido. Consideramos aqui como critério a condição

$$\left| \boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k-1)} \right| < \epsilon, \quad \epsilon > 0,$$

em que  $\boldsymbol{\theta}^{(k)} = \left( \theta_1^{(k)}, \dots, \theta_p^{(k)} \right)^T$ .

**Demonstração:** Supondo primeiramente que o espaço paramétrico de (4.49)

seja contínuo, para  $k \geq 2$ , defina-se

$$\begin{aligned}
 L_1^k &= L(\theta_1^k, \theta_2^{k-1}, \dots, \theta_p^{k-1}; \mathfrak{D}^*), \\
 L_2^k &= L(\theta_1^k, \theta_2^k, \theta_3^{k-1}, \dots, \theta_p^{k-1}; \mathfrak{D}^*), \\
 &\dots \\
 L_j^k &= L(\theta_1^k, \dots, \theta_j^k, \theta_{j+1}^{k-1}, \dots, \theta_p^{k-1}; \mathfrak{D}^*), \quad j = 2, \dots, p-1, \\
 &\dots \\
 L_p^k &= L(\theta_1^k, \dots, \theta_p^k; \mathfrak{D}^*).
 \end{aligned}$$

A suposição de log-concavidade implica que

$$L_1^2 \leq \dots \leq L_p^2 \leq L_1^3 \leq \dots \leq L_p^3 \leq \dots \leq L_1^k \leq \dots \leq L_p^k \leq \dots,$$

e a seqüência

$$\{L_j^k\} = L_1^2, \dots, L_p^2, L_1^3, \dots, L_p^3, \dots, L_1^k, \dots, L_p^k, \dots$$

é não decrescente, e  $\lim_{k \rightarrow \infty} \{L_j^k\} = L(\hat{\theta}_1, \dots, \hat{\theta}_p; \mathfrak{D}^*)$ ,  $j = 1, \dots, p$ . Ainda que o espaço paramétrico de (4.49) seja discreto para pelo menos um  $\theta_j$ ,  $j = 1, \dots, p$ , o resultado é imediato. ■

No código R, a função *optimize* também apresenta problemas ao maximizar funções produto, como é o caso da função de verossimilhança. Optamos então por trabalhar com a função log-verossimilhança, que transforma produtos numa soma de log-funções. Nos parece que a função *optimize* apresenta maior facilidade para maximizar funções do tipo soma.



## Capítulo 5

# Inferência Bayesiana Objetiva

A inferência bayesiana objetiva (Berger, 2006) considera somente distribuições *a priori* não-informativas, como por exemplo, as distribuições *a priori* de Jeffreys (Jeffreys, 1961) e distribuições *a priori* de referência (Bernardo, 1979; Berger *et al.*, 2009). Neste trabalho, e portanto neste e no próximo capítulo, consideramos somente distribuições *a priori* de Jeffreys, associadas às funções de verossimilhança apresentadas no Capítulo 4. Em cada caso, por (4.8) e (4.9), é fácil verificar que a distribuição conjunta *a posteriori* relacionada é própria, dado que pode ser majorada por uma função limitante, cuja integral converge. Por isso não vemos a necessidade de uma demonstração mais detalhada. Mas desenvolvemos toda a inferência bayesiana na forma de proposições e corolários.

Nas duas primeiras seções, obtemos a distribuição *a priori* de Jeffreys associada às funções (4.1) e (4.16), e as respectivas distribuições *a posteriori* conjunta, seguidas das distribuições condicionais marginais *a posteriori* correspondentes. Sob a suposição de censura não-informativa, usamos a função de verossimilhança (4.1) e respectiva matriz de informação de Fisher, sob as expressões (4.13)-(4.15), dado  $\eta$ . Sob a suposição de censura informativa, usamos a função de verossimilhança (4.16) e respectiva matriz de informação de Fisher dada em (4.19).

Na Seção 5.3 obtemos a distribuição *a priori* de Jeffreys associada as funções

de verossimilhança com dados ampliados, em (4.34) e (4.45), e respectivas matrizes de informação de Fisher. Embora estas duas funções não sejam usadas na inferência clássica, com relação aos estudos apresentados no Capítulo 6 (veja segundo parágrafo do Capítulo 4), nesta Seção 5.3 desenvolvemos as distribuições *a posteriori* relacionada, das quais o caso com censura informativa é usado nas inferências apresentadas no Capítulo 6.

Na Seção 5.4 apresentamos uma inferência bayesiana empírica como uma metodologia bayesiana alternativa, a qual é usada também nos estudos apresentados no Capítulo 6. Os estudos e resultados com dados simulados e com dados reais realizados sob esta metodologia são importantes contribuições deste trabalho. No entanto esta metodologia requer um estudo futuro detalhado, como comentamos nas conclusões finais apresentadas no Capítulo 7.

## 5.1 Modelo Com Censura Não-Informativa

Sob a função de verossimilhança (4.1), na matriz de informação de Fisher associada, dada pelas expressões (4.13)-(4.15), devemos considerar  $\eta$  conhecido. Nas duas proposições a seguir determinamos a distribuição *a priori* de Jeffreys associada e a respectiva distribuição conjunta *a posteriori*.

**Proposição 5.1** *Sob o modelo (4.1), a distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$  é imprópria e é dada por*

$$J(\lambda, p_0) \propto \frac{1}{\lambda p_0^{1/2}} J_{p_0}, \quad (5.1)$$

em que

$$J_{p_0} = \sqrt{\left(\frac{1}{1+\eta} - \eta\varphi_{11}\right) (\varphi_{221} - \varphi_{222}) - \left[\frac{\eta}{\lambda p_0}\varphi_{12}\right]^2},$$

$$\varphi_{11} = \varphi\left(1 + \eta, 3, \frac{1-p_0}{p_0}\right), \quad \varphi_{12} = \varphi\left(1 + \eta, 2, \frac{1-p_0}{p_0}\right),$$

$$\varphi_{221} = \varphi\left(\eta, 1, \frac{1-p_0}{p_0}\right), \quad \varphi_{222} = \varphi\left(1 + \eta, 1, \frac{1-p_0}{p_0}\right).$$

**Demonstração.** Considerar a matriz de informação de Fisher, dada pelos elementos em (4.13)-(4.15). Calcular a raiz quadrada do determinante. Para provar que a distribuição em (5.1) é imprópria basta verificar que

$$\int_{\Theta} \frac{1}{\lambda p_0^{1/2}} d\lambda dp_0 = \infty,$$

em que  $\Theta$  é dado por (4.3). ■

**Proposição 5.2** *Sob o modelo (4.1), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ , a função de distribuição a posteriori conjunta é própria e é dada por*

$$\begin{aligned} \pi(\lambda, p_0 | \mathcal{D}) &\propto p_0^{-1/2} (1 - p_0)^{\sum \delta_i} \lambda^{\sum \delta_i - 1} e^{-\lambda \sum t_i \delta_i} \\ &\times \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{(1 - \delta_i)} J_{p_0}. \end{aligned} \quad (5.2)$$

**Demonstração.** A prova de (5.2) é imediata. Para verificar que é própria temos que

$$\pi(\lambda, p_0 | \mathcal{D}) \leq p_0^{-1/2} (1 - p_0)^{\sum \delta_i} \lambda^{n-1} e^{-\lambda \sum t_i \delta_i} J_{p_0},$$

em que, dado  $p_c$ , por (4.8) e (4.9),  $J_{p_0} \leq \zeta$ , para algum  $\zeta > 0$ . ■

**Corolário 5.1** *Sob o modelo (4.1), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ , as distribuições a posteriori condicionais são dadas por*

$$\pi(\lambda | p_0, \mathcal{D}) \propto \lambda^{n-1} e^{-\lambda \sum t_i \delta_i} \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{(1 - \delta_i)} \quad (5.3)$$

e

$$\pi(p_0 | \lambda, \mathcal{D}) \propto p_0^{-1/2} (1 - p_0)^{\sum \delta_i} \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{(1 - \delta_i)} J_{p_0}. \quad (5.4)$$

## 5.2 Modelo Com Censura Informativa

**Proposição 5.3** *Sob o modelo (4.16), a distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$  é imprópria e é dada por*

$$J^{**}(\lambda, p_0, \eta) \propto \frac{1}{\lambda p_0^{1/2}} \left[ \frac{p_0 + \eta}{1 + \eta} \right]^{1/2} J_{p_0\eta}^{**}, \quad (5.5)$$

em que

$$J_{p_0\eta}^{**} = \sqrt{A^{**}(\varphi_{221} - \varphi_{222}) - \frac{\lambda + \eta}{\lambda p_0} \varphi_{12}},$$

$$A^{**} = \frac{\lambda + \eta}{\lambda(1 - p_0)} - \frac{\eta(\lambda + \eta)\varphi_{11}}{\lambda},$$

$$\varphi_{11} = \varphi\left(1 + \eta, 3, \frac{1 - p_0}{p_0}\right), \quad \varphi_{12} = \varphi\left(1 + \eta, 2, \frac{1 - p_0}{p_0}\right),$$

$$\varphi_{221} = \varphi\left(\eta, 1, \frac{1 - p_0}{p_0}\right) \quad e \quad \varphi_{222} = \varphi\left(1 + \eta, 1, \frac{1 - p_0}{p_0}\right).$$

**Demonstração.** Sob as expressões em (4.20)-(4.25), calcular a raiz quadrada do determinante da matriz de informação (4.19). Para verificar que é imprópria usar procedimento similar ao da demonstração da Proposição 5.1, considerando que  $\Theta^{**}$  em (4.18) é o espaço paramétrico de  $\lambda$ ,  $p_0$  e  $\eta$ . ■

**Proposição 5.4** *Sob o modelo (4.16), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$ , a função de distribuição a posteriori conjunta é própria e é dada por*

$$\begin{aligned} \pi(\lambda, p_0, \eta | \mathcal{D}) &\propto p_0^{-1/2} (1 - p_0)^{\sum \delta_i} \lambda^{n-1} \exp\left\{-\lambda \left[\sum t_i \delta_i + \eta \sum t_i\right]\right\} \\ &\times \left[\frac{\eta + p_0}{1 + \eta}\right]^{1/2} \\ &\times \eta^{\sum(1-\delta_i)} \prod_{i=1}^n [p_0 + (1 - p_0) e^{-\lambda t_i}]^{(1-\delta_i)} J_{p_0\eta}^{**}. \end{aligned} \quad (5.6)$$

**Demonstração.** Similar à demonstração da Proposição 5.2. ■

**Corolário 5.2** *Sob o modelo (4.16), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$ , as distribuições a posteriori condicionais são dadas por*

$$\pi(\lambda|p_0, \eta, \mathcal{D}) \propto \lambda^{n-1} e^{-\lambda[\sum t_i \delta_i + \eta \sum t_i]} \prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)}, \quad (5.7)$$

$$\begin{aligned} \pi(p_0|\lambda, \eta, \mathcal{D}) &\propto p_0^{-1/2} (1-p_0)^{\sum \delta_i} [\eta + p_0]^{1/2} \\ &\times \prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)} J_{p_0 \eta}^{**}, \end{aligned} \quad (5.8)$$

e

$$\pi(\eta|\lambda, p_0, \mathcal{D}) \propto \eta^{\sum (1-\delta_i)} e^{-\lambda \eta \sum t_i} \left[ \frac{\eta + p_0}{1 + \eta} \right]^{1/2} J_{p_0 \eta}^{**}. \quad (5.9)$$

## 5.3 Modelo Com Dados Ampliados

As funções (4.34) e (4.45) apresentam expressões fatoradas, em  $\lambda$  e  $p_0$ . Isso proporciona uma vantagem computacional também na inferência bayesiana. Estudos com dados simulados e com dados reais, apresentados no Capítulo 6, são feitos sob o modelo com censura não-informativa, apresentados na Subseção a seguir.

### 5.3.1 Modelo Com Censura Não-Informativa

**Proposição 5.5** *Sob o modelo (4.34), a distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$  é imprópria e é dada por*

$$J_a(\lambda, p_0) \propto \frac{1}{\lambda p_0^{1/2} (1-p_0)^{1/2}}. \quad (5.10)$$

**Demonstração.** Calcular a raiz quadrada do determinante da matriz de informação em (4.42). Para provar que é imprópria o procedimento é similar ao usado na demonstração da Proposição 5.1, levando-se em conta o espaço paramétrico de  $\lambda$  e  $p_0$ . ■

**Proposição 5.6** *Sob o modelo (4.34), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$ , a função de distribuição a posteriori conjunta é própria e é dada por*

$$\begin{aligned} \pi(\lambda, p_0; \mathbf{D}_a) &\propto \lambda^{\sum_{i=1}^n \delta_i} e^{\sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i} \\ &\times p_0^{\sum_{i=1}^n (1 - \delta_i) w_i + 1/2} (1 - p_0)^{n - \sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i + 1/2}. \end{aligned} \quad (5.11)$$

**Demonstração.** Similar à demonstração da proposição 5.2. ■

**Corolário 5.3** *Sob o modelo (4.34), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$  e  $p_0$ , as distribuições a posteriori condicionais são dadas por*

$$\lambda | p_0, \mathbf{D}_a \sim \Gamma \left[ \sum_{i=1}^n \delta_i, \sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i \right], \quad (5.12)$$

e

$$p_0 | \lambda, \mathbf{D}_a \sim \text{Beta} \left[ \sum_{i=1}^n (1 - \delta_i) w_i + 0.5, n - \sum_{i=1}^n (1 - \delta_i) w_i + 0.5 \right]. \quad (5.13)$$

### 5.3.2 Modelo Com Censura Informativa

**Proposição 5.7** *Sob o modelo (4.45), a distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$  é imprópria e é dada por*

$$J_a^{**}(\lambda, p_0, \eta) \propto \frac{1}{\lambda \eta p_0^{1/2} (1 - p_0)^{1/2}}, \quad (5.14)$$

**Demonstração.** Calcular a raiz quadrada do determinante da matriz de informação em (4.48). Para provar que é imprópria usar procedimento similar ao da demonstração da Proposição 5.1, considerando-se o espaço paramétrico de  $\lambda$ ,  $p_0$  e  $\eta$ . ■

**Proposição 5.8** *Sob o modelo (4.45), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$ , a função de distribuição a posteriori*

conjunta é própria e é dada por

$$\begin{aligned} \pi(\lambda, p_0, \eta; \mathbf{D}_a) &\propto p_0^{\sum (1-\delta_i)w_i - 1/2} (1-p_0)^{n - \sum (1-\delta_i)w_i - 1/2} \\ &\times \lambda^{n-1} e^{-\lambda \sum [1 - (1-\delta_i)w_i] t_i} \eta^{n - \sum \delta_i - 1} e^{-\lambda \eta \sum t_i}. \end{aligned} \quad (5.15)$$

**Demonstração.** Similar à demonstração da Proposição 5.2. ■

**Corolário 5.4** *Sob o modelo (4.45), considerando-se distribuição a priori de Jeffreys para os parâmetros  $\lambda$ ,  $p_0$  e  $\eta$ , as distribuições a posteriori condicionais são dadas por*

$$\lambda | p_0, \eta, \mathbf{D}_a \sim \Gamma \left[ n, \sum_{i=1}^n [1 - (1 - \delta_i) w_i] t_i + \eta \sum t_i \right], \quad (5.16)$$

$$p_0 | \lambda, \eta, \mathbf{D}_a \sim \text{Beta} \left[ \sum_{i=1}^n (1 - \delta_i) w_i + 0.5, n - \sum_{i=1}^n (1 - \delta_i) w_i + 0.5 \right], \quad (5.17)$$

e

$$\eta | \lambda, p_0, \mathbf{D}_a \sim \Gamma \left[ n - \sum \delta_i, \lambda \sum t_i \right]. \quad (5.18)$$

A vantagem do uso das funções de verossimilhança com dados ampliados é que as respectivas distribuições *a posteriori* conjunta apresenta distribuições condicionais marginais conhecidas, o que permite o uso do algoritmo de Gibbs Sampling, de fácil implementação computacional. Essa simplicidade não ocorre nas distribuições *a posteriori* dadas em (5.2) e (5.6).

## 5.4 Uma Inferência Bayesiana Empírica

Sob a Proposição 4.1 temos o estimador condicional de  $\eta$ , dado um valor fixo para  $\lambda$ . Substituindo este estimador condicional em (4.16), esta última se reduz à função de verossimilhança sob censura não-informativa em (4.1). Isto nos motivou a particionar a matriz de informação em (4.19) nos parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ . Sob essa partição, obtemos a distribuição *a priori* de Jeffreys condicional

dos parâmetros  $\lambda$  e  $p_0$ . Propomos então uma nova função de distribuição *a posteriori* condicional conjunta para os parâmetros  $\lambda$  e  $p_0$ , sob suposição de censura informativa, que obtemos e apresentamos nesta seção.

**Proposição 5.9** *Sob o modelo (4.16), dado  $\eta$ , a matriz de informação de Fisher condicional para os parâmetros  $\lambda$  e  $p_0$  é dada por*

$$IF_p(\lambda, p_0, \eta) = \begin{bmatrix} I_{\lambda\lambda} & I_{\lambda p_0} \\ I_{p_0\lambda} & I_{p_0 p_0} \end{bmatrix}, \quad (5.19)$$

em que

$$I_{\lambda\lambda} = \frac{n(1-p_0)}{\lambda^2} \left[ \frac{1}{1-p_0} - \eta\varphi\left(1 + \eta, 3, \frac{1-p_0}{p_0}\right) \right], \quad (5.20)$$

$$I_{\lambda p_0} = I_{p_0\lambda} = \frac{n\eta}{\lambda p_0} \left[ \varphi\left(1 + \eta, 2, \frac{1-p_0}{p_0}\right) \right], \quad (5.21)$$

$$I_{p_0 p_0} = \frac{n\eta}{(1-p_0)p_0} \left[ \varphi\left(\eta, 1, \frac{1-p_0}{p_0}\right) - \varphi\left(1 + \eta, 1, \frac{1-p_0}{p_0}\right) \right]. \quad (5.22)$$

**Demonstração.** Sob as expressões (4.20)-(4.25), particionar a matriz de informação (4.19) nos parâmetros  $\lambda$  e  $p_0$ . ■

**Proposição 5.10** *Sob o modelo (4.1) e matriz de informação condicional em (5.19), a distribuição a priori de Jeffreys condicional para os parâmetros  $\lambda$  e  $p_0$  é imprópria e é dada por*

$$J_{cond}(\lambda, p_0) \propto \frac{1}{\lambda p_0^{1/2}} J_{pp_0}, \quad (5.23)$$

em que

$$J_{pp_0} = \sqrt{\left(\frac{1}{1-p_0} - \eta\varphi_{11}\right) (\varphi_{221} - \varphi_{222}) - \frac{\eta}{p_0} \varphi_{12}^2},$$

$$\varphi_{11} = \varphi\left(1 + \eta, 3, \frac{1-p_0}{p_0}\right), \quad \varphi_{12} = \varphi\left(1 + \eta, 2, \frac{1-p_0}{p_0}\right),$$

$$\varphi_{221} = \varphi\left(\eta, 1, \frac{1-p_0}{p_0}\right) \quad e \quad \varphi_{222} = \varphi\left(1 + \eta, 1, \frac{1-p_0}{p_0}\right).$$



**Demonstração.** Sob as expressões (5.20)-(5.22), calculando a raiz quadrada do determinante da matriz (5.19), temos que a distribuição *a priori* de Jeffreys condicional para os parâmetros  $\lambda$  e  $p_0$  é dada por

$$J_{cond}(\lambda, p_0) \propto \frac{1}{\lambda p_0^{1/2}} \sqrt{\left(\frac{1}{1-p_0} - \eta \varphi_{11}\right) (\varphi_{221} - \varphi_{222}) - \frac{\eta}{p_0} \varphi_{12}^2},$$

em que, dado  $\eta$ ,  $J_{pp_0}$  só depende de  $p_0$ . Para verificar que a distribuição (5.23) é imprópria, o procedimento é similar ao da demonstração da Proposição 5.1. ■

**Proposição 5.11** *Sob o modelo (4.1), considerando-se distribuição a priori de Jeffreys condicional (5.23) para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ , a função de distribuição a posteriori conjunta é própria e é dada por*

$$\begin{aligned} \pi(\lambda, p_0 | \mathcal{D}) &\propto p_0^{-12} (1-p_0)^{\sum \delta_i} \lambda^{n-1} e^{-\lambda[\sum t_i \delta_i + \eta \sum t_i]} \\ &\quad \times \prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)} J_{pp_0}. \end{aligned} \quad (5.24)$$

**Demonstração.** Similar à da Proposição 5.2. ■

**Corolário 5.5** *Sob o modelo (4.1), considerando-se distribuição a priori de Jeffreys condicional (5.23) para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ , as funções de distribuição a posteriori condicionais para  $\lambda$  e  $p_0$  são dadas por*

$$\pi(\lambda | p_0, \mathcal{D}) \propto \lambda^{n-1} e^{-\lambda[\sum t_i \delta_i + \eta \sum t_i]} \prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)} \quad (5.25)$$

e

$$\pi(p_0 | \lambda, \mathcal{D}) \propto p_0^{-1/2} (1-p_0)^{\sum \delta_i} \prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)} J_{pp_0}. \quad (5.26)$$

Para eliminar o fator  $\prod_{i=1}^n [p_0 + (1-p_0) e^{-\lambda t_i}]^{(1-\delta_i)}$ , e facilitar o trabalho computacional, multiplicamos a função de distribuição *a posteriori* conjunta (5.24) pelo fator

$$\prod_{i=1}^n \left\{ \rho_i^{w_i} (1-\rho_i)^{(1-w_i)} \right\}^{(1-\delta_i)}, \quad (5.27)$$

função de verossimilhança da variável latente dada em (4.31)-(4.32), relacionada aos indivíduos censurados. Este procedimento é semelhante ao usado em (4.33). A diferença é que em (4.33) a função de verossimilhança (4.1) é multiplicada pelo fator (5.27), enquanto que neste caso este mesmo fator está sendo multiplicado por uma função de distribuição *a posteriori*. Sob este procedimento, dado  $\eta$ , a função de distribuição conjunta *a posteriori* para  $\lambda$  e  $p_0$  em (5.24), dada na sua versão ampliada, é expressa pela proposição a seguir.

**Proposição 5.12** *Considere-se o modelo (4.1) e a distribuição a priori de Jeffreys condicional (5.23) para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ . Sob o conjunto de dados  $\mathcal{D}_a$  em (4.35), a função de distribuição a posteriori conjunta é própria e é dada por*

$$\begin{aligned} \pi(\lambda, p_0 | \mathcal{D}; \mathcal{D}_a) &\propto p_0^{\sum w_i(1-\delta_i)-1/2} (1-p_0)^{n-\sum w_i(1-\delta_i)} \\ &\times \lambda^{n-1} e^{-\lambda[(1+\eta)\sum t_i - \sum t_i w_i(1-\delta_i)]} J_{pp_0}. \end{aligned} \quad (5.28)$$

**Demonstração.** Multiplicar a função (5.24) pelo fator

$$\prod_{i=1}^n \left\{ \rho_i^{w_i} (1-\rho_i)^{(1-w_i)} \right\}^{(1-\delta_i)},$$

em que a distribuição de  $W_i$  é dada por (4.31)-(4.32). Para verificar que é própria o procedimento é semelhante ao da demonstração da Proposição 5.2. ■

**Corolário 5.6** *Considere-se o modelo (4.1) e a distribuição a priori de Jeffreys condicional (5.23) para os parâmetros  $\lambda$  e  $p_0$ , dado  $\eta$ . Sob o conjunto de dados  $\mathcal{D}_a$  em (4.35), as funções de distribuição a posteriori condicionais para  $\lambda$  e  $p_0$  são dadas por*

$$\lambda | p_0, \mathcal{D}; \mathcal{D}_a \sim \Gamma \left[ n, (1+\eta) \sum t_i - \sum t_i w_i (1-\delta_i) \right], \quad (5.29)$$

e

$$\pi(p_0 | \lambda, \mathcal{D}; \mathcal{D}_a) \propto p_0^{\sum w_i(1-\delta_i)-1/2} (1-p_0)^{n-\sum w_i(1-\delta_i)} J_{pp_0}. \quad (5.30)$$

Nas funções de distribuição *a posteriori* condicionais para  $\lambda$  e  $p_0$  em (5.29)-(5.30), a implementação do MCMC (Monte Carlo with Markov chains), requer um valor atribuído previamente para  $\eta$  em cada iteração. Nos estudos com dados simulados que realizamos no Capítulo 6 a seguir, este valor de  $\eta$  é dado a partir dos valores de entrada, através das expressões em (3.20) e (3.21)-(3.22), Teorema 3.2. Um detalhe importante neste estudo é que, para cada amostra de tamanho  $n$  simulada, é usado o mesmo valor de  $\eta$  em todas as iterações.

No mesmo Capítulo 6, após o estudo com dados simulados, realizamos o estudo com o conjunto de dados reais E1690 (Chen *et al.*, 2002), em que a implementação do MCMC para as distribuições *a posteriori* condicionais (5.29) e (5.30) ocorreu sob dois procedimentos. No primeiro, foi necessário um uso prévio do conjunto E1690 para, sob a função (4.16) (censura informativa), obter a estimativa de máxima verossimilhança de  $\eta$ . Com o uso dessa estimativa como valor dado para  $\eta$  é implementado o MCMC. Um detalhe importante é que neste caso a amostra E1690 é usada duas vezes.

No segundo procedimento, é obtida previamente a estimativa de máxima verossimilhança de  $\eta$  a partir de um outro conjunto de dados, o E1684 (Chen *et al.*, 2002). Essa estimativa, feita também sob a função (4.16), é usada como valor dado para  $\eta$  na implementação do MCMC. Neste caso, a amostra de interesse, o conjunto E1690, é usada apenas uma vez, na própria implementação do MCMC, em que é assumido um valor prévio para  $\eta$ .

Um procedimento similar a este segundo é visto em Chen *et al.* (2002), em que, a partir do conjunto de dados E1684 é elicitada a distribuição *a priori* dos parâmetros de interesse, que é usada na implementação do MCMC, tendo como amostra de interesse o conjunto E1690. Neste texto também foi usada uma segunda amostra, o E1684, para obter informações *a priori* a respeito dos parâmetros de interesse. Esta discussão nos motivou a considerar este MCMC como uma inferência bayesiana empírica. Como referência para inferência bayesiana empírica citamos apenas Bernardinelli & Montomoli (2007).

## Capítulo 6

# Estudo de Simulação e Aplicação com Dados Reais

Vale lembrar que no modelo (3.4),  $S_p(y) = p_0 + (1 - p_0)e^{-\lambda y}$ , os parâmetros de interesse são  $\lambda$  e  $p_0$ , em que  $\lambda > 0$  e  $0 \leq p_0 < p_c$ . Quanto ao parâmetro  $\eta$ , atribuímos a este um interesse secundário. Deixamos como proposta de trabalhos futuros um estudo com dados simulados em que  $\eta$  também seja parâmetro de interesse. Nesta tese é usado como um recurso auxiliar (censura informativa), que resolve o problema da sua presença nas expressões (4.13)-(4.15). Na Seção 6.5, com réplicas de dados simulados, desenvolvemos estudos de inferência sobre  $\lambda$  e  $p_0$ , usando abordagens clássica e bayesiana objetiva com distribuições *a priori* de Jeffreys. O estudo envolve basicamente a teorização vista nos Capítulos 4 e 5. Sob um número estipulado de réplicas de dados simulados calculamos a média das estimativas pontuais, estimativas do comprimento médio dos intervalos, e estimativas das respectivas probabilidades de cobertura. Na abordagem clássica usamos intervalos de confiança de 95% e na bayesiana intervalos de credibilidade de 95%. Sob estes estudos é feita uma varredura no espaço paramétrico de  $p_0$ .

Na Seção 6.6 consideramos um conjunto de dados reais, para o qual obtemos as estimativas pontuais de  $\lambda$  e  $p_0$ , com estimativas do comprimento dos intervalos. Todos os resultados numéricos e gráficos deste capítulo são obtidos com o uso

de programas em código R (R Development Core Team, 2009). Um detalhe importante é que os gráficos são apresentados em cores diferenciadas, e portanto devem ser vistos na interface do PDF ou em impressão a cores.

## 6.1 Algoritmo de Simulação de Dados

Para simular um conjunto de dados implementamos um algoritmo de geração, que descrevemos abaixo. Sob os modelos (3.4) e (3.14), o objetivo é simular observações das variáveis  $T$  e  $\Delta$ ,  $T$  definida em (2.42) e  $\Delta$  em (2.43). Descrevemos o algoritmo de simulação de dados, particularizado para os seguintes valores de entrada:  $\lambda = 0.25$ ,  $p_0 = 0.30$ ,  $n = 400$  e  $p_c = 0.50$ . Neste caso, simulamos uma amostra de tamanho  $n = 400$ , de uma população em que a probabilidade de cura é de 30%, e entre os não curados a taxa de falha é de 25%. Nessa amostra 50% dos indivíduos são censurados.

- Valores de entrada:  $\lambda = 0.25$ ,  $p_0 = 0.30$ ,  $p_c = 0.5$  (50% de censura) e  $n = 400$ . O valor de  $\lambda_c$ , em (3.14), pode ser obtido por (3.20).
- Sob a condição em (3.13), definir  $Y_0 = Y|M = 0$ .
- Sob a distribuição condicional em (3.12), definir  $Y_1 = Y|M = 1$ .
- Sob o modelo (3.4) definir

$$Y = Y_0 I_{\{m=0\}}(m) + Y_1 I_{\{m=1\}}(m), \quad (6.1)$$

em que  $I_{\{m\}}(\cdot)$  é a função indicadora,  $m = 0, 1$ .

- Gerar amostra de tamanho  $n$ , de  $M \sim Bernoulli(1 - p_0)$ .
- Gerar amostra de tamanho  $n$  de  $Y_0$ , em que  $Y_{0i} = \infty$ ,  $i = 1, \dots, n$ .
- Gerar amostra de tamanho  $n$ , de  $Y_1 \sim exp(\lambda)$ .
- Sob a equação (6.1), obter amostra de tamanho  $n$  de  $Y$ .
- Gerar amostra de tamanho  $n$ , de  $X \sim exp(\lambda_c)$ .

- Obter obter amostra de tamanho  $n$  de  $T$ ,  $T_i = \min \{Y_i, X_i\}$ ,  $i = 1, \dots, n$ .
- Obter amostra de tamanho  $n$  de  $\Delta$ , em que
  - Se  $Y_i \leq X_i$ ,  $\Delta_i = 1$ ,
  - Se  $Y_i > X_i$ ,  $\Delta_i = 0$ .

Sobre este algoritmo devemos fazer uma observação, dado que estamos usando o código R (R Development Core Team, 2009). Neste código podemos obter amostras de tamanho  $n$  de  $Y_0$  digitando

```
> Y0<-rep(Inf,n),
```

veja Apêndice C. Na ausência de um comando equivalente, uma opção é usar, por exemplo,  $Y_{0i} = 10^{20}$ ,  $i = 1, \dots, n$ , um valor suficientemente grande para ser sempre censurado, e então válido como uma forma de representar o infinito.

Na simulação de cada conjunto de dados consideramos  $\lambda = 0.25$  e  $p_c = 0.5$  (50% de indivíduos censurados). Sob abordagem clássica usamos  $n = 40$  e  $n = 400$  como tamanho das amostras, e apenas amostras de tamanho  $n = 40$  na abordagem bayesiana (veja Seções 6.2 e 6.3 a seguir).

Na Seção 6.4 são listados e definidos os procedimentos usados nos estudos com dados simulados e com dados reais. Na definição de cada procedimento, é dito se a abordagem é clássica ou bayesiana, e se a suposição de censura é informativa ou não informativa. Na abordagem clássica é dito se os intervalos são obtidos a partir da matriz de informação de Fisher ou se a partir da matriz de informação observada.

## 6.2 Metodologia Sob Abordagem Clássica

Na abordagem clássica usamos  $n = 40$  e  $n = 400$ , e o conjunto de valores

$$\{0.01, 0.02, \dots, 0.48, 0.49\} \quad (6.2)$$

para  $p_0$ . Neste conjunto temos um *grid*, ou malha, de valores, varrendo o espaço paramétrico numa variação uniforme de 1%.

Sob cada procedimento (definidos na Seção 6.4 a seguir), dado o tamanho da amostra, para cada valor de  $p_0$  em (6.2) usamos 10000 réplicas (amostras), para as quais são obtidas as estimativas pontuais dos parâmetros de interesse,  $\lambda$  e  $p_0$ , e respectivos intervalos de confiança assintóticos de 95% . Em cada amostra e para cada parâmetro, um comando contador do programa em código R determina se o verdadeiro valor do parâmetro é coberto ou não pelo respectivo intervalo de confiança (veja Definição 1.1).

Para cada parâmetro é calculada a média das 10000 estimativas pontuais, a porcentagem de coberturas obtida do contador de coberturas e as médias dos limites inferior e superior dos intervalos. Com relação aos limites dos intervalos, a medida de interesse é o comprimento médio dos intervalos, que mensuramos como a diferença entre as médias dos dois limites. Os gráficos resultantes, em duas dimensões, apresentam no eixo da abscissa os valores de  $p_0$  em (6.2). No eixo das ordenadas temos as médias frequentistas dos valores das estimativas das medidas de interesse, tais como cobertura, comprimento do intervalo e estimativas pontuais.

### 6.3 Metodologia Sob Abordagem Bayesiana

Na abordagem bayesiana, devido ao custo computacional, estabelecemos duas diferenças com relação ao caso clássico. A primeira é que usamos o conjunto de valores

$$\{0.01, 0.04, 0.07, \dots, 0.46, 0.49\} \quad (6.3)$$

para  $p_0$ , varrendo o espaço paramétrico numa variação uniforme de 3%. A segunda é que, para cada valor de  $p_0$  em (6.3) usamos apenas 1000 réplicas de tamanho  $n = 40$ , contra 10000 no caso clássico. Estas duas restrições não chegam a comprometer a visualização gráfica do comportamento médio das inferências.

Uma terceira limitação na abordagem bayesiana é que, dado  $p_0$  em (6.3), sob cada réplica, o MCMC (Monte Carlo with Markov chains) foi implementado com apenas uma cadeia. Não verificamos a convergência por meio dos métodos usuais (Paulino *et al.*, 2003), mas, sob as 1000 réplicas, usamos a média das estimativas pontuais, e os resultados destas são similares aos da abordagem clássica, o que consideramos razoável, dado que estamos usando distribuições *a priori* de Jeffreys, que é não informativa. É esta similaridade com a abordagem clássica que usamos como um meio empírico de validar os resultados da inferência bayesiana (veja gráfico das estimativas pontuais).

São obtidas também as médias dos limites inferior e superior dos intervalos de credibilidade de 95%, sob os quais, a medida de interesse é o comprimento médio dos intervalos, dados pela diferença entre as médias dos dois limites. Um comando contador verifica se o intervalo de credibilidade contém ou não o valor verdadeiro do parâmetro, e assim é estimada a porcentagem de cobertura dos intervalos de credibilidade, sob as 1000 réplicas. Assim como na abordagem clássica, e, para efeito de comparação, os gráficos resultantes, em duas dimensões, apresentam no eixo da abscissa os valores de  $p_0$  em (6.3). No eixo das ordenadas as médias dos valores das estimativas das medidas de interesse, tais como porcentagem de cobertura dos intervalos de credibilidade de 95%, comprimento médio dos mesmos e média das estimativas pontuais.

Para cada conjunto de dados simulados, na inferência dos parâmetros foram obtidas 5010 iterações de MCM, com salto de tamanho 3 para amenizar a correlação entre as iterações. Ou seja, de 5010 iterações aceitas, são consideradas  $1670 = \frac{5010}{3}$  iterações para cada parâmetro, das quais são descartadas as primeiras 334, totalizando  $1336 = 1670 - 334$  amostras de MCMC para cada parâmetro. É destas 1336 amostras que é obtida a estimativa pontual de MCMC, bem como o intervalo de credibilidade de 95%, e respectiva cobertura.



## 6.4 Apresentação Gráfica e em Tabela

Para proporcionar uma análise comparativa bastante simplificada dos procedimentos de inferência usamos uma notação padronizada para as legendas. Estas legendas referem-se às funções de verossimilhança (4.1), (4.16), (4.34) e (4.45), e respectivas matriz de informação, observada ou de Fisher. Abaixo, a partir da Figura 6.4.1, definimos cada procedimento, com respectivas notações das legendas que serão usadas nos gráficos e tabelas deste capítulo.

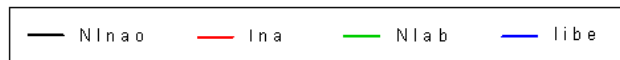


Figura 6.1. *Legendas utilizadas.*

### 6.4.1 Notações dos Procedimentos

- **Dados não-ampliados:** Conjunto de dados definido em (4.2), usado nas funções (4.1) e (4.16).
- **Dados ampliados:** Conjunto de dados definido em (4.35), usado nas funções (4.34) e (4.45).
- **emv:** abreviação para estimativa de máxima verossimilhança.

- **Caso Não Informativo**

1. **Nlnao:** (Não-informativa, dados não-ampliados, matriz de informação observada) Função de verossimilhança (4.1), emv de  $\lambda$  e  $p_0$ , e intervalos de confiança assintóticos de 95%, obtidos da respectiva matriz de informação observada.
2. **Nlnae:** (Não-informativa, dados não-ampliados, matriz de informação de Fisher) Função de verossimilhança (4.1), emv de  $\lambda$  e  $p_0$ , e intervalos de confiança assintóticos de 95%, obtidos da respectiva matriz de informação de Fisher, dada pelas expressões em (4.13)-(4.15). Este procedimento é usado somente com dados reais na Seção 6.6, e para tal é necessário um valor dado para  $\eta$ .
3. **NIab:** (Não-informativa, dados ampliados, bayesiana) Estimativas bayesianas a partir da distribuição conjunta *a posteriori* (5.11), com intervalos de credibilidade de 95%.

- **Caso Informativo**

1. **Ina:** (Informativa, dados não-ampliados e matriz de informação de Fisher) Função de verossimilhança (4.16), emv de  $\lambda$  e  $p_0$  e  $\eta$ , e intervalos de confiança assintóticos de 95%, obtidos a partir da respectiva matriz de informação de Fisher, dada em (4.19).
2. **Ibe:** (Informativa, dados ampliados, inferência bayesiana empírica) Inferência bayesiana a partir da distribuição condicional *a posteriori* (5.28) e intervalos de credibilidade de 95%. Este procedimento requer um valor dado (conhecido) para o parâmetro perturbador  $\eta$ .

## 6.5 Estudos com Dados Simulados

Nas Subseções 6.5.1 e 6.5.2 a seguir, para os parâmetros  $\lambda$  e  $p_0$ , respectivamente, apresentamos os resultados das simulações em representações gráficas. Em cada subseção temos uma figura com seis quadros gráficos, dispostos em duas linhas e três colunas. Na primeira linha (quadros gráficos da parte superior) temos resultados para amostras de tamanho  $n = 40$ , e na segunda, resultados para amostras de tamanho  $n = 400$  (quadros gráficos da parte inferior). Na primeira coluna temos a porcentagem estimada de cobertura dos intervalos de confiança assintóticos de 95%, e de 95% de credibilidade. Na segunda, temos o comprimento médio destes intervalos, e na terceira a média das estimativas pontuais dos parâmetros. Na parte superior de cada figura temos as legendas.

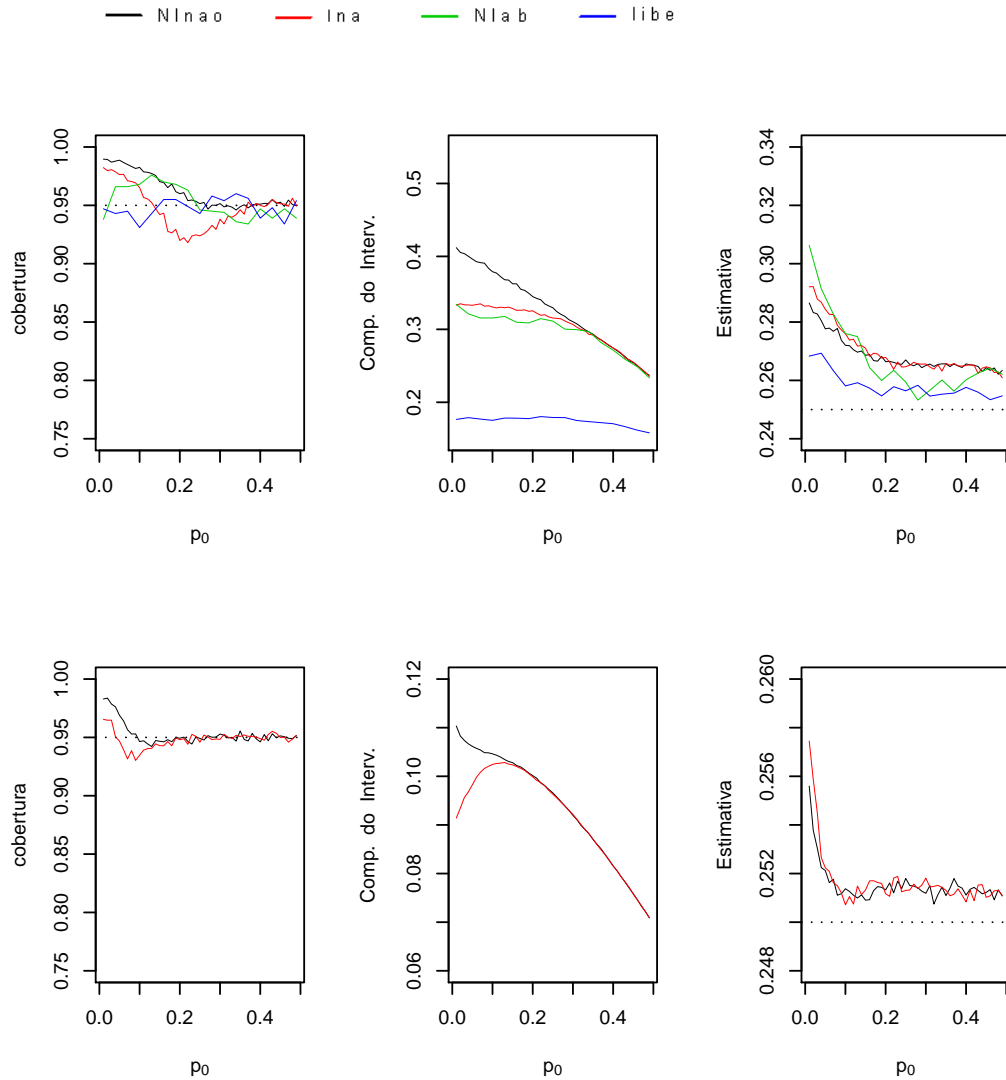
6.5.1 Inferência Sobre  $\lambda$ 

Figura 6.2. Inferências sobre  $\lambda$ : Na primeira linha temos dados simulados de tamanho  $n = 40$ , e na segunda, de tamanho  $n = 400$ ; nas colunas, da esquerda para a direita, plot da porcentagem de coberturas, do comprimento médio dos intervalos e da média frequentista das estimativas pontuais.

De (3.16) temos que  $0 < p_0 < p_c$ . Portanto, dentro das nossas condições iniciais de simulação (veja primeiro parágrafo da Seção 6.2), o espaço paramétrico de  $p_0$  é definido pelo intervalo de variação  $0 < p_0 < 0.50$ . Analizando os gráficos apresentados na Figura 6.2, na região à direita deste espaço paramétrico de  $p_0$ , nota-se que os procedimentos **NInao** e **Ina** são equivalentes. Se compararmos

sob os dois tamanhos de amostra,  $n = 40$  e  $n = 400$ , essa região à direita é menor para  $n = 40$ , e os gráficos da segunda linha ( $n = 400$ ) sugerem que o limite inferior dessa região recua para a esquerda, em direção à 0, com o aumento do tamanho da amostra.

Este fato leva à conjectura de que, com dados reais, para valores de  $p_0$  nessa região, é irrelevante a escolha da suposição de censura, sob a abordagem clássica. Esta conjectura é reforçada com os resultados para dados reais apresentados na última seção. Os resultados para  $n = 400$  sugerem que essa região aumenta com o tamanho  $n$  da amostra. Para  $p_0$  na região oposta, à esquerda, o procedimento **Ina** apresenta menor comprimento de intervalo, mas maior vício de estimativa pontual. Sob o procedimento **NInao**, o maior comprimento de intervalo pode explicar o excesso de cobertura nessa região, que é adjacente ao limite inferior do espaço paramétrico de  $p_0$ .

Para amostras de tamanho  $n = 40$ , numa região de transição, entre estas duas citas acima, nota-se uma queda na porcentagem estimada de cobertura dos intervalos sob o procedimento **Ina**. A nossa conjectura é de que existe nessa região intermediária uma competição entre os dois mecanismos de censura definidos em (2.49 e (3.24), não detectada sob censura não-informativa, e que, sob censura informativa se evidencia na forma de queda na probabilidade estimada de cobertura. Sob esta hipótese, a explicação de como essa competição provoca queda na cobertura nessa região de transição seria assunto para investigação em trabalhos futuros. A partir dessa região de transição, em direção à  $p_0 = 0$ , o procedimento **NInao** já apresenta crescimento do comprimento do intervalo, com relação à metodologia **Ina**. Com o aumento do tamanho da amostra, essa região de transição recua para a esquerda, e tende a se confundir com a região adjacente ao limite inferior do espaço paramétrico, e os gráficos mostram que essa queda na probabilidade estimada de cobertura tende a diminuir. Se essa queda na diferença entre a probabilidade esperada de 95% e a probabilidade estimada tende a diminuir para 0 também é caso para investigação futura. Sob abordagem clássica analisamos três regiões distintas do espaço paramétrico de  $p_0$ : a região adjacente ao limite inferior (à esquerda), a intermediária (do meio) e a região

adjacente ao limite superior (à direita).

Analisamos então os procedimentos bayesianos, estes apenas com amostras de tamanho  $n = 40$ . Na região à esquerda, o procedimento **NIab** apresenta considerável excesso na probabilidade estimada de cobertura, apesar de o comprimento médio dos intervalos ser ligeiramente menor, se comparado ao do procedimento **Ina**. Na região à direita a probabilidade estimada de cobertura é ligeiramente menor que a esperada, com comprimento médio dos intervalos equivalente aos dos procedimentos clássicos. Na região à esquerda apresenta maior vício na estimativa pontual, enquanto que nas regiões do meio e à direita, o segundo menor vício.

Os melhores resultados do procedimento **Iibe** são devidos ao fato de ser atribuído um valor conhecido para  $\eta$ , obtido por (3.19) e (3.21), dados os valores de entrada de  $p_c$  e  $p_0$ . Este valor conhecido de  $\eta$  constitui considerável informação *a priori* nas inferências sobre  $\lambda$  e  $p_0$ . Nos estudos com dados reais vistos na Seção 6.6, usamos a estimativa de máxima verossimilhança de  $\eta$  como o valor conhecido.

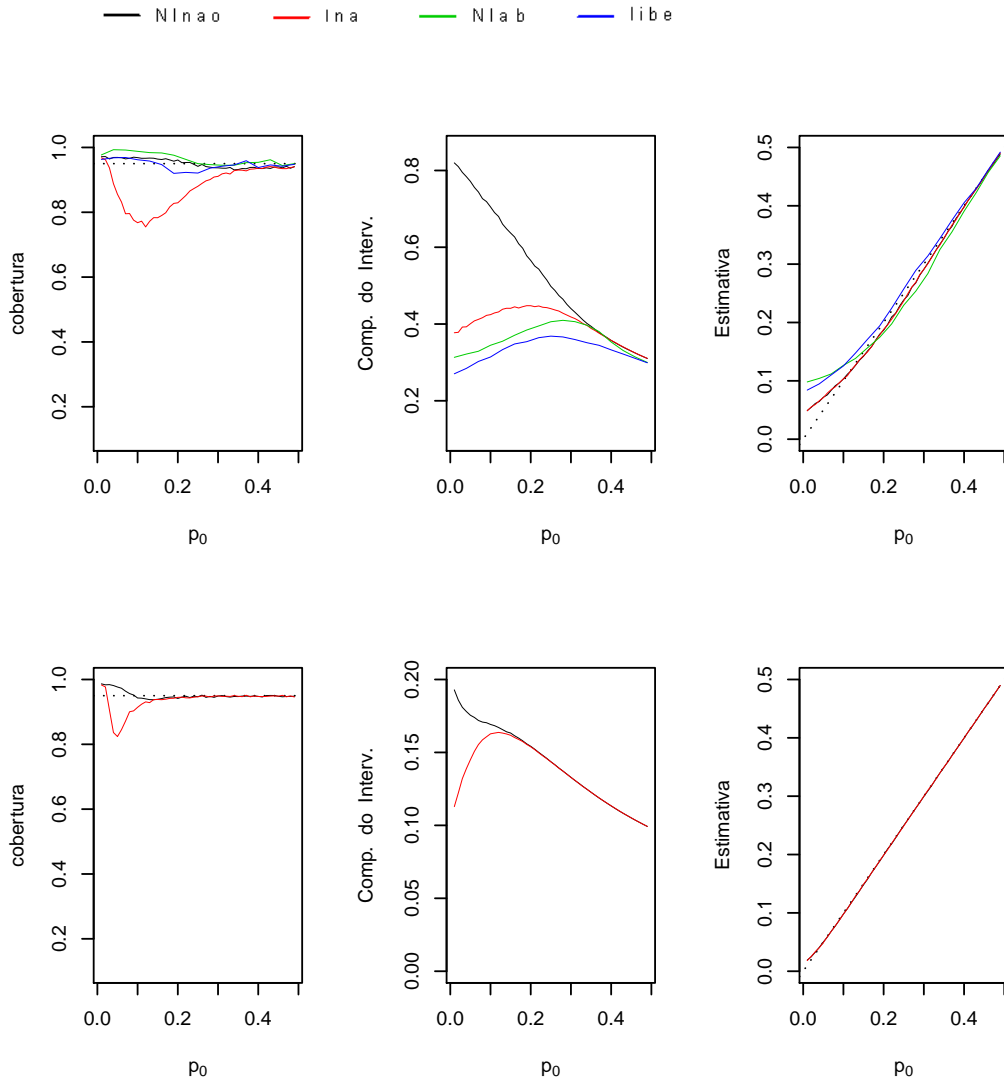
6.5.2 Inferência Sobre  $p_0$ 

Figura 6.3. Inferências sobre  $p_0$ : Na primeira linha temos dados simulados de tamanho  $n = 40$ , e na segunda, de tamanho  $n = 400$ ; nas colunas, da esquerda para a direita, plot da porcentagem de coberturas, do comprimento médio dos intervalos e da média frequentista das estimativas pontuais.

Para amostras de tamanho  $n = 40$ , os procedimentos **NInao**, **NIab** e **Iibe** apresentam probabilidade estimada de cobertura similares. No entanto, o procedimento **NIab** produz comprimento de intervalos significativamente menores nas regiões adjacente ao limite inferior e do meio, do espaço paramétrico de  $p_0$ , enquanto que sob o procedimento **Iibe** o comprimento dos mesmos é menor em

todo este espaço paramétrico. Com relação às estimativas pontuais, os procedimentos **NInao** e **Ina** apresentam menor vício, se comparadas com os outros dois procedimentos. Mas, de uma forma geral, os resultados do procedimento **Iibe** são mais precisos, consequência da informação *a priori* dada pelo parâmetro  $\eta$ . Sob o procedimento **Ina**, a queda da probabilidade estimada da cobertura na região esquerda do espaço paramétrico de  $p_0$  é maior, se comparada com a mesma para o parâmetro  $\lambda$ . Em amostras de tamanho  $n = 400$ , em regiões distantes do limite inferior do espaço paramétrico de  $p_0$ , os procedimentos clássicos são equivalentes. Um detalhe importante, é que, sob os dois procedimentos clássicos, as médias das estimativas pontuais coincidem em todo o espaço paramétrico, como o mesmo comportamento de maior vício na região adjacente à  $p_0 = 0$ .

### 6.5.3 Algumas Considerações

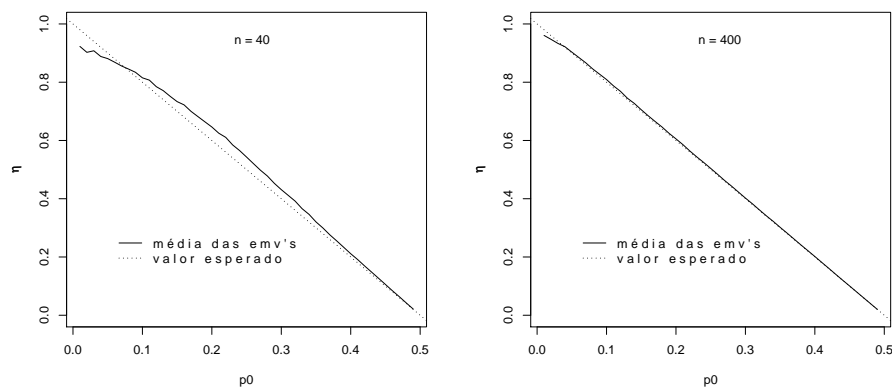


Figura 6.4. Média das estimativas de máxima verossimilhança de  $\eta$  (linha cheia) e valor esperado de  $\eta$  (linha pontilhada), para amostras de tamanho  $n = 40$  e  $n = 400$ , obtida destas simulações, sob o procedimento **Ina**.

Um detalhe que devemos observar é o problema de como obter o verdadeiro valor de  $\eta$ , quando se trata de dados reais. A Figura 6.4 mostra que o vício da sua emv sob o procedimento **Ina** é pequeno, mesmo para amostras de tamanho  $n = 40$ . Isso nos motiva, nas inferências com dados reais analisadas na Seção 6.6 a seguir, a usar o emv de  $\eta$  nos procedimentos **NInae** e **Iibe**. Mas falta

implementar os estudos com dados simulados para amostras com altos valores de  $p_c$ , acima de 50%. Nestes caso  $\eta$  tem valor máximo maior que 1, e não temos como fazer conjecturas quanto ao comportamento, ou vício, do seu estimador de máxima verossimilhança, dado em (4.26), quando  $p_c$  é grande.

Uma segunda observação é que, sob as duas suposições de censura, quanto ao uso das funções de verossimilhança com dados ampliados em (4.34) e (4.45) (e suas respectivas matriz de informação de Fisher), as inferências pontuais para ambos os parâmetros são equivalentes às obtidas pelo procedimento **NIao**, em todo o espaço paramétrico de  $p_0$ . Porém, ao longo deste mesmo espaço paramétrico de  $p_0$ , o gráfico para ambos os parâmetros mostra comprimentos de intervalos muito pequenos, o que provoca uma queda considerável na cobertura. As duas suposições de censura se mostraram equivalentes nos resultados. O péssimo resultado das coberturas foi o motivo de não termos apresentado aqueles resultados graficamente nesta seção. Descartamos então as funções em (4.34) e (4.45), e respectivas matriz de informação de Fisher, como procedimento em inferências, mas evidenciamos a necessidade de estudos mais detalhados, que verifiquem concretamente aqueles resultados, isso como proposta de trabalho futuro.

Na conclusão desta seção de estudos com dados simulados devemos estar atento à uma última observação. Em estudos com dados simulados, neste caso visando comparação de metodologias de inferência, uma orientação que nos ocorre de imediato é, sob cada valor de  $p_0$  em (6.2) e cada um dos quatro procedimentos, usar as mesmas réplicas de dados simulados. Ou seja, as mesmas 10000 réplicas deveriam ser usadas sob os procedimentos **NIao**, **Ina**, **NIab** e **Iibe**.

No entanto, nota-se nos programas em R apresentados nos Apêndices C.2 e C.3, que no uso de cada metodologia, para cada tamanho  $n$  da amostra e para cada de valor de  $p_0$ , foram geradas novas 10000 réplicas. O objetivo foi deversificar a quantidade total de conjunto de dados, de 490000 réplicas para cada metodologia. Os resultados gráficos parecem indicar que o não uso das mesmas réplicas não compromete os resultados aqui apresentados.



## 6.6 Dados Reais: Melanoma Maligno E1690

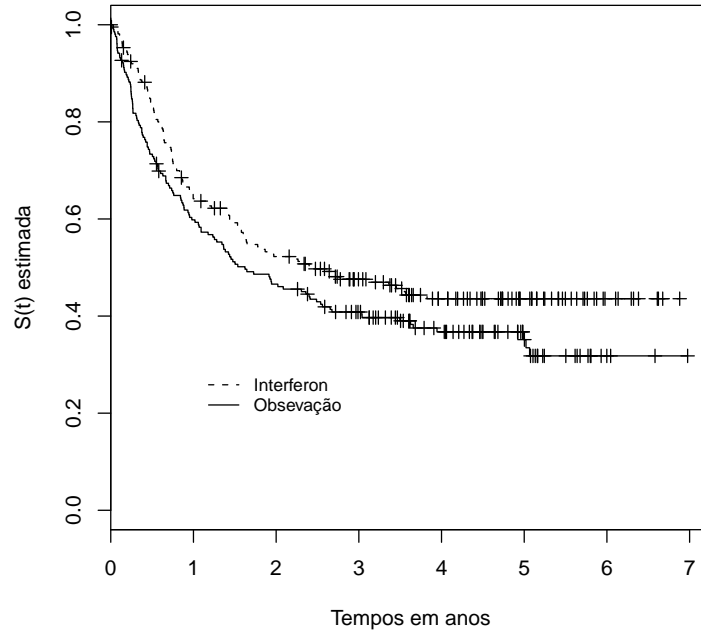


Figura 6.5. Estimativas de Kaplan-Meier sob os dois grupos dos dados E1690, em que o símbolo + indica a ocorrência de censura.

Em Chen *et al.* (2002) é usado o conjunto de dados de melanoma maligno E1690, obtido de um estudo conduzido pela Eastern Cooperative Oncology Group. São considerados 427 pacientes do grupo E1690, divididos em dois grupos: o grupo de observação, com 212 pacientes, recebeu o tratamento de quimioterapia convencional, e o grupo tratamento, com 215 pacientes, recebeu altas doses de interferon alpha-2b. A variável de interesse nos dois grupos é o tempo até a reicidência da doença, e a variável indicadora dos grupos é usada como covariável.

Para cada grupo, é apresentada a estimativa de máxima verossimilhança da fração de cura,  $p_0$ , sob o modelo (2.9), com distribuição de Poisson para os fatores de risco e distribuição exponencial para o tempo de falha. Também é apresentado os resultados e análises da inferência bayesiana para a fração de cura, sob versões de três diferentes modelos, com distribuições *a priori* elicidadas. As estimativas são comparadas com as estimativas de máxima verossimilhança. Para

mais detalhes, veja Chen *et al.* (2002). Neste trabalho, também consideramos os dois grupos e a mesma variável de interesse, o tempo até a reincidência da doença, com o detalhe de que usamos o modelo de mistura padrão exponencial e tempo de censura também com distribuição exponencial.

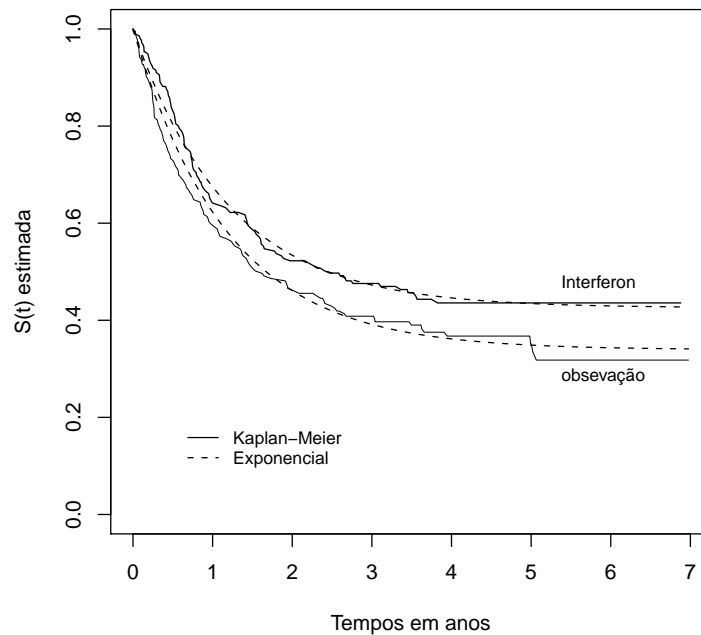


Figura 6.6. Ajuste das estimativas de Kaplan-Meier ao modelo de mistura padrão exponencial (3.4), modelo de Barkson-Gage, sob os emv's de  $\lambda$  e  $p_0$  com procedimento **Ina**.

Na Figura 6.5, as estimativas de Kaplan-Meier (Kaplan & Meier, 1958) para o tempo de reincidência mostra que os indivíduos do grupo que recebeu altas doses de interferon têm maior probabilidade de sobreviver até o tempo  $t$ , e que, em ambos os grupos, as censuras ocorrem com maior intensidade após dois anos do início do experimento. Em Chen *et al.* (2002), para cada modelo usado no texto, os resultados das inferências sobre  $p_0$  são apresentados para cada grupo, separadamente (veja Tabelas 1, 2 e 3 daquele texto), assim como os gráficos das estimativas de Kaplan-Meier e ajustes sob os modelos. Neste trabalho seguimos esse mesmo procedimento, inclusive nas apresentações gráficas das Figuras 6.5, 6.6 e 6.7.

Na Figura 6.6 temos, para os dois grupos, o ajuste da função de sobrevivência sob a estimativa de Kaplan-Meier com as respectivas estimativas pelo modelo de mistura padrão exponencial. As estimativas da função de sobrevivência pelo modelo de mistura padrão exponencial são obtidas sob o modelo (3.4), dado pela função  $S(t) = p_0 + (1 - p_0)e^{-\lambda t}$ , em que os valores de  $\lambda$  e  $p_0$  são dados pelas respectivas estimativas de máxima verossimilhança (emv), sob a condição **Ina**. Esta forma de estimar a função de sobrevivência, segundo um dado modelo, é visto em Colosimo & Giolo (2006), mais exatamente na página 96.

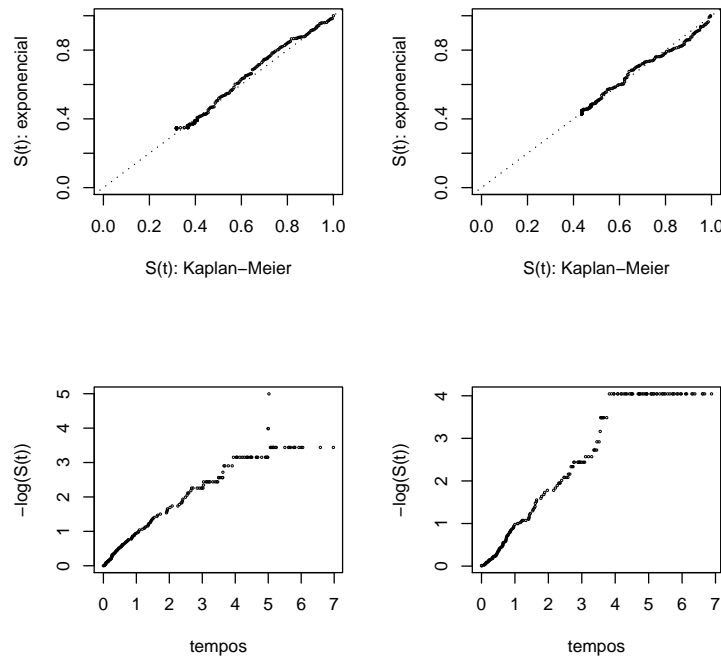


Figura 6.7. Ajuste ao modelo exponencial, sob dois métodos alternativos; na primeira linha, método 1: plot de Kaplan-Meier versus  $S(t)$ : exponencial; na segunda linha, método 2: plot de tempos versus  $-\log\left(\frac{ekm-p_0}{1-p_0}\right)$ , em que  $ekm$  é a estimativa de Kaplan-Meier. Na primeira coluna, ajustes para o grupo observação; na segunda, ajustes para o grupo interferon. O ajuste ao modelo exponencial é feito sob os emv's de  $\lambda$  e  $p_0$ , com o procedimento **Ina**.

Na Figura 6.7 temos dois outros métodos gráficos que confirmam o ajuste do modelo exponencial para os dois grupos. Estes métodos gráficos são vistos em Colosimo & Giolo (2006), página 96, mas para o modelo de sobrevivência

Tabela 6.1. Comprimento  $I_\lambda$  e  $I_{p_0}$ , dos intervalos de  $\lambda$  e de  $p_0$ , respectivamente, e estimativas  $\tilde{\lambda}$  de  $\lambda$ , e  $\tilde{p}_0$  de  $p_0$ . Na penúltima coluna, o procedimento **Iibe** é implementado com o uso da emv de  $\eta$  a partir dos dados E1690. Na última coluna, o procedimento **Iibe\*** é implementado com o uso da emv de  $\eta$  a partir dos dados E1684.

Grupo Observação

Parâmetros	NInao	NInae	Ina	NIab	Iibe	Iibe*
$I_\lambda$	0.3681	0.3684	0.3683	0.3625	0.2585	0.3208
$I_{p_0}$	0.1425	0.1562	0.1562	0.1457	0.1405	0.1359
$\tilde{\lambda}$	0.8455	0.8455	0.8455	0.8397	0.8455	1.1178
$\tilde{p}_0$	0.3392	0.3392	0.3392	0.3378	0.3393	0.3547

Grupo Interferon

$I_\lambda$	0.3839	0.3926	0.3926	0.3930	0.2524	0.2991
$I_{p_0}$	0.1449	0.1599	0.1599	0.1433	0.1451	0.1354
$\tilde{\lambda}$	0.8351	0.8351	0.8349	0.8274	0.8353	1.0640
$\tilde{p}_0$	0.4258	0.4258	0.4257	0.4246	0.4245	0.4393

usual, e neste trabalho estendemos para modelos de longa duração. Sob o modelo exponencial são feitas as inferências sob os procedimentos apresentados na Seção 6.4. Os resultados são apresentados na Tabela 6.1.

Analisamos então os resultados apresentados na Tabela 6.1. Nos procedimentos **NInae** e **Iibe**, o parâmetro  $\eta$  está presente na matriz de informação de Fisher e distribuição *a priori* de Jeffreys, respectivamente. É substituído pela sua estimativa de máxima verossimilhança (emv), a partir do mesmo conjunto de dados E1690. Sob o procedimento **NInae**, nas expressões (4.13)-(4.15), se atribuirmos valores para  $\eta$  diferentes desta emv, o comprimento do intervalo para ambos os parâmetros de interesse fica alterado, se comparado com os mesmos sob os procedimentos **NInao**, **Ina** e **NIab**. Sob o procedimento **Iibe**, valores de  $\eta$  diferentes da sua emv altera não só o comprimento dos intervalos de credibilidade

dos parâmetros, mas também a estimativa pontual dos mesmos. É o que se verifica no mesmo procedimento (última Coluna), que denotamos por **Iibe\***, em que o valor de  $\eta$  dado é obtido sob estimativa de máxima verossimilhança a partir de um outro conjunto de dados, o E1684. É intuitivo pensar que  $\eta$  diferente da sua emv compromete também a probabilidade de cobertura dos intervalos, sob os procedimentos **NInae** e **Iibe**.

O resultado do procedimento **NInae** nessa tabela requer um comentário. No período de implementação dos algoritmos usados no estudo de simulação não nos ocorreu estudar o procedimento **NInae**, para o qual o valor de  $\eta$ , nas expressões (4.13)-(4.15), pode ser obtido por meio dos valores de entrada. Mas estes resultados a partir de dados reais ensejam a necessidade de um estudo de simulação com este procedimento, e uma hipótese que interessa verificar é a de que, substituindo  $\eta$  pelo seu valor verdadeiro nas expressões (4.13)-(4.15), isso equivale a usar a censura informativa.

Voltando à análise desta tabela, nota-se, para os dois parâmetros, que os procedimentos **NInao**, **NInae**, **Ina** e **NIab** são equivalentes. Isto é previsto pelos resultados da simulação quando a fração de cura está na região à direita do seu espaço paramétrico, pois, sob censura informativa, para o grupo observação, com 40.09% de dados censurados, a fração de cura estimada é de 33.92%, e o grupo interferon, com 46.98% de dados censurados, apresenta uma fração de cura estimada em 42.57%. Mas o detalhe que deve ser observado é que os procedimentos **NIab** e **Iibe** não foram simulados com amostras de tamanho  $n = 400$ ,  $n$  grande, enquanto que, como vimos no parágrafo anterior, o **NInae** ficou fora dos estudos de simulação. O procedimento **NIab** apresenta pequenas diferenças, na ordem da terceira casa decimal, tanto no comprimento dos intervalos quanto nas estimativas pontuais dos parâmetros.

Quanto à abordagem bayesiana com censura informativa, na estimativa pontual de  $p_0$ , o procedimento **Iibe** é equivalente aos **NInao**, **NInae** e **Ina**. Com relação ao comprimento do intervalo de credibilidade, este mesmo procedimento **Iibe** apresenta ligeira diferença, se comparado aos procedimentos **NInae** e **Ina**.

Na inferência sobre  $\lambda$ , o procedimento **Iibe** resulta num comprimento de intervalo significativamente menor, se comparado aos procedimentos **NIInao**, **NIInae** e **Ina**. Isso é previsto nos estudos de simulação apresentados na seção anterior, pois o valor de  $\eta$  dado é uma informação com influência significativa na inferência de  $\lambda$ , e esta influência é explicada pelo fato de que  $\eta$  relaciona o parâmetro  $\lambda$ , do tempo de falha, com  $\lambda_c$ , este último, parâmetro da variável de censura. Ou seja,  $\eta = \frac{\lambda_c}{\lambda}$ . No entanto, nesta Tabela 6.1, o procedimento **Iibe** requer o uso preliminar da amostra para obter a emv de  $\eta$ , e isso implica que a amostra é usada duas vezes.

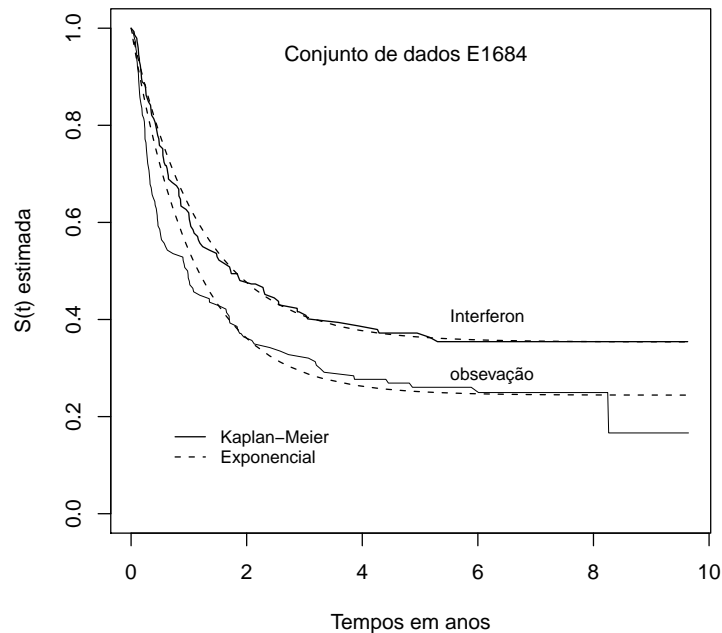


Figura 6.8. Sob o conjunto de dados E1684, ajuste das estimativas de Kaplan-Meier ao modelo de mistura padrão exponencial (3.4), modelo de Barkson-Gage, sob os emv's de  $\lambda$  e  $p_0$  com procedimento **Ina**.

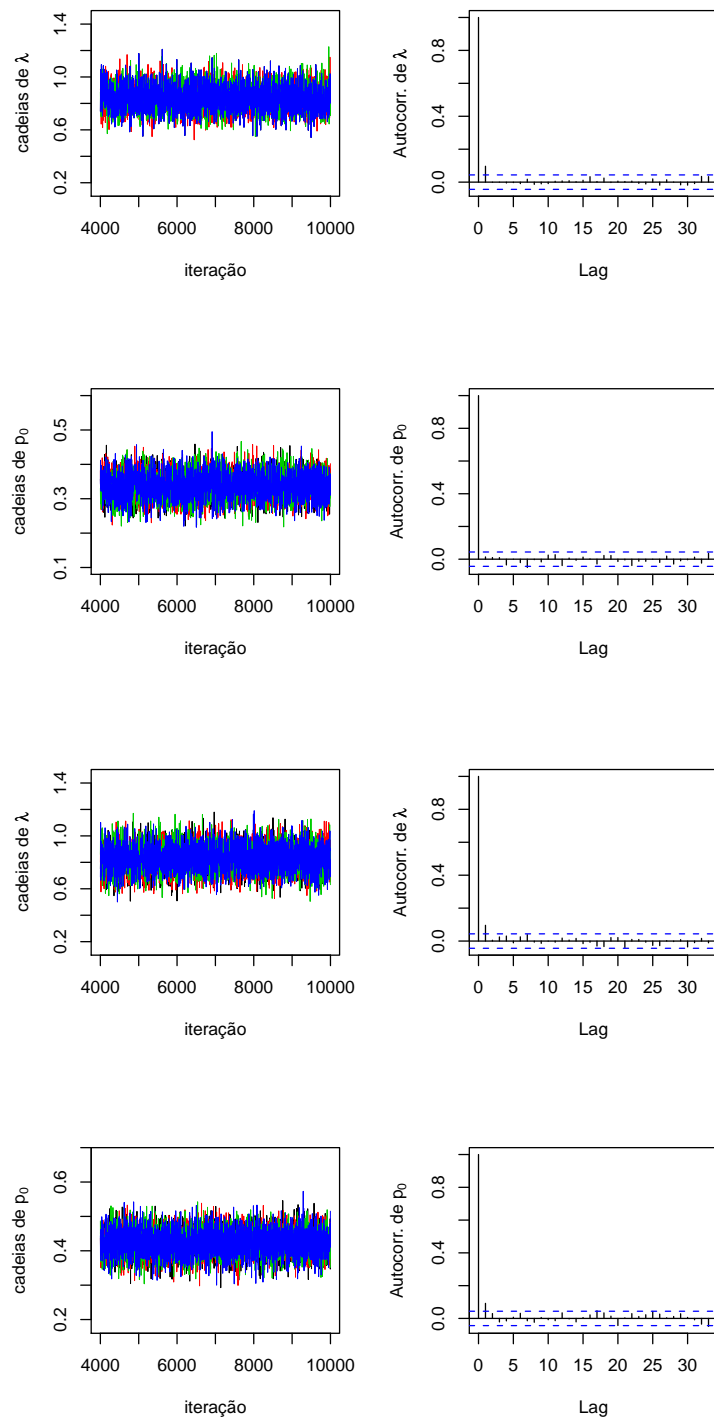


Figura 6.9. MCMC da distribuição conjunta a posteriori (5.11): Na primeira e segunda linhas, grupo observação; na terceira e quarta, grupo interferon. Para cada gráfico, na primeira coluna, históricos das quatro cadeias, na segunda, autocorrelação das cadeias.

# Capítulo 7

## Conclusões e Propostas de Trabalhos Futuros

A análise de cobertura realizada no Capítulo 6, com dados simulados, pode é feita de forma comparativa, sob duas perspectivas. Primeiramente, consideramos o estudo da cobertura dos intervalos sob as abordagens clássica e bayesiana, sob a expectativa de que a abordagem bayesiana iria produzir coberturas mais estáveis, em torno de 95%. Com o uso de conceitos frequentistas foi estimada a probabilidade de cobertura dos intervalos no espaço paramétrico da fração de cura, sob dois procedimentos clássicos e dois bayesianos. Estas estimativas, associadas à análise do comprimento dos intervalos sugere que a abordagem bayesiana confirma essa expectativa, no caso de censura informativa, desde que o valor verdadeiro de  $\eta$  seja conhecido.

Com dados reais, uma opção imediata é trabalhar com a função de verossimilhança sob suposição de censura informativa, com expressão em (4.16), e respectiva matriz de informação de Fisher em (4.19), que é o procedimento **Ina**. Uma segunda opção é o procedimento **Iibe**, em que  $\eta$  deve ser dado pela sua emv, esta obtida sob o procedimento **Ina**. O problema é que para obter esta estimativa de  $\eta$  é necessária a implementação do procedimento **Ina**, e desta forma usar a amostra numa primeira vez, antes de fazer o estudo de inferência pelo procedimento **Iibe**,



que requer o uso da amostra pela segunda vez.

Uma solução que propomos é o procedimento **Iibe\***, e estimar  $\eta$  a partir de um outro conjunto de dados pré-existente, com o mesmo perfil do conjunto de dados atual. Isso é feito com o conjunto de dados reais E1690, Seção 6.6, mas os dois grupos (observação e interferon) do conjunto de dados pré-existente E1684, não têm o mesmo perfil dos do E1690. Essa diferença de perfil pode ser verificada empiricamente, comparando os gráficos das estimativas de Kaplan-Meier do conjunto de dados E1684 e E1690, sob os dois grupos, observação e interferon. Veja Figuras 6.5 e 6.8.

Sob o procedimento **Ina**, as emv's de  $\eta$  a partir do conjunto de dados E1684 são  $\eta = 0.1130$  e  $\eta = 0.1426$ , para os grupos Obsevação e Interferon, respectivamente, enquanto que as mesmas emv's a partir do E1690 são, na mesma sequência,  $\eta = 0.2270$  e  $\eta = 0.2317$ . Em Chen *et al.* (2002), na elicitação de distribuições *a priori* dos parâmetros de interesse, também é executado um procedimento em que a informação obtida do conjunto de dados E1684 (através da função de verossimilhança usada) exerce um peso (probabilidade) na distribuição *a priori* dos parâmetros de interesse, e essa probabilidade varia no intervalo  $[0, 1]$ . Naquele texto também pode ser observado que o fato de os dois conjunto de dados, E1684 e E1690, serem de populações diferentes, altera os valores das inferências dos parâmetros de interesse.

Na segunda perspectiva, foram comparadas as suposições de censura não-informativa e de censura informativa, também com a análise das coberturas, e através dos procedimentos **NInao** e **Ina**, em que se verificou que a suposição de censura não-informativa oculta incertezas decorrentes do aspecto dual da censura (veja equações em (3.24)), principalmente da concorrência dessa dualidade na região “do meio” do espaço paramétrico da fração de cura. Nessa região (veja Figuras 6.2 e 6.3), sob amostras pequenas, a probabilidade estimada de cobertura de  $\lambda$  e de  $p_0$  cai significativamente sob a metodologia **Ina**. Esta queda sob o procedimento **Ina** é corrigido sob o procedimento **Iibe**.

No entanto, com relação aos estudos com dados simulados, consideramos par-

cial, não completos, estes resultados sob estas duas perspectivas, dado que o estudo de inferência sob as duas suposições de censura, sob as duas abordagens, clássica e bayesiana, com e sem variável latente, foi apresentado numa forma incompleta. Para ser objetivo, na abordagem clássica, nos estudos com dados simulados, os resultados referentes às funções de verossimilhança com dados ampliados, sob as duas suposições de censura, dadas em (4.34) e (4.45), foram omitidos neste texto, e requerem estudo detalhado. Também devemos considerar a necessidade do mesmo estudo para o procedimento **NInae**.

Sob a abordagem bayesiana, com relação aos estudos com dados simulados, não foram estudadas as distribuições condicionais de  $\lambda$  e  $p_0$  dadas em (5.3) e (5.4), respectivamente, sob a suposição de censura não-informativa. Neste caso devemos considerar  $\eta$  conhecido, e evidenciamos a possibilidade de comparação com os resultados do procedimento **NInae**. Em três dimensões, faltou desenvolver os mesmos estudos para as distribuições condicionais de  $\lambda$ ,  $p_0$  e  $\eta$ , em (5.7), (5.8) e (5.9), respectivamente.

Com relação às abordagens, clássica ou bayesiana objetiva, sob suposição de censura não-informativa e de censura informativa, com dados ampliados e não ampliados, devemos considerar também a opção de estender os estudos com dados simulados ao espaço paramétrico de  $\lambda$  e de  $p_c$ , este último, um parâmetro perturbador. Com  $p_c > 50\%$ , o valor de  $\eta$  é maior que 1 para valores de  $p_0$  próximo de 0, e o interesse então é estudar o vício e o comportamento do seu estimador de máxima verossimilhança nessa proximidade. Essa discussão nos motiva a dar continuidade aos estudos com dados simulados, como uma proposta de trabalhos futuros.

## 7.1 Sobre censura Informativa

Não podemos deixar de evidenciar dois resultados importantes deste trabalho. Obtemos a matriz de informação de Fisher, sob as duas suposições de censura, Teoremas 4.2 e 4.3. Em Maller & Zhou (1996) é sugerido o uso da matriz

observada relativa à função (4.1), sugestão que usamos no procedimento **NI**nao. O uso da matriz de informação de Fisher, sob censura não-informativa, com elementos dados em (4.13)-(4.15), e da mesma matriz sob suposição de censura informativa, dada em (4.19) é uma contribuição deste trabalho.

Um segundo resultado importante é a função de verossimilhança em (2.54), Teorema 2.10. Note-se que este teorema pressupõe a independência entre a variável de censura e a variável de interesse, mas pode ser estendido para o caso em que a suposição de independência seja violada. Se retirarmos a suposição de independência, a função (2.54) pode ser vista como uma extensão, para o modelo unificado, da função de verossimilhança proposta em Lagakos (1979), texto este do qual seguimos as mesmas notações. Devemos então estabelecer as diferenças entre o Teorema 2.10 e os resultados apresentados em Lagakos (1979).

Em Lagakos (1979) é considerada a função de sobrevivência usual, sem fração de cura, que vale para a função  $S_p^*(y)$  em (2.11) e (2.12). Aquele texto considera que a censura é informativa quando, de (2.51) e (2.52),  $a(t_i)$  e  $B(t_i)$  são variáveis aleatórias, cuja soma,  $a(t_i) + B(t_i)$ , é não degenerada e tem esperança em  $S_p^*(\cdot)$  dada por

$$\int_0^\infty [a(t_i) + B(t_i)] dF_p^*(t_i) = 1. \quad (7.1)$$

Segundo aqueles resultados, sob (7.1), não é suficiente ocorrer relação de dependência entre a variável de falha  $Y$  e a de censura  $X$  para que a censura seja informativa. Em Williams & Lagakos (1977) é apresentado um contra exemplo que consolida essa insuficiência. Em Williams & Lagakos (1977), Lagakos & Williams (1978) e Lagakos (1979) temos uma sequência de três textos que abordam a censura informativa, seguindo esta mesma restrição em (7.1).

Uma diferença entre a nossa abordagem e a vista em Lagakos (1979) é que a nossa abordagem considera que a função de verossimilhança em (2.54) deve ser considerada, a menos que  $\eta$  seja conhecido nas expressões (4.13)-(4.15). Uma interpretação empírica da abordagem em Lagakos (1979), é que nesta a censura é informativa se a influência da variável de censura afeta significativamente as estimativas pontuais dos parâmetros de interesse, enquanto que a nossa abordagem

se justifica na presença do parâmetro perturbador  $\eta$  na matriz de informação de Fisher. Nos estudos com dados simulados apresentados neste trabalho, sob os procedimentos clássicos **NI**nao e **I**na, a diferença entre as médias das estimativas pontuais só é significativa para  $\lambda$ , e somente na região do espaço paramétrico de  $p_0$  adjacente à  $p_0 = 0$ . A Figura 6.2 sugere que essa região tende a diminuir com o aumento do tamanho da amostra.

Observando as Figuras 6.2 e 6.3, tendo como referência o espaço paramétrico de  $p_0$  no eixo das abscissas, a instabilidade das inferências na região adjacente à  $p_0 = 0$  e a competição entre os dois mecanismos de censura na região do “meio”, já comentada, parecem sugerir uma tendência natural de  $p_0$  estar na região contrária, adjacente à  $p_0 = p_c$ . Esta conjectura, tão subjetiva a ponto de ser uma elucubração (ou divagação), mas também empírica, motivada pela observação intensa nos estudos com dados simulados, pode ter um terceiro respaldo na dificuldade de se obter conjunto de dados com  $p_c$  alto, maior que 50%, e  $p_0$  próximo de 0.

Sobre censura informativa, existe também uma diferença entre a nossa abordagem e a de Lawless (1982), que considera que o vetor de parâmetros da variável de censura deve depender de parâmetros da variável de interesse. Consideração semelhante se verifica em Huang & Wolfe (2002). Neste texto evidenciamos a necessidade de considerar  $\eta$  um terceiro parâmetro a ser estimado. O estimador de máxima verossimilhança de  $\eta$ , dado pela expressão em (4.26), é obtido maximizando a função de verossimilhança sob suposição de censura informativa, que é a função 4.16.

Considerando a sobrevivência usual, sem fração de cura, suponha-se  $Y \sim \exp(\lambda)$  e  $X \sim \exp(\lambda\eta)$ , em que  $\eta = \frac{p_c}{1-p_c}$ . Os estimadores de máxima verossimilhança, em  $v$ 's de  $\lambda$ ,  $\hat{\lambda}_{ni}$  sob suposição de censura não-informativa, e  $\hat{\lambda}_{inf}$  sob suposição de censura informativa, são dados respectivamente por

$$\hat{\lambda}_{ni} = \frac{\sum \delta}{\sum t} \quad \text{e} \quad \hat{\lambda}_{inf} = \frac{n}{(1 + \eta) \sum t}, \quad (7.2)$$

com informações de Fisher para  $\lambda$  dadas por

$$I_{ni} = \frac{n}{(1 + \eta)\lambda^2} \quad \text{e} \quad \widehat{I}_{inf} = \frac{n}{\lambda^2}, \quad (7.3)$$

respectivamente. Observe-se que, sob a suposição de censura não-informativa, a informação de  $\lambda$ , dada por  $I_{ni} = \frac{n}{(1+\eta)\lambda^2}$ , contém o parâmetro perturbador  $\eta$ , enquanto que, sob a suposição de censura informativa, a mesma informação, dada por  $\widehat{I}_{inf} = \frac{n}{\lambda^2}$ , não contém o parâmetro perturbador.

## 7.2 Outros Resultados

Os Teoremas 2.22 e 2.23 definem as funções de risco da população e dos não-curados, respectivamente, como a soma esperada dos fatores de risco no momento  $Y = y$ . É possível que exista classes de estruturas de dependência entre os fatores de risco, que potencializem os riscos, da população e dos indivíduos não-curados, ou que neutralizem parte do risco esperado.

No Capítulo 3, temos teoremas que estabelecem as relações paramétricas entre os parâmetros dos modelos (3.4) e (3.14). Por último, o algoritmo de maximização iterativa que apresentamos na última seção do Capítulo 4.

## 7.3 Propostas de Trabalhos Futuros

Uma proposta que consideramos importante é, sob suposição de censura informativa, relacionar a fração de cura com covariáveis, como em Chen *et al.* (2002), com realização de estudos de simulação, e uso do mesmo conjunto de dados E1690. Neste caso obter a matriz de informação, tratável computacionalmente, requer maior investimento.

As covariadas a inserir na fração de cura é da forma

$$p_{0i} = e^{-exp(x'_i\beta)}. \quad (7.4)$$

O conjunto de dados será então dado por

$$\mathcal{D} = [\boldsymbol{\delta}, \mathbf{t}, \mathbf{x}],$$

em que

$$\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top, \quad \mathbf{t} = (t_1, \dots, t_n),$$

e  $\mathbf{x}$  é uma matriz  $n \times (p + 1)$ , dada por

$$\mathbf{x} = [1, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p], \quad \mathbf{1} = (1, \dots, 1)^\top, \quad \mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})^\top, \quad j = 1, \dots, p.$$

Importante nesta proposta é estudar a sensibilidade das covariadas na probabilidade de cobertura dos intervalos de confiança e de credibilidade da fração de cura e taxa de falha. Também é importante considerar as duas suposições de censura, e sob estas, comparar abordagens clássica e bayesiana objetiva nas inferências para o vetor  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ , para  $\lambda$  e para  $p_0$ .

# Apêndice A

## Demonstração do Teorema 3.4

Sob suposição de censura não-informativa, os elementos da matriz de informação de Fisher em (4.5)-(4.7), são apresentados em Maller & Zhou (1996), baseados no teorema que apresentamos a seguir. Este teorema é formulado a partir do Lema 2 de Ghitany *et al.* (1994) e Lema 2.1 de Zhou & Maller (1995), e é usado neste apêndice como base para demonstração do Teorema 3.4.

**Teorema A.1** (Maller & Zhou, 1996). *Suponha-se que  $t_i^*$  são tempos de sobrevivência independentes com distribuição acumulada  $(1 - p_i) F_i(t_i)$ ,  $p_i \in [0, 1]$  e  $u_i$  variáveis aleatórias independentes, não negativas, com distribuição  $G_i(t)$ ,  $t_i = \min(t_i^*, u_i)$ ,  $t_i^*$  e  $u_i$  independentes, para  $i = 1, \dots, n$ . Então, para qualquer função mensurável positiva  $Q : \mathbb{R} \rightarrow \mathbb{R}_+$ ,*

$$E[\delta_i Q(t_i)] = (1 - p_i) E \left\{ \int_0^{u_i} Q(t) dF_i(t) \right\} \quad (\text{A.1})$$

e

$$E[(1 - \delta_i) Q(t_i)] = E[\{1 - (1 - p_i) F_i(u_i)\} Q(u_i)]. \quad (\text{A.2})$$

## A.1 Demonstração de Teorema 3.4

A função log-verossimilhança associada à função de verossimilhança (4.16) é dada por

$$l(p_0, \lambda, \eta) = \sum_{i=1}^n [\log(1 - p_0) \delta_i + \log(\lambda) - \lambda \delta_i t_i + \log(\eta) (1 - \delta_i) - \lambda \eta t_i] \\ + \sum_{i=1}^n (1 - \delta_i) \log [p_0 + (1 - p_0) e^{-\lambda t_i}]. \quad (\text{A.3})$$

As derivadas segundas de (A.3) são dadas por

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \lambda^2} = \sum_{i=1}^n \left[ \frac{1}{\lambda^2} + (1 - \delta_i) \frac{p_0 (1 - p_0) t_i^2 e^{-\lambda t_i}}{[p_0 + (1 - p_0) e^{-\lambda t_i}]^2} \right], \quad (\text{A.4})$$

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial p_0^2} = \sum_{i=1}^n \left[ \frac{\delta_i}{(1 - p_0)^2} + (1 - \delta_i) \frac{1 - e^{-\lambda t_i}}{[p_0 + (1 - p_0) e^{-\lambda t_i}]^2} \right], \quad (\text{A.5})$$

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \eta^2} = \sum_{i=1}^n \frac{1 - \delta_i}{\eta^2}, \quad (\text{A.6})$$

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \lambda \partial p_0} = -\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial p_0 \partial \lambda} = \sum_{i=1}^n \left[ (1 - \delta_i) \frac{t_i e^{-\lambda t_i}}{[p_0 + (1 - p_0) e^{-\lambda t_i}]^2} \right], \quad (\text{A.7})$$

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \lambda \partial \eta} = -\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \eta \partial \lambda} = \sum_{i=1}^n t_i, \quad (\text{A.8})$$

$$-\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial p_0 \partial \eta} = -\frac{\partial^2 l(p_0, \lambda, \eta)}{\partial \eta \partial p_0} = 0. \quad (\text{A.9})$$

Os elementos da matriz de informação (4.19) são obtidos com a aplicação do Teorema A.1 nas derivadas segundas (A.4)-(A.9), e em seguida, com a aplicação da transformação  $\varphi$ -paramétrica (4.8), concluindo assim a prova do teorema. ■

Podemos obter  $I_{\eta\eta}^{**}$  e  $I_{\lambda\eta}^{**}$  em dois procedimentos alternativos. Em (A.6),  $E[\sum_{i=1}^n (1 - \delta_i)] = n p_c$ , e das relações (3.16) e (3.22) calculamos  $I_{\eta\eta}^{**}$ . Em (A.8), a partir da distribuição de  $T$  em (3.31), obtemos  $I_{\lambda\eta}^{**}$ .



# Apêndice B

## Funções Geradoras

Examinamos aqui as relações diferenciais de  $P(s)$  com a média da variável  $M$ . Temos inicialmente que

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}, \quad (\text{B.1})$$

que converge no intervalo  $-1 < s < 1$ . Para  $s = 1$  temos

$$E(M) = \sum_{k=1}^{\infty} k p_k.$$

Se  $E(M) < \infty$ , então  $P'(s)$  é contínua no intervalo  $-1 \leq s \leq 1$ . Se  $\sum_{k=1}^{\infty} k p_k$  diverge então  $P'(s) \rightarrow \infty$  quando  $s \rightarrow 1$ , e neste caso escrevemos que  $P'(s) = E(M) = \infty$ . Existem também relações diferenciais entre  $P(s)$  e a variância da variável  $M$ , que omitimos neste trabalho. Para mais detalhes ver Feller (1967).

Da Definição 2.1 temos que

$$A(s) = a_0 + a_1 s + a_2 s^2 + \dots = E[s^M] = E[e^{\log(s)M}],$$

que é a função geradora de momentos da v.a.  $M$  no ponto  $\log(s)$ .

# Apêndice C

## Programas em R

Apresentamos neste apêndice dois dos programas em R no trabalho. Inicialmente, Seção **C.1**, apresentamos o programa gerador de dados simulados. Na Seção **C.2**, sob a função de verossimilhança (4.16), Procedimento **Ina**, apresentamos o programa para o cálculo das médias frequentistas das coberturas dos intervalos assintóticos, dos seus comprimentos, e das estimativas pontuais. Na Seção **C.3**, sob a distribuição *a posteriori* (5.28), apresentamos o programa para o cálculo das médias frequentistas das coberturas dos intervalos de credibilidade, dos seus comprimentos, e das estimativas pontuais.

### C.1 Geração de dados

Sob os modelos (3.4) e (3.14), e suposição de independência entre  $Y$  e  $X$ , este programa gera 10000 conjunto de dados de tamanho 40, com 50% de censura, fração de cura de 30% e taxa de falha  $\lambda = 0.25$ .

```
> rm(list=ls(all=TRUE))
> D<-10000          # número de réplicas (amostras)
> n<-40            # tamanho de cada réplica
> Pc<-numeric()   # porcentagem de censura estimada
> for (k in 1:D){  # abre chave do D
+   p0<-0.3        # fração de cura a ser estimada
```

```

+   Pcens<-0.50      # porcentagem de censura a ser estimada
+   scale1y<-1      # escala a ser estimada
+   y0<-rep(Inf,n)  # representa o infinito repetido n vezes
+   y1<-rexp(n,scale1y)
+   y<-numeric()
+   t<-numeric()
+ N<-rbinom(n,1,1-p0)
+ delta <- rep(0,n)
+ for (i in 1:n) {      # abre chave do n
+   if (N[i]==0) y[i]<-y0[i]
+   else y[i]<-y1[i]
+   t[i]=min(y[i],c[i])
+   if (y[i]<=c[i]) delta[i] <- 1
+ }                      # fecha chave do n
+ Pc[k]=1-mean(delta)  #Porcentagem de censura amostral calculada para a réplica k
+ }                      # fecha chave do D
> mean(Pc) # média dos Pc's
> ## END

```

## C.2 Inferência sob Censura Informativa, Função de Verossimilhança (4.16)

Para cada valor de  $p_0$ , de 1% a 49%, à razão de 1%, são gerados 10000 conjuntos de dados de tamanho 40, com 50% de censura e taxa de falha  $\lambda = 0.25$ . É feito o estudo frequentista das coberturas, dos comprimentos de intervalos assintóticos e das estimativas de  $p_0 \in \{0.01, 0.02, \dots, 0.48, 0.49\}$  e de  $\lambda = 0.25$ .

```

> rm(list=ls(all=TRUE))
> ptm <- proc.time()
> D<-300
> n<-40
> ka=numeric()kav=numeric();etav=numeric()
> Pc<-numeric()
> theta<-numeric();alpha<-numeric();eta<-numeric()
> alphafinal<-numeric();thetafinal<-numeric();etafinal<-numeric()
> ICalpha1<-numeric();ICalpha2<-numeric();ICtheta1<-numeric();ICtheta2<-numeric()
> ICetaa1<-numeric();ICetaa2<-numeric()
> ICalpha<-numeric();ICtheta<-numeric();ICetaa<-numeric()
> iteracao<-numeric()
> I11<-numeric();I111<-numeric();I12<-numeric();I1212<-numeric();I22<-numeric()
> I222<-numeric();d11<-numeric();d12<-numeric();d22<-numeric()
> dc12<-numeric();dc32<-numeric();dc33<-numeric()

```

```

> In11<-numeric();In22<-numeric();In33<-numeric()
> ncalphae=numeric();ncalphad=numeric();nctetae=numeric();nctetad=numeric()
> ncetae=numeric();ncetad=numeric()
> calfa=numeric();cteta=numeric();ceta=numeric()
> liminfalpha=numeric();limsupalpha=numeric()
> liminftheta=numeric();limsuptheta=numeric()
> liminfeta=numeric();limsupeta=numeric()
> mediaalpha=numeric();mediatheta=numeric();mediaeta=numeric();meanPc=numeric()
> Ialpha=numeric();Itheta=numeric();Ieta=numeric()
> p0<-seq(0.01,0.49,0.01)      # fração de cura a ser estimada
> Pcens<-0.50
> scale1y<-0.25                # taxa de falha a ser estimada
> for (i0 in 1:length(p0)){     # abre chave do p0
+ ka[i0]=(Pcens-p0[i0])/(1-p0[i0])
+ kav[i0]<-ka[i0]              # valor verdadeiro de k
+ scale2c[i0]=scale1y*kav[i0]/(1-kav[i0]) # lambda c: regula a porcentagem de censura
+ etav[i0]<-kav[i0]/(1-kav[i0])
+ for (k in 1:D){             # abre chave do D
+   y0<-rep(Inf,n)
+   y1<-rexp(n,scale1y)
+   y<-numeric()
+   c<-rexp(n,scale2c[i0])
+   t<-numeric()
+   N<-rbinom(n,1,1-p0[i0])
+   delta <- rep(0,n)
+   #####
+   # Algoritmo para simulação de cada conjunto de dados
+   #####
+   for (i in 1:n) {          # abre chave do n
+     if (N[i]==0) y[i]<-y0[i]
+     else y[i]<-y1[i]
+     t[i]=min(y[i],c[i])
+     if (y[i]<=c[i]) delta[i] <- 1
+   }
+   # fecha chave do n
+
+   Pc[k]<-1-mean(delta)      # Porcentagem de censura amostral
+   #####
+   ## ---- Inferência -----
+   #####
+   n=length(t)
+   epsilon=0.00001
+   alpha=0.1 # valor inicial de theta
+   theta=0.1 # valor inicial de alpha
+   eta=0.5 # valor inicial de alpha
+   M=2000
+   alph=c()
+   thet=c()
+   ita=c()
+   for (j in 2:2) {

```

```

+ u1=function(alpha1)
+ n*log(alpha1)-alpha1*sum(t*delta)-alpha1*eta[j-1]*sum(t)+sum((1-delta)*log(theta[j-1]+
+   (1-theta[j-1])*exp(-alpha1*t))) # continuação da linha anterior
+ alph = optimize(u1,c(0,10000), maximum = T)$maximum
+ alpha=c(alpha, alph)
+ u2 = function(theta1)
+ sum(delta)*log(1-theta1)-alpha[j]*sum(t*delta)+
+   sum((1-delta)*log(theta1+(1-theta1)*exp(-alpha[j]*t))) # continuação da linha anterior
+ thet = optimize(u2,c(0,1), maximum = T)$maximum
+ theta=c(theta, thet)
+ u3 = function(eta1)
+ log(eta1)*sum(1-delta)-alpha[j]*eta1*sum(t)
+ ita = optimize(u3,c(0,10000), maximum = T)$maximum
+ eta=c(eta,ita)
+ }
+
+ for (j in 3:M) {
+ if ((abs(alpha[j-1]-alpha[j-2])>epsilon)&(abs(theta[j-1]-theta[j-2])>epsilon)&
+   (abs(eta[j-1]-eta[j-2])>epsilon)) { # continuação da linha anterior
+ u1=function(alpha1)
+ n*log(alpha1)-alpha1*sum(t*delta)-alpha1*eta[j-1]*sum(t)+
+   sum((1-delta)*log(theta[j-1]+(1-theta[j-1])*exp(-alpha1*t))) # continuação da
+   # linha anterior
+ alph = optimize(u1,c(0,10000), maximum = T)$maximum
+ alpha=c(alpha, alph)
+ u2 = function(theta1)
+ sum(delta)*log(1-theta1)-alpha[j]*sum(t*delta)+
+   sum((1-delta)*log(theta1+(1-theta1)*exp(-alpha[j]*t))) # continuação da
+   # linha anterior
+ thet = optimize(u2,c(0,1), maximum = T)$maximum
+ theta=c(theta, thet)
+ u3 = function(eta1)
+ log(eta1)*sum(1-delta)-alpha[j]*eta1*sum(t)
+ ita = optimize(u3,c(0,10000), maximum = T)$maximum
+ eta=c(eta,ita)
+ }
+ else break
+ cat(j, "nn") }
+ m=length(theta)
+ alphafinal[k]=alpha[m];thetafinal[k]=theta[m];etafinal[k]=eta[m];iteracao[k]=m
+
+ ## ---- 1.1 - calculando os elementos da matriz de informação ----
+ rmc1<-function(u){
+ (u^2)*(exp(-u*(etafinal[k]+1)))/(1+((1-thetafinal[k])/thetafinal[k])*exp(-u))
+ }
+ I11<-integrate(rmc1, lower = 0, upper = Inf)
+ I11[k]=as.numeric(I11[1]) +d11[k]=((1-thetafinal[k])/(alphafinal[k]^2))*
+ ((etafinal[k]+1)/((1-thetafinal[k])))*((1/(etafinal[k]+1))- # continuação da
+ (etafinal[k]/(etafinal[k]+1))*(1-thetafinal[k])*I11[k]) # linha anterior

```

```

+
+ rmc2<-function(u){
+ u*(exp(-u*(etafinal[k]+1)))/(1+((1-thetafinal[k])/thetafinal[k])*exp(-u))
+ }
+ I12<-integrate(rmc2, lower = 0, upper = Inf)
+ I1212=as.numeric(I12[1])
+ d12[k]=etafinal[k]*(1/alphafinal[k])*(1/(thetafinal[k]))*I1212
+ rmc3<-function(u){
+ ((exp(-u*etafinal[k])-(exp(-u*(etafinal[k]+1)))))/(1+
+ ((1-thetafinal[k])/thetafinal[k])*exp(-u)) # continuação da linha anterior
+ }
+ I22<-integrate(rmc3, lower = 0, upper = Inf)
+ I222=as.numeric(I22[1])
+ d22[k]=etafinal[k]*(1/(thetafinal[k]*(1-thetafinal[k])))*I222
+ dc12[k]<-((1-thetafinal[k])/alphafinal[k])*(1/(etafinal[k]+1))+
+ thetafinal[k]/(alphafinal[k]*etafinal[k]) # continuação da linha anterior
+ dc32[k]<-0
+ dc33[k]<-((thetafinal[k]+etafinal[k])/(etafinal[k]^2)*(1+etafinal[k]))
+ Iesperada=matrix(c(n*d11[k],n*d12[k],n*dc12[k],n*d12[k],
+ n*d22[k],n*dc32[k],n*dc12[k],n*dc32[k],n*dc33[k]),nrow=3,ncol=3) # continuação da
+ # linha anterior
+ Iinv=solve(Iesperada)
+ In11[k]<-Iinv[1,1]
+ In22[k]<-Iinv[2,2]
+ In33[k]<-Iinv[3,3]
+ ICalpha1[k]=alphafinal[k]-qnorm(0.975)*sqrt(In11[k])
+ ICalpha2[k]=alphafinal[k]+qnorm(0.975)*sqrt(In11[k])
+ ICtheta1[k]=thetafinal[k]-qnorm(0.975)*sqrt(In22[k])
+ ICtheta2[k]=thetafinal[k]+qnorm(0.975)*sqrt(In22[k])
+ ICetaa1[k]=etafinal[k]-qnorm(0.975)*sqrt(In33[k])
+ ICetaa2[k]=etafinal[k]+qnorm(0.975)*sqrt(In33[k])
+ ICalpha[k]=ICalpha2[k]-ICalpha1[k]
+ ICtheta[k]=ICtheta2[k]-ICtheta1[k]
+ ICetaa[k]=ICetaa2[k]-ICetaa1[k]
+ }
+ #####
+ ## ---- Verificação da frequência das coberturas ----
+ #####
+ j0=0;j1=0;k0=0;k1=0;L0=0;L1=0
+ for(i in 1:D){
+ if (ICalpha1[i]>scale1y) j0=j0+1
+ if (ICalpha2[i]<scale1y) j1=j1+1
+ if (ICtheta1[i]>p0[i0]) k0=k0+1
+ if (ICtheta2[i]<p0[i0]) k1=k1+1
+ if (ICetaa1[i]>etav[i0]) L0=L0+1
+ if (ICetaa2[i]<etav[i0]) L1=L1+1
+ }
+ ncalphae[i0]=j0/D;ncalphad[i0]=j1/D;ncthetae[i0]=k0/D;ncthetad[i0]=k1/D
+ ncetae[i0]=L0/D;ncetad[i0]=L1/D;

```

```

+ calfa[i0]<-1-ncalphae[i0]-ncalphad[i0];cteta[i0]<-1-nctetae[i0]-nctetad[i0]
+ ceta[i0]<-1-ncetae[i0]-ncetad[i0]
+ liminfalpha[i0]=mean(ICalpha1); limsupalpha[i0]=mean(ICalpha2)
+ liminftheta[i0]=mean(ICtheta1); limsuptheta[i0]=mean(ICtheta2)
+ liminfeta[i0]=mean(ICetaa1);limsupeta[i0]=mean(ICetaa2)
+ mediaalpha[i0]<-mean(alfafinal);mediatheta[i0]<-mean(thetafinal)
+ mediaeta[i0]<-mean(etafinal);meanPc[i0]=mean(Pc)
+ } # fecha chave do p0
>
> Ialpha=limsupalpha-liminfalpha;Itheta=limsuptheta-liminftheta;Ieta=limsupeta-liminfeta
> cobert=matrix(c(p0,ka,mediaalpha,ncalphae,ncalphad,calfa,Ialpha,mediatheta,nctetae,
+ nctetad,cteta,Itheta,mediaeta,ncetae,ncetad,ceta,Ieta,liminfalpha, # continuação
+ limsupalpha,liminftheta,limsuptheta,liminfeta,limsupeta), # da linha
+ length(p0),23) # anterior
> write.table(cobert, file = "IF com eta alpha025 3Dn20.txt",
+ quote = TRUE, sep = " ", dec = ".",row.names = # continuação da
+ FALSE, col.names = TRUE) # linha anterior
> proc.time() - ptm
> timeh=(proc.time() - ptm)/3600
> #####
> # ---- RESULTADOS ----
> #####
> timeh # tempo de duração da execução do programa
> meanPc
> D
> n
> scalely
> i0
> print(cobert, digits=4)
> # END

```

### C.3 Estudo de Cobertura dos Intervalos de Credibilidade Associados à Distribuição *A Posteriori* conjunta (5.28)

Para cada valor de  $p_0$ , de 1% a 49% à razão de 3%, são gerados 1000 conjunto de dados de tamanho 40, com 50% de censura e taxa de falha  $\lambda = 0.25$ . É feito o estudo frequentista das coberturas e dos comprimentos dos intervalos de credibilidade. Em razão do custo computacional, para o caso Bayesiano geramos apenas 1000 réplicas para cada valor de  $p_0$ .

```

> rm(list=ls(all=TRUE))
> ptm <- proc.time()
> D<-1000;n<-40
> af=numeric();ka=c();kav=numeric()
> Pc<-numeric()
> theta<-numeric();alpha<-numeric();thetafinal<-numeric();alphafinal<-numeric()
> ICalpha1<-numeric();ICalpha2<-numeric();ICtheta1<-numeric();ICtheta2<-numeric()
> ICalpha<-numeric();ICtheta<-numeric();alphaf<-numeric();thetaf<-numeric()
> af=numeric();ka=c();kav=numeric();eta=numeric()
> ncalphae=numeric();ncalphad=numeric();nctetae=numeric();nctetad=numeric()
> ncalfa=numeric();ncteta=numeric();calfa=numeric();cteta=numeric()
> liminfalpha=numeric();limsupalpha=numeric()
> liminftheta=numeric();limsuptheta=numeric()
> mediaalpha=numeric();mediatheta=numeric()
> meanPc=numeric()
> INF11=numeric();INF12=numeric()
> INF22=numeric();V11=numeric()
> V12=numeric();V22=numeric()
> p0<-seq(0.01,0.49,0.03)           # fração de cura a ser estimada
> Pcens<-0.50                       # porcentagem de censura
> scale1y=0.25                      # taxa de falha a ser estimada
> for (i0 in 1:length(p0)) {        # abre chave do p0
+ ka[i0]=(Pcens-p0[i0])/(1-p0[i0])
+ kav[i0]<-ka[i0]                   # verdadeiro valor de k
+ af[i0]=scale1y*kav[i0]/(1-kav[i0])
+ scale2c=af[i0]                   # lambda c regula o valor de k
+ for (k in 1:D){                  # abre chave do D
+   y0<-rep(Inf,n)
+   y1<-rexp(n,scale1y)
+   y<-numeric()
+   c<-rexp(n,scale2c)
+ t<-numeric()
+ N<-rbinom(n,1,1-p0[i0])
+ delta <- rep(0,n)
+ #####
+ # Algoritmo de simulação de cada conjunto de dados
+ #####
+ for (i in 1:n) {                  # abre chave do n
+   if (N[i]==0) y[i]<-y0[i]
+   else y[i]<-y1[i]
+   t[i]=min(y[i],c[i])
+   if (y[i]<=c[i]) delta[i] <- 1
+ }                                  # fecha chave do n
+ Pc<-1-sum(delta)/n                #Porcentagem de censura
+ #####
+ ## ---- 1 - INFERÊNCIA ----
+ #####
+ n=length(t)
+ alpha=0.5 # valor inicial de alpha

```



```

+ theta=0.5 # valor inicial de theta
+ vetoralpha=alpha;vetortheta=theta
+ p=c();w=c()
+ M=5000 #número de iterações
+ matrizp=c();matrizw=c()
+ iteralpha=numeric();itertheta=numeric()
+ theta1=numeric()
+ theta1=0.5;vetortheta1=theta1
+ I111<-numeric();I1111<-numeric();d111<-numeric()
+ I121<-numeric();I12121<-numeric();d121<-numeric()
+ I221<-numeric();I2221<-numeric();d221<-numeric()
+ IF11<-numeric();IF21<-numeric();IF1<-numeric()
+ J1=numeric()
+
+ I112<-numeric();I1112<-numeric();d112<-numeric()
+ I122<-numeric();I12122<-numeric();d122<-numeric()
+ I222<-numeric();I2222<-numeric();d222<-numeric()
+ IF12<-numeric();IF22<-numeric();IF2<-numeric()
+ J2=numeric()
+ r1t=numeric();r2t=numeric();lrt=numeric()
+ Ualfgt=numeric();G=numeric();Ualfat=numeric()
+ Ualfgt=1
+ G=1
+ Ualfat=0.1 # Ualfat inicial deve ser sempre menor que G inicial
+ thet=numeric()
+ thet=1
+ #####
+ # ----- 1.1 - MCMC -----
+ #####
+ for (j in 1:M) {
+ #####
+ for (i in 1:n) {
+ p[i]=theta/(theta+(1-theta)*exp(-alpha*y[i]))
+ w[i]=rbinom(1,1,p[i])
+ }
+ #### Geração de alfa
+ a=n
+ s=sum(t*delta)+(ka[i0]/(1-ka[i0]))*sum(t)+sum((1-w)*(1-delta)*t)
+ vetoralpha=c(vetoralpha, rgamma(1,shape=a, scale=1/s))
+ alpha=vetoralpha[j+1]
+ #####
+ #####
+ #### Geração de teta
+ st1=sum(w*(1-delta))+0.5
+ st2=n-sum(w*(1-delta))+1
+ vetortheta1<-c(vetortheta1,rbeta(1,st1,st2))
+ theta1=vetortheta1[j+1]
+ rmc11<-function(u){
+ (u^2)*(exp(-u/(1-ka[i0])))/(1+((1-theta1)/theta1)*exp(-u))

```

```

+ }
+ I111<-integrate(rmc11, lower = 0, upper = Inf)
+ I1111[j]=as.numeric(I111[1])
+ d111[j]=(1/(1-theta1))-I1111[j]
+ rmc21<-function(u){
+ u*(exp(-u/(1-ka[i0])))/(1+((1-theta1)/theta1)*exp(-u))
+ }
+ I121<-integrate(rmc21, lower = 0, upper = Inf)
+ I12121[j]=as.numeric(I121[1])
+ d121[j]=I12121[j]
+ rmc31<-function(u){
+ ((exp(-ka[i0]*u/(1-ka[i0]))-(exp(-u/(1-ka[i0])))))/
+ (1+((theta1)/theta1)*exp(-u)) # continuação da linha anterior
+ }
+ I221<-integrate(rmc31, lower = 0, upper = Inf)
+ I2221[j]=as.numeric(I221[1])
+ d221[j]=I2221[j]
+ #####
+ IF11[j]<-d111[j]*d221[j]
+ IF21[j]<- (ka[i0]/(1-ka[i0]))*(1/theta1)*(d121[j]^2)
+ IF1[j]<-IF11[j]-IF21[j]
+ J1[j]=sqrt(IF1[j])
+ #####
+ rmc12<-function(u){
+ (u^2)*(exp(-u/(1-ka[i0])))/(1+((1-theta)/theta)*exp(-u))
+ }
+ I112<-integrate(rmc12, lower = 0, upper = Inf)
+ I1112[j]=as.numeric(I112[1])
+ d112[j]=(1/(1-theta))-I1112[j]
+ rmc22<-function(u){
+ u*(exp(-u/(1-ka[i0])))/(1+((1-theta)/theta)*exp(-u))
+ }
+ I122<-integrate(rmc22, lower = 0, upper = Inf)
+ I12122[j]=as.numeric(I122[1])
+ d122[j]=I12122[j]
+ rmc32<-function(u){
+ ((exp(-ka[i0]*u/(1-ka[i0]))-(exp(-u/(1-ka[i0])))))/(1+((theta)
+ /theta)*exp(-u)) # continuação da linha anterior
+ }
+ I222<-integrate(rmc32, lower = 0, upper = Inf)
+ I2222[j]=as.numeric(I222[1])
+ d222[j]=I2222[j]
+ #####
+ IF12[j]<-d112[j]*d222[j]
+ IF22[j]<- (ka[i0]/(1-ka[i0]))*(1/theta)*(d122[j]^2)
+ IF2[j]<-IF12[j]-IF22[j]
+ J2[j]=sqrt(IF2[j])
+ #####
+ r1t[j]=log(J1[j])

```

```

+ r2t[j]=log(J2[j])
+ lrt[j]=r1t[j]-r2t[j]
+ Ualftg=c(Ualftg,runif(1,0,1))
+ G[j]<-log(Ualftg[j])
+ Ualfat[j]<-min(0,lrt[j])
+ if (G[j]<=Ualfat[j]) thet[j]=theta1
+ else thet[j]=theta
+ vetortheta=c(vetortheta,thet[j])
+ theta=vetortheta[j+1]
+ cat(j,"nn")
+ } # chave do M
+ burnin=M-(4*M/5)
+ salto=3
+ indice=seq(burnin+1,M,by=salto)
+ ICalpha1[k]=quantile(vetoralpha[indice],c(0.025))
+ ICalpha2[k]=quantile(vetoralpha[indice],c(0.975))
+ ICtheta1[k]=quantile(vetortheta[indice],c(0.025))
+ ICtheta2[k]=quantile(vetortheta[indice],c(0.975))
+ ICalpha[k]=ICalpha2[k]-ICalpha1[k]
+ ICtheta[k]=ICtheta2[k]-ICtheta1[k]
+ alphaf[k]=mean(vetoralpha[indice])
+ thet[k]=mean(vetortheta[indice])
+ #####
+ ## Fim do MCMC
+ #####
+ } # fecha chave do D
+ #####
+ ### ----- 2 - Verificação da porcentagem de coberturas -----
+ #####
+ j0=0;j1=0;k0=0;k1=0
+ for(i in 1:D){ # abre de novo chave do D
+           if (ICalpha1[i]>scale1y) j0=j0+1
+           if (ICalpha2[i]<scale1y) j1=j1+1
+           if (ICtheta1[i]>p0[i0]) k0=k0+1
+           if (ICtheta2[i]<p0[i0]) k1=k1+1
+ } # fecha chave do D
+ ncalphae[i0]=j0/D;ncalphad[i0]=j1/D;ncthetae[i0]=k0/D;ncthetad[i0]=k1/D
+ calfa[i0]<-1-(ncalphae[i0]+ncalphad[i0])
+ cteta[i0]<-1-(ncthetae[i0]+ncthetad[i0])
+ #####
+ ## ---- 3 - Médias relativas aos intervalos e estimativas e matriz
+ ##           de resultados
+ #####
+ liminfalpha[i0]=mean(ICalpha1); limsupalpha[i0]=mean(ICalpha2)
+ liminftheta[i0]=mean(ICtheta1); limsuptheta[i0]=mean(ICtheta2)
+ mediaalpha[i0]<-mean(alphaf)
+ mediatheta[i0]<-mean(thetaf)
+ meanPc[i0]=mean(Pc)
+ } # fecha chave do p0

```

```

> Ialpha=numeric();Itheta=numeric()
> Ialpha=limsupalpha-liminfalpha # amplitude
> Itheta=limsuptheta-liminftheta # amplitude
> cobert= matrix(c(p0,ka,mediaalpha,ncalphae,ncalphad,calfa,Ialpha,
> mediatheta,nctetae,nctetad,cteta,Itheta,liminfalpha, # continua
> limsupalpha,liminftheta,limsuptheta),length(p0),16) # linha anterior
> write.table(cobert, file = "IF com eta alpha025 n20.txt", # continua
> quote = TRUE, sep = " ", dec = ".",#row.names = FALSE, # linha anterior
> col.names = TRUE) # continua linha anterior
> proc.time() - ptm
> timeh=(proc.time() - ptm)/3600
> #####
> ##### ---- 4 - Resultados ----
> #####
> mediaalpha
> mediatheta
> ncalphae
> ncalphad
> nctetae
> nctetad
> calfa
> cteta
> #####
> # ---- ---- 5 - Gráficos ---- ----
> #####
> print(cobert, digits=4)
> par(mfrow=c(2,4))
> plot(p0, calfa, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(p0, cteta, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(p0, Ialpha, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(p0, Itheta, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(p0, mediaalpha, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(p0, mediatheta, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(calfa, cteta, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> plot(1-calfa,1- cteta, type="l", pch=21, bg=par("bg"),
> col = "blue", cex=1)
> ## End

```

## C.4 Sobre as Inferências em Código R

Sob abordagem clássica, as estimativas de máxima verossimilhança, em  $v$ 's, foram obtidas com a implementação do algoritmo de maximização iterativa, descrito no Capítulo 4, Seção 4.4. Portanto, este algoritmo é usado nos procedimentos **NInao**, **NInae** e **Ina**, dos quais apresentamos apenas a implementação do **Ina**, veja a sua codificação na Seção C.2. Na Seção C.2, o algoritmo de maximização iterativa tem a sua codificação vista nas páginas 99 e 100.

Quanto aos métodos de maximização pelo algoritmo EM (veja Seção 4.3), não apresentamos a codificação em R da sua implementação, dado que os respectivos resultados com dados simulados não foram apresentados nesta tese. Mas devemos relatar que as estimativas pontuais obtidas pelo algoritmo de maximização iterativa e pelo algoritmo EM são equivalentes. São dois algoritmos que maximizam as mesmas funções de verossimilhança: a função 4.1, sob censura não-informativa, e a função 4.16 sob censura informativa. Então, os resultados sob as funções 4.34 e 4.45, e respectivas matrizes de informação, são discutíveis apenas com relação à cobertura dos intervalos. Mas, exatamente por esse motivo decidimos não apresentar os resultados e deixar o seu estudo como proposta de trabalhos futuros.

No entanto, a inferência bayesiana, sob as duas suposições de censura, é implementada com o uso de variável latente, através do conjunto de dados em (4.35). Quanto à implementação dos estudos com dados simulados e com dados reais, com a implementação do MCMC para as distribuições condicionais *a posteriori* em (5.3)-(5.4) e em (5.7)-(5.9), deixamos como proposta de trabalhos futuros, dado os prazos que tínhamos que cumprir.

# Referências

- Abramowitz, M. & Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Washington, D.C.
- Berger, J. (2006). The case for objective bayesian analysis. *Bayesian Analysis*, **1**, 385–402.
- Berger, J., Bernardo, J. M. & Sun, D. (2009). The formal difinition of reference priors. *The Annals of Statistics*, **37**, 905–938.
- Berkson, J. & Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501–515.
- Bernardinelli, L. & Montomoli, C. (2007). Empirical bayes versus fully bayesian analysis of geographical variation in disiaese risk. *Statistics in Medicine*, **11**, 983–1007.
- Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society B*, **1**, 113–147.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*, **11**, 15–53.
- Brent, R. (1973). *Algorithms for Minimization Without Derivatives*. Prentice-Hall, Englewood Cliffs N. J.
- Brown, L. D., Cai, T. T. & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*, **16**, 101–133.
- Chen, M.-H. & Ibrahim, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates. *Biometrics*, **57**, 43–52.
- Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909–919.
- Chen, M.-H., Harrington, D. P. & Ibrahim, J. G. (2002). Bayesian cure rate model for malignant melanoma: a case-study of Eastern Cooperative Oncology Group trial E1690. *Applied Statistics*, **51**, 135–150.

- Colosimo, E. A. & Giolo, S. R. (2006). *Análise de Sobrevida Aplicada*. Edgard Blücher, São Paulo.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, **34**, 187–220.
- de Castro, M., Cancho, V. G. & Rodrigues, J. (2010). A hands-on approach for fitting long-term survival models under the GAMLSS framework. *Computer Methods and Programs in Biomedicine*, **97**, 168–177.
- Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of Royal Statistical Society B*, **39**, 1–38.
- Emoto, S. E. & Matthews, P. C. (1990). A Weibull model for dependent censoring. *The Annals of Statistics*, **18**, 1556–1577.
- Ewell, M. & Ibrahim, J. G. (1997). The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis*, **3**, 5–12.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041–1046.
- Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *The Canadian Journal of Statistics*, **14**, 257–262.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. 1. Wiley, New York, third edition.
- Ghitany, M. E., Maller, R. A. & Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of Multivariate Analysis*, **49**, 218–241.
- Goldman, A. I. (1984). Survivorship analysis when cure is a probability: a Monte Carlo study. *Statistics in Medicine*, **3**, 153–163.
- Gradshteyn, I. S. & Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Edited by A. Jeffrey and D. Zwillinger. Academic Press, New York, 7th edition.
- Gray, R. J. & Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics*, **45**, 899–904.
- Greenhouse, J. B. & Wolfe, R. A. (1984). A competing risks derivation of a mixture model for the analysis of survival. *Communications in Statistics - Theory and Methods*, **13**, 3133–3154.
- Halpern, J. & Brown, B. W. (1987a). Cure rate models: power of the log rank and generalized Wilcoxon tests. *Statistics in Medicine*, **6**, 483–489.
- Halpern, J. & Brown, B. W. (1987b). Designing clinical trials with arbitrary specification of survival functions and for the log rank or generalized Wilcoxon test. *Controlled Clinical Trials*, **8**, 177–189.

- Huang, X. & Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics*, **58**, 510–520.
- Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London, third edition.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, second edition.
- Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.
- Kuk, A. Y. C. & Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531–541.
- Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, **35**, 139–156.
- Lagakos, S. W. & Williams, J. S. (1978). Models for censored survival analysis: a cone class of variable-sum models. *Biometrika*, **65**, 181–189.
- Laska, E. M. & Meisner, M. J. (1992). Nonparametric estimation and testing in a cure rate model. *Biometrics*, **48**, 1223–1234.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- Link, W. A. (1989). A model for informative censoring. *Journal of the American Statistical Association*, **84**, 749–752.
- Maller, R. A. & Zhou, S. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York.
- Mizoi, M. F. (2004). *Influência Local em Modelos de Sobrevida com Fração de Cura*. Tese de doutorado, IME – Universidade de São Paulo, São Paulo, Brasil.
- Paulino, C. D., Turkman, M. A. A. & Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rodrigues, J., Cancho, V. G. & de Castro, M. (2008). *Unified Theory of Survival Analysis*. Brazilian Statistical Association, São Paulo. (in Portuguese).
- Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009). On the unification of the long-term survival models. *Statistics and Probability Letters*, **79**, 753–759.



- Spoto, R., Sather, H. N. & Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients of who are cured. *Biometrics*, **48**, 87–99.
- Stangl, D. K. & Greenhouse, J. B. (1998). Assessing placebo response using bayesian hierarchical survival models. *Lifetime Data Analysis*, **4**, 5–28.
- Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899–907.
- Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**, 1063–1078.
- Williams, J. S. & Lagakos, S. W. (1977). Models for censored survival analysis: constant-sum and variable-sum models. *Biometrika*, **64**, 215–224.
- Yakovlev, A. Y. (1994). Parametric versus nonparametric methods for estimating cure rates based on censored survival-data. *Statistics in Medicine*, **13**, 983–985.
- Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of 'permanent employment' in Japan. *Journal of the American Statistical Association*, **87**, 284–292.
- Yau, K. K. W. & Ng, A. S. K. (2001). Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in Medicine*, **20**, 1591–1607.
- Yin, G. & Ibrahim, J. G. (2005). Cure rate models: a unified approach. *The Canadian Journal of Statistics*, **33**, 559–570.
- Zheng, M. & Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**, 127–138.
- Zhou, S. & Maller, R. A. (1995). The likelihood ratio test for the presence of immunes in a censored sample. *Statistics*, **27**, 181–201.