

# Modelagem de Partição Bayesiana para Dados de Sobrevivência de Longa Duração

JHON FRANKY BERNEDO GONZALES

UFSCar - São Carlos/SP

Dezembro/2009

Universidade Federal de São Carlos  
Centro de Ciências Exatas e de Tecnologia  
Departamento de Estatística

# Modelagem de Partição Bayesiana para Dados de Sobrevivência de Longa Duração

JHON FRANKY BERNEDO GONZALES

ORIENTADORES:

PROF<sup>A</sup>. DR<sup>A</sup>. VERA LUCIA D. TOMAZELLA

PROF. DR. MÁRIO DE CASTRO A. FILHO

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar como parte dos requisitos para obtenção do título de Mestre em Estatística.

UFSCar - São Carlos/SP

Dezembro/2009

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

B525mp

Bernedo Gonzales, Jhon Franky.

Modelagem de partição bayesiana para dados de sobrevivência de longa duração / Jhon Franky Bernedo Gonzales. -- São Carlos : UFSCar, 2009.  
64 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2009.

1. Análise de sobrevivência. 2. Modelos de partição. 3. Modelos estatísticos. 4. Tesselação. I. Título.

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

**Jhon Franky Bernedo Gonzales**

**Modelagem de partição bayesiana para dados de sobrevivência de longa duração**

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

Aprovada em 27 de novembro de 2009.

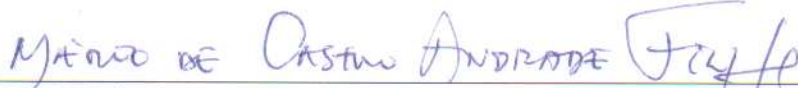
**BANCA EXAMINADORA**

Presidente



Prof. Dra. Vera Lucia Damasceno Tomazella (DEs-UFSCar/ Orientadora)

1º Examinador



Prof. Dr. Mário de Castro Andrade Filho (ICMC-USP/ Co-Orientador)

2º Examinador



Prof. Dra. Roseli Aparecida Leandro (ESALQ-USP)

3º Examinador



Prof. Dra. Teresa Cristina Martins Dias (DEs-UFSCar)

# Agradecimentos

Primeiramente agradeço a Deus, que me dá saúde e força para superar os obstáculos e provas todos os dias.

A meu pai, mãe, pelo constante apoio e ânimo em minha vida, às minhas irmãs Hayme e Gleny pela compreensão e ajuda nos momentos difíceis.

A todos os meus amigos que sempre estiveram carinhosamente presentes, contribuindo com críticas, sugestões e paciente tolerância.

À minha orientadora Vera Lucia D. Tomazella e a meu co-orientador Mário de Castro A. Filho, pela orientação e incentivo na elaboração e condução do trabalho. Foi um prazer trabalhar com eles, que são inspiração em minha vida para continuar estudando.

Aos professores do Departamento de Estatística da Universidade Federal de São Carlos que me abriram as portas e me ofereceram o ambiente acolhedor e sadio para que eu pudesse realizar o meu mestrado.

Aos diversos professores do Departamento de Matemática da Universidad Nacional de San Agustin-Arequipa, em especial os professores Vladimir Rosas, Percy Ticona, Luis Guerra e Daniel Roque, pelo incentivo para estudar estatística.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio concedido para este trabalho.

# Resumo

Neste trabalho apresentamos uma abordagem bayesiana para modelos de sobrevivência com fração de cura na presença de covariáveis. Nesta perspectiva, a modelagem é uma extensão direta do modelo de longa duração (Chen *et al.*, 1999). Este modelo é considerado flexível no sentido de que os efeitos das covariáveis são medidos localmente, utilizando o modelo de partição bayesiana desenvolvido por Holmes *et al.* (1999). O modelo de partição bayesiana é uma abordagem genérica para problemas de classificação e regressão, em que o espaço das covariáveis é dividido em regiões disjuntas definidas por uma estrutura de tesselação. A extensão para modelagem local mantém a estrutura de riscos proporcionais, que é intrínseca ao modelo de longa duração (tempo de promoção) (Rodrigues *et al.*, 2009a). Aplicações desta teoria aparecem em várias áreas, como por exemplo, em Finanças, Biologia, Engenharia, Economia e Medicina. Neste trabalho, apresentamos um estudo de simulação e aplicamos a metodologia a um conjunto de dados na área de estudos clínicos.

**Palavras-chave:** Análise de sobrevivência, Modelos de partição, Modelos de longa duração, Tesselação por hiperplanos.

# Abstract

In this work we present a bayesian approach for the survival model with cure rate in the presence of covariates. In this perspective, the modelling is a direct extension of the long-term model of (Chen *et al.*, 1999). This model is considered flexible in the sense that the effects of the covariates are measured locally using the bayesian partition model developed by Holmes *et al.* (1999). The bayesian partition model is a generic approach to problems of classification and regression where the space of covariates is divided in disjoint regions defined by a structure of tessellation. The extension to modelling local maintains the structure of the proportional hazards model that it is intrinsic of the long-term model(promotion time) (Rodrigues *et al.*, 2009a). Application of this theory appears in several areas, for example in finance, biology, engineering, economics and medicine. We present a simulation study and apply the methodology to a set of data on the clinical studies.

**Keywords:** Survival analysis, Partition model, Long-term model, Tessellation for hyperplanes .

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Conceitos básicos de análise de sobrevivência	3
1.1.1	Funções básicas	4
1.1.2	Método de máxima verossimilhança	5
1.2	Algumas distribuições	6
1.2.1	Distribuição exponencial	6
1.2.2	Distribuição Weibull	7
1.2.3	Distribuição Log-normal	9
1.2.4	Distribuição Log-logística	11
1.3	Objetivos do trabalho	12
<b>2</b>	<b>Modelos de Longa Duração</b>	<b>14</b>
2.1	Modelo de longa duração unificado	15
2.1.1	Função de sobrevivência	15
2.1.2	Função de verossimilhança	19
2.2	Modelo de mistura padrão	20
2.3	Modelo de tempo de promoção	23
2.3.1	Modelo de tempo de promoção com distribuição Weibull	24



---

<b>3</b>	<b>Modelo de Partição Bayesiano</b>	<b>26</b>
3.1	Introdução	26
3.2	Modelo de partição bayesiano	29
3.2.1	Especificação <i>a priori</i>	31
3.2.2	Análise <i>a posteriori</i> e estratégia computacional	33
3.2.3	Predição	34
3.2.4	Tesselação por hiperplanos ortogonais	35
3.3	Extensão local para modelos de longa duração	38
3.3.1	Função de verossimilhança	39
3.3.2	Especificação <i>a priori</i> e verossimilhança marginal	40
3.3.3	Análise <i>a posteriori</i>	42
3.3.4	Predição	44
<b>4</b>	<b>Aplicações</b>	<b>45</b>
4.1	Estudo de simulação	45
4.2	Exemplo de aplicação	50
<b>5</b>	<b>Considerações finais e propostas</b>	<b>59</b>
	<b>Referências</b>	<b>61</b>

# Capítulo 1

## Introdução

As técnicas de análise de sobrevivência consistem, em sua essência, em uma coleção de procedimentos estatísticos utilizados para analisar dados relacionados ao tempo até a ocorrência de um determinado evento de interesse (morte, cura, contração ou recidiva de uma doença) (Lawless, 2002), a partir de um tempo inicial pré-estabelecido. Na área financeira o evento de interesse pode ser, por exemplo, o abandono de um cliente, ou o não pagamento de empréstimos, a ocorrência de um sinistro, etc. Neste contexto, a análise de sobrevivência/confiabilidade, permite determinar quais variáveis afetam a forma da função de risco e também obter estimativas destas funções para cada indivíduo.

A principal característica relacionada a estes dados diz respeito à presença de observações censuradas, que consistem na observação parcial da resposta, o que se dá geralmente pelo abandono do tratamento por alguns pacientes (ou carteira de clientes), morte por outros motivos além do estudado, ou quando, ao término da pesquisa, muitos pacientes ou clientes deixaram de experimentar o evento de interesse.

Em análise de sobrevivência existem vários estudos sobre esses tipos de dados. Contudo, modelos típicos assumem que todos os pacientes eventualmente morrem devido à uma doença, ou que todos os fregueses deixam o banco depois da ocorrência de um particular evento, como por exemplo, o pagamento de um empréstimo. Entretanto, existem situações em que acreditamos que uma

proporção de fregueses não se desligam do banco depois do pagamento de um empréstimo e que existe uma proporção de pacientes curados em relação à doença de interesse. Neste contexto, uma metodologia usada em análise de sobrevivência que considera uma proporção de curados é a de modelos de longa duração, também chamados de modelos com fração de cura.

Os modelos de análise de sobrevivência com longa duração possuem uma vantagem em relação aos modelos de sobrevivência usuais, no sentido de incorporarem a heterogeneidade de duas subpopulações (susceptíveis e imunes); são, por isso conhecidos como modelos de mistura.

A suposição de que alguns pacientes nunca experimentaram o evento de interesse é baseada em considerações científicas ou empíricas, como a presença de um grande número de sobreviventes de longa duração (alta proporção de censura). O estimador Kaplan-Meier (Kaplan & Meir, 1958) é uma boa forma de evidenciar essa presença, uma vez que um grande número de censuras pode ser observado na cauda, ou seja, já que pode ser testada a existência de pacientes “curados”. O gráfico desse estimador deve apresentar uma cauda em um nível aproximadamente constante e estritamente maior que zero, por um período de tempo considerável.

Existe uma extensa literatura sobre os modelos de longa duração. Podemos citar Maller & Zhou (1996), Yakovlev & Tsodikov (1996), Chen *et al.* (1999), Rodrigues *et al.* (2008), entre outros. Dentre esses modelos, o tipo mais utilizado é o modelo de mistura padrão (Boag, 1949; Berkson & Gage, 1952). O modelo de mistura de Berkson & Gage (1952) é muito usado na literatura estatística, no entanto, tem muitas desvantagens que são discutidas em Chen *et al.* (1999). Estes autores, fazem uso de um modelo com algumas vantagens em relação ao modelo de mistura padrão. Este modelo é conhecido na literatura como modelo de tempo de promoção.

Neste trabalho utilizamos o modelo proposto por Hoggart & Griffin (2001), que é uma extensão do modelo de tempo de promoção. Nesta extensão fazemos uso da metodologia de partição bayesiana proposta por Holmes *et al.* (1999). Nesse modelo a fração, de cura depende naturalmente das covariáveis através de um modelo linear generalizado.

O modelo de partição bayesiana (MPB) é uma abordagem genérica para problemas de classificação e regressão, assumindo independência entre as regiões de partição. Regressão e classificação estão entre os problemas mais pesquisados em estatística. Recentemente, têm sido desenvolvidas pesquisas envolvendo extensões bayesianas para modelos clássicos. Exemplos desse tipo, incluem Chipman *et al.* (1998) e Denison *et al.* (1998). Adaptar uma abordagem bayesiana para o problema de classificação e regressão, envolve determinar a distribuição preditiva *a posteriori*. Para desenvolver a modelagem preditiva, devemos requerer uma distribuição *a priori* para o parâmetro de interesse, pois, a função de verossimilhança, é usualmente definida de forma natural a partir do contexto do problema.

O modelo de partição bayesiano é motivado pelo princípio básico de que pontos próximos em um espaço preditor, provêm de uma mesma distribuição local, isto é, em cada região do espaço preditor, é assumida uma distribuição de probabilidade. Amostras da distribuição de probabilidade *a posteriori* do espaço do modelo de partição bayesiano, são gerados usando métodos de Monte Carlo via Cadeia de Markov (MCMC) (Gamerman & Lopes, 2006). Modelar dados usando partições não é uma idéia nova, podemos encontrar exemplo de seu uso em Barry & Hartigan (1993), Green (1995) e Heikkinen (1998).

Na Seção 1.1 apresentamos alguns conceitos básicos de análise de sobrevivência. Na Seção 1.2 apresentamos alguns modelos básicos de análise de sobrevivência e na Seção 1.3 apresentamos os objetivos da tese.

## 1.1 Conceitos básicos de análise de sobrevivência

A análise de sobrevivência avançou bastante nestas últimas décadas, devido ao desenvolvimento e aprimoramento de técnicas estatísticas combinadas com o avanço computacional. Esta área reúne um conjunto de métodos estatísticos úteis na análise de dados relacionados ao tempo até a ocorrência de um evento de interesse. Esses dados são chamados de tempos de sobrevivência, de falha ou simplesmente de tempos de vida. Analisar o tempo de vida é um assunto

extremamente interessante para pesquisadores das áreas de biologia, medicina, engenharia, entre outras áreas. Exemplos de tempos de falha incluem os períodos de vida de máquinas industriais até a quebra, e os tempos de sobrevivência de pacientes em um estudo clínico até a morte ou cura.

### 1.1.1 Funções básicas

O comportamento da variável aleatória contínua e não negativa, tempo de sobrevivência,  $T \geq 0$ , pode ser definido pelas seguintes funções: função densidade de probabilidade  $f(t)$ , função de sobrevivência  $S(t)$ , e função de risco  $h(t)$ .

Essas três funções podem ser utilizadas para descrever os diferentes aspectos apresentados pelo conjunto de dados de tempos de sobrevivência (Lawless, 2002).

A função densidade de probabilidade é definida como o limite da probabilidade de um indivíduo falhar (morrer) no intervalo de tempo  $[t, t + \Delta t)$  por unidade de tempo, e é expressa por

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t},$$

em que  $f(t) \geq 0$ , para todo  $t$  e tem a área abaixo da curva igual a 1.

A função de sobrevivência  $S(t)$  é definida como a probabilidade de um indivíduo sobreviver mais que um determinado tempo  $t$ , ou seja, não falhar após um certo tempo, e é expressa por

$$S(t) = P(T > t) = 1 - F(t),$$

em que  $F(t) = \int_0^t f(u) du$  representa a função de distribuição acumulada. Quando  $t \rightarrow \infty$ ,  $S(t) = 0$  e quando  $t = 0$ ,  $S(t) = 1$ .

A função de risco  $h(t)$ , também conhecida como taxa de falha, é definida como o limite da probabilidade de um indivíduo falhar no intervalo de tempo  $[t, t + \Delta t]$ , dado que o mesmo tenha sobrevivido até o tempo  $t$ , e é expressa por

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

Esta função representa a probabilidade de falha (risco) de um determinado indivíduo e descreve como a probabilidade instantânea de falha se modifica com o passar do tempo. Pode apresentar diferentes comportamentos, como crescente, decrescente, constante ou uma combinação de comportamentos.

### 1.1.2 Método de máxima verossimilhança

Supondo que o tempo de sobrevivência  $T$  tem função densidade de probabilidade  $f(t)$ , a função de verossimilhança  $L$ , obtida para todos os tempos observados, se não houver censura, será

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i, \boldsymbol{\theta}).$$

Entretanto, cada observação censurada nos informa apenas que o tempo de falha é maior que o tempo de censura observado e, portanto, que a sua contribuição para  $L$  é a sua função de sobrevivência  $S(t)$ . No contexto de censura tipo I, as observações podem ser divididas em dois conjuntos, as  $r$  primeiras são as não censuradas  $\{1, 2, \dots, r\}$  e, as  $n - r$  seguintes, são as censuradas  $\{r + 1, r + 2, \dots, n\}$ .

No processo de construção de uma função de verossimilhança para dados censurados, uma suposição fundamental a respeito dos tempos em que há censura, é a de que eles são independentes dos tempos de ocorrência do evento. Essa suposição equivale a dizer que as censuras são não-informativas ou não estão relacionadas ao evento em questão.

Assim, a função de verossimilhança considerando censura de tipo I, é dada por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f(t_i, \boldsymbol{\theta})]^{\delta_i} [S(t_i, \boldsymbol{\theta})]^{1-\delta_i}. \quad (1.1)$$

equivalente a

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [h(t_i, \boldsymbol{\theta})]^{\delta_i} [S(t_i, \boldsymbol{\theta})], \quad (1.2)$$

em que  $\delta_i$  é a variável indicadora de censura.

Os estimadores de máxima verossimilhança são os valores de  $\boldsymbol{\theta}$ , vetor de parâmetros que maximizam  $L$ , ou de forma equivalente:  $\log(L(\boldsymbol{\theta})) = l(\boldsymbol{\theta})$ . Os estimadores são encontrados resolvendo

$$U(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0.$$

## 1.2 Algumas distribuições

Nesta seção apresentamos algumas distribuições de probabilidade que são bastante usadas para modelar, de forma plausível, dados relacionados aos tempos de sobrevivência.

### 1.2.1 Distribuição exponencial

A distribuição exponencial é uma das mais simples e importantes distribuições utilizadas na modelagem de dados que representam o tempo até a ocorrência de algum evento de interesse. A mesma tem sido utilizada intensivamente no campo de sobrevivência. Uma variável aleatória não-negativa e contínua  $T$ , tem distribuição exponencial,  $T \sim \text{Exp}(\lambda)$ , com parâmetro  $\lambda > 0$ , se sua função densidade é dada por

$$f(t|\lambda) = \lambda e^{-\lambda t}, \quad (1.3)$$

com funções de sobrevivência e de risco dadas, respectivamente, por

$$S(t|\lambda) = e^{-\lambda t} \quad \text{e} \quad h(t|\lambda) = \lambda. \quad (1.4)$$

em que  $\lambda$  representa a taxa de falha da distribuição.

Observamos que a função de risco é constante, isto é, não depende do tempo. Apresentamos os gráficos das funções da distribuição exponencial, densidade-Figura 1.1(a); sobrevivência Figura-1.1(b); risco-Figura 1.1(c), considerando diferentes valores para o parâmetro  $\lambda$ .

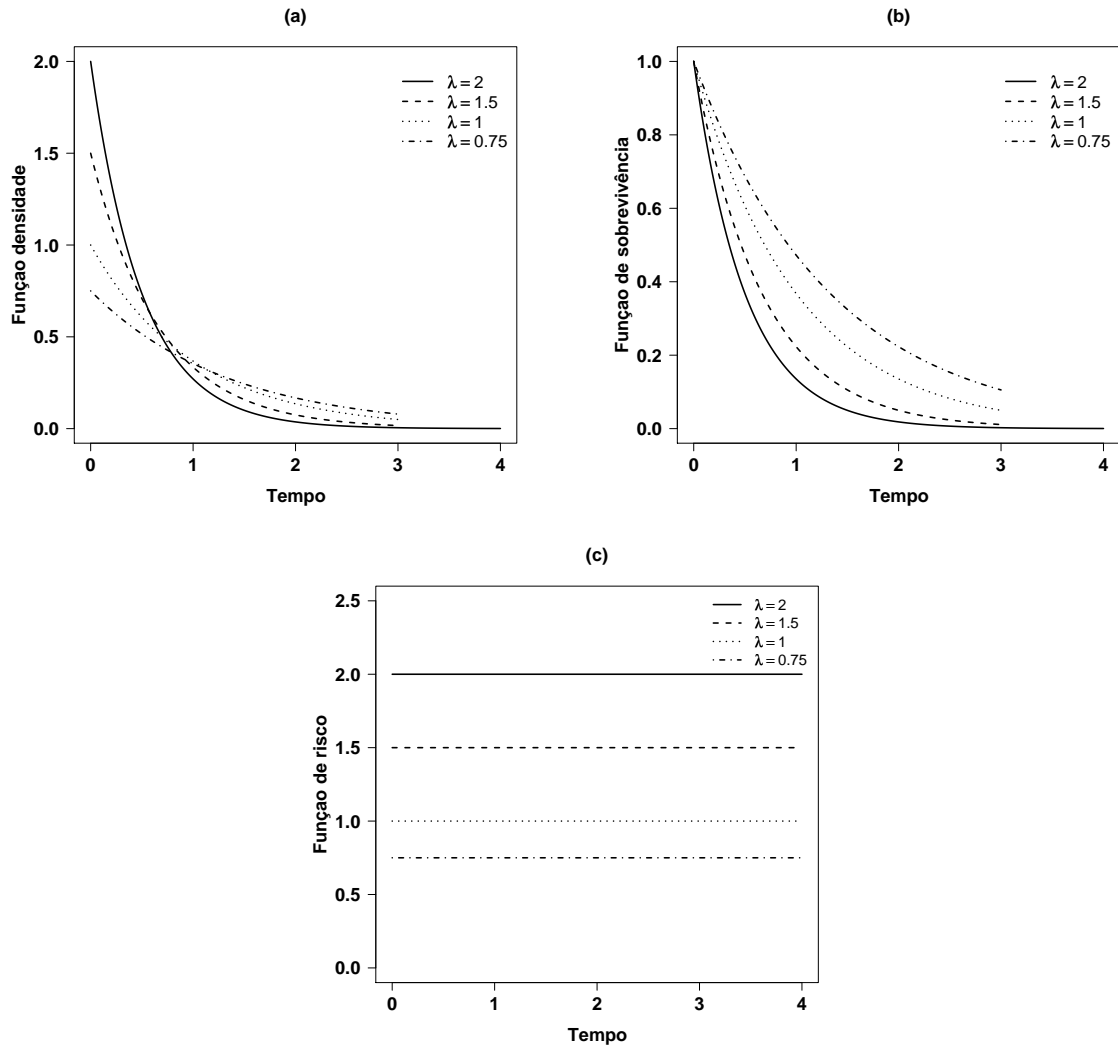


Figura 1.1: (a) Função densidade, (b) função de sobrevivência, (c) função de risco, da distribuição exponencial para diferentes valores de  $\lambda$ .

### 1.2.2 Distribuição Weibull

A distribuição Weibull é amplamente utilizada como modelo em aplicações biométricas, industriais e em muitas outras áreas. Uma das características importantes da distribuição Weibull na modelagem de tempos de sobrevivência ou confiabilidade, está relacionada à sua flexibilidade em acomodar diferentes formas na taxa de falha (função de risco), apenas variando os valores de seus parâmetros.

Uma variável aleatória não-negativa e contínua  $T$ , tem distribuição Weibull



com parâmetros  $\alpha > 0$  e  $\lambda > 0$ ,  $\text{Wei}(\alpha, \lambda)$ , se sua função densidade é

$$f(t|\alpha, \lambda) = \alpha \lambda t^{\alpha-1} \exp(-\lambda t^\alpha). \quad (1.5)$$

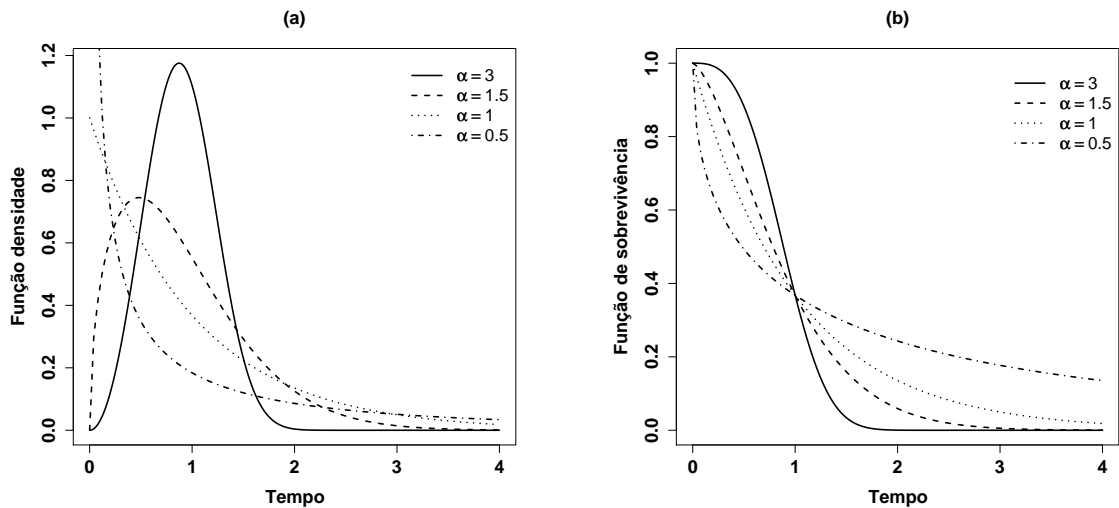
As funções de sobrevivência e de risco da distribuição Weibull, são dadas, respectivamente, por

$$S(t|\alpha, \lambda) = \exp(-\lambda t^\alpha) \quad \text{e} \quad h(t|\alpha, \lambda) = \alpha \lambda t^{\alpha-1}. \quad (1.6)$$

Considerando a parametrização  $\lambda = \exp(\lambda^*)$ ,  $\lambda^* \in \mathbb{R}$ , a função densidade 1.5 é reescrita como

$$f(t|\alpha, \lambda^*) = \alpha t^{\alpha-1} \exp(\lambda^* - t^\alpha e^{\lambda^*}). \quad (1.7)$$

Determinados valores do parâmetro  $\alpha$  fornecem alguns casos particulares, por exemplo:  $\alpha = 1$  leva à distribuição exponencial,  $\alpha = 2$ , corresponde à distribuição Rayleigh e, assim, obtemos uma aproximação das distribuições Log-Normal e Normal para  $\alpha = 2, 5$  e  $\alpha = 3, 6$ , respectivamente. Os comportamentos das funções densidade, de sobrevivência e de risco da distribuição de Weibull, considerando  $\lambda = 1$  estão representados na Figura 1.2(a),(b) e (c).



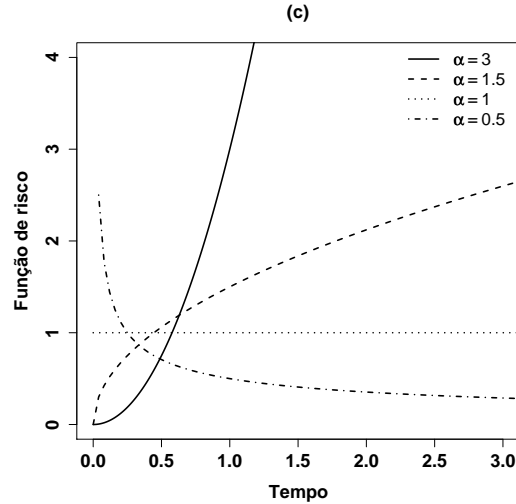


Figura 1.2: (a) Função densidade, (b) função de sobrevivência, (c) função de risco, da distribuição Weibull para diferentes valores de  $\alpha$ .

### 1.2.3 Distribuição Log-normal

Uma variável aleatória não-negativa e contínua  $T$ , tem distribuição log-normal com parâmetros  $\mu \in \mathbb{R}$  e  $\sigma > 0$ , se sua função densidade é

$$f(t|\mu, \sigma) = \frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, \quad (1.8)$$

em que  $\mu$  é a média do logaritmo do tempo de falha, e  $\sigma$  é o desvio-padrão.

As funções de sobrevivência e de risco da distribuição log-normal são dadas, respectivamente, por

$$S(t|\mu, \sigma) = \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right) \quad \text{e} \quad h(t|\mu, \sigma) = \frac{f(t|\mu, \sigma)}{S(t|\mu, \sigma)}, \quad (1.9)$$

em que  $\Phi$  é a função de distribuição acumulada de uma distribuição normal padrão.

Os comportamentos das funções densidade, de sobrevivência e de risco da distribuição de probabilidade log-normal, considerando que  $\mu = 0$ , estão representados na Figura 1.3(a),(b) e (c).

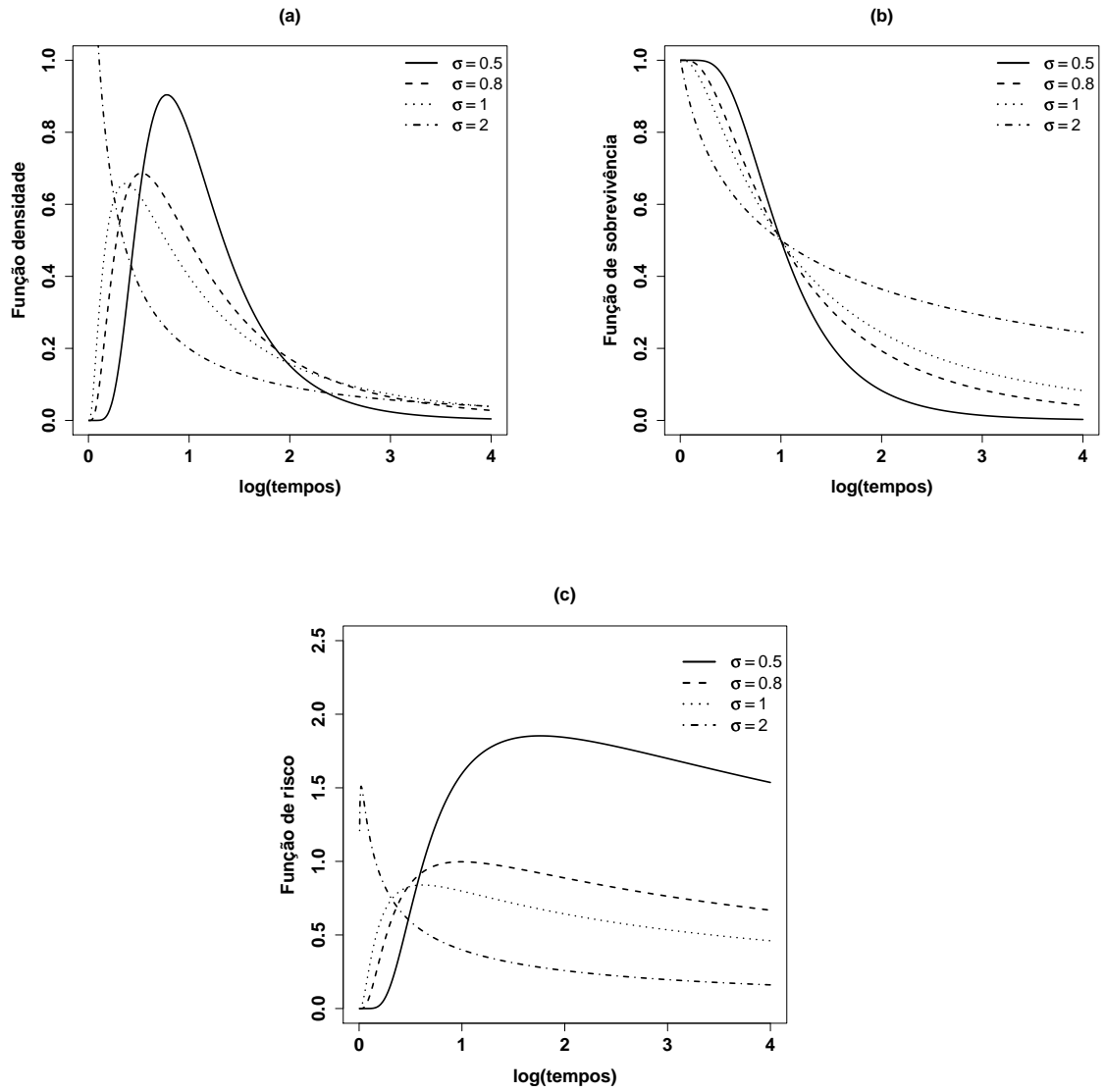


Figura 1.3: (a) Função densidade, (b) função de sobrevivência, (c) função de risco, da distribuição log-normal para diferentes valores de  $\sigma$ .

### 1.2.4 Distribuição Log-logística

Uma variável aleatória não negativa e contínua  $T$ , tem distribuição log-logística com parâmetro de forma  $\alpha > 0$  e de escala  $\gamma > 0$ , se sua função densidade é dada por

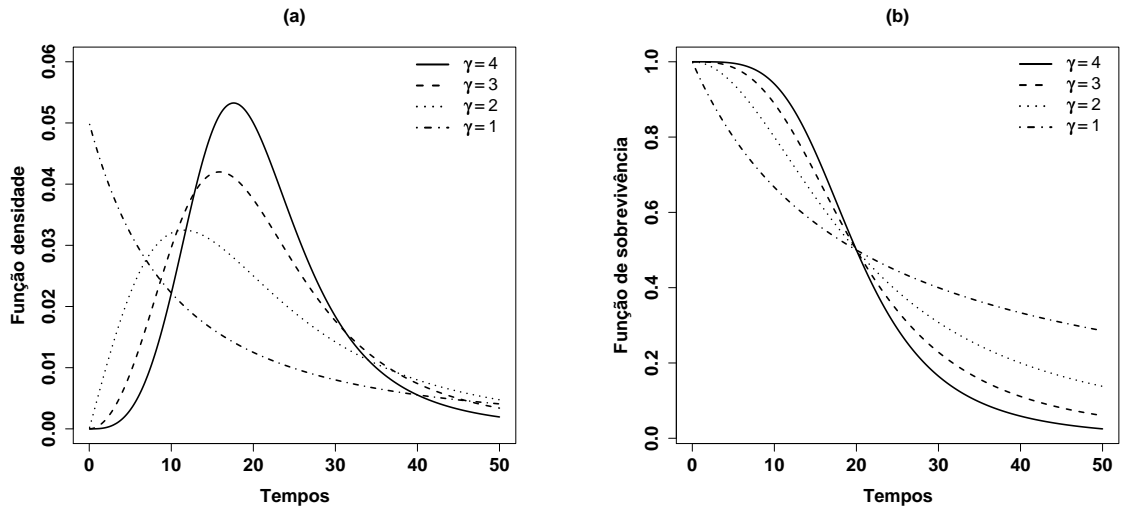
$$f(t|\alpha, \gamma) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} (1 + (t/\alpha)^\gamma)^{-2}. \quad (1.10)$$

As funções de sobrevivência e de risco da distribuição de probabilidade log-logística são dadas, respectivamente, por

$$S(t|\alpha, \gamma) = \frac{1}{1 + (t/\alpha)^\gamma} \quad \text{e} \quad h(t|\alpha, \gamma) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha [1 + (t/\alpha)^\gamma]}. \quad (1.11)$$

É importante observar que se  $T$  tem distribuição log-logística, então, o logaritmo de  $T$ ,  $\log(T)$ , segue a distribuição de probabilidade logística.

Os comportamentos das funções densidade, de sobrevivência e de risco da distribuição de probabilidade log-logística considerando  $\alpha = 20$ , estão representados na Figura 1.4(a),(b) e (c).



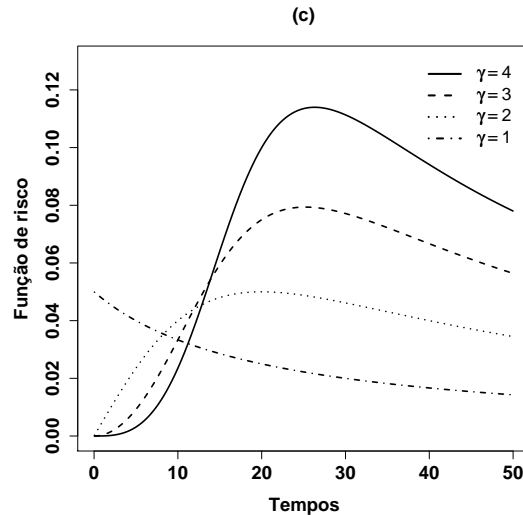


Figura 1.4: (a) Função densidade, (b) função de sobrevivência, (c) função de risco, da distribuição log-logística para diferentes valores de  $\gamma$ .

### 1.3 Objetivos do trabalho

Para a análise de dados de longa duração, isto é, quando se admite uma porcentagem de não-ocorrência do evento de interesse na população, vários modelos foram formulados sendo alguns mais antigos, como por exemplo, Berkson & Gage (1952) e outros mais recentes, como os apresentados por Chen *et al.* (1999) e Rodrigues *et al.* (2009b).

Enfatizamos o objetivo de desenvolver uma modelagem bayesiana para modelos de sobrevivência de longa duração na presença de covariáveis. Neste contexto, apresentamos a metodologia do modelo de partição bayesiano desenvolvida por Holmes *et al.* (1999) e a extensão dos modelos de longa duração, utilizando o modelo de partição bayesiano proposto por Hoggart & Griffin (2001), metodologia na qual fazer predição é central e, portanto, também o é neste trabalho. Desenvolvemos, ainda, uma estratégia computacional para esta extensão, considerando variáveis contínuas e categóricas, em que a estratégia computacional escolhe de forma natural as variáveis preditoras que realmente têm efeito na variável resposta. Desenvolvemos um estudo de simulação para verificar o comportamento do modelo e aplicamos a metodologia a um conjunto de dados reais (Kirkwood

---

*et al.*, 2000; Ibrahim *et al.*, 2001)

Este trabalho está organizado da seguinte forma: no Capítulo 2, apresentamos os fatos relevantes dos modelos de longa duração, do ponto de vista da teoria unificada de análise de sobrevivência desenvolvido por Rodrigues *et al.* (2009a). No Capítulo 3, apresentamos o modelo de partição bayesiana proposto por Holmes *et al.* (1999) e a extensão local para o modelos de longa duração. No Capítulo 4, apresentamos o estudo de simulação e um exemplo com dados reais, utilizando a metodologia descrita ao longo do trabalho. No Capítulo 5, encontram-se as considerações finais e as propostas de trabalho.

## Capítulo 2

# Modelos de Longa Duração

Neste capítulo apresentamos uma introdução aos modelos de longa duração por meio da teoria unificada de análise de sobrevivência proposta por Rodrigues *et al.* (2009a).

Os modelos de sobrevivência de longa duração têm grande importância em análise de dados de sobrevivência e confiabilidade, e surgem em várias áreas, tais como: biomedicina, finanças, criminologia, confiabilidade industrial, entre outras. Por isso, diferentes métodos para ajustar tais modelos têm sido publicados na literatura. Diversos artigos têm abordado a questão dos dados de longa duração. Em confiabilidade industrial, o evento de interesse pode ser a falha de placas de circuito, devido a diferentes fatores de risco ou ao desgaste por uso, (Meeker & Escobar, 1998). Em dados financeiros, o evento de interesse pode ser o desligamento do cliente de um banco devido a várias causas (Hoggart & Griffin, 2001). Em dados biomédicos, o evento de interesse pode ser a morte de um paciente submetido a um certo tratamento, devido a diferentes causas competitivas ou à recorrência do tumor pela presença de um número desconhecido de células cancerígenas (Yakovlev & Tsodikov, 1996; Chen *et al.*, 1999; Tsodikov *et al.*, 2003).

## 2.1 Modelo de longa duração unificado

A teoria proposta por Rodrigues *et al.* (2009a), tem por objetivo unificar a análise de sobrevivência com o modelo clássico de Boag (1949) e Berkson & Gage (1952), e com os modelos mais recentes de longa duração de Yakovlev & Tsodikov (1996) e Chen *et al.* (1999). A unificação foi obtida através de uma composição da função geradora de probabilidade do número de causas de ocorrência do evento de interesse e da função de sobrevivência dos pacientes em risco (Feller, 1968). Também é mostrado que a função geradora de longa duração formulada, satisfaz a propriedade de riscos proporcionais se, e somente se, o número de causas relacionado à ocorrência do evento de interesse segue uma distribuição de Poisson.

Apresentamos os resultados e fatos mais relevantes desta teoria. No entanto, a definição fundamental desta teoria é a definição da função geradora de uma sequência de números reais encontradas em Feller (1968). Uma boa referência desta teoria, pode ser vista em Rodrigues *et al.* (2008) e Iritani (2008).

### 2.1.1 Função de sobrevivência

Seja  $N$  uma variável aleatória, que representa o número de causas ou riscos para um particular evento de interesse, com distribuição de probabilidade

$$p_{n^*} = P[N = n^*], \quad n^* = 0, 1, 2, \dots,$$

em que  $N$  é uma variável aleatória latente. Condicionado a  $N = n^*$ , sejam  $Z_v$ ,  $v = 1, \dots, n^*$ , variáveis aleatórias contínuas não-negativas e independentes, com função de distribuição  $F(\cdot) = 1 - S(\cdot)$ , sendo que  $N$  é independente de  $Z_v$ , onde  $Z_v$  representa o tempo de ocorrência de um particular evento de interesse, devido a  $v$ -ésima causa ou risco. O tempo de ocorrência do evento de interesse é definido como:

$$T = \min \{Z_1, Z_2, \dots, Z_N\}, \quad (2.1)$$

em que  $P[Z_0 = +\infty] = 1$ , leva a uma proporção  $p_0$  da população não-sujeita à ocorrência do evento de interesse. As variáveis  $Z_v$  são latentes e  $T$  é uma variável



aleatória observável ou censurada.

A função de sobrevivência da variável aleatória  $T$ , chamada de função de sobrevivência da população, é dada por:

$$S_{pop}(t) = P[T > t]. \quad (2.2)$$

Seguindo Feller (1968), apresentamos a definição da função geradora de uma seqüência de números reais, que permite o desenvolvimento da teoria unificada de análise de sobrevivência.

**Definição 2.1.** Seja  $\{a_n\}$  uma seqüência de números reais. Se para qualquer  $s \in [0, 1]$ ,

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

converge, então  $A(s)$  é definida como a função geradora da seqüência  $\{a_n\}$ .

**Teorema 2.1.** Dada uma função de sobrevivência própria  $S(t)$ , a função de sobrevivência da variável aleatória  $T$ , definida em (2.2), é dada por

$$S_{pop}(t) = A_p(S(t)) = \sum_{n^*=0}^{+\infty} p_{n^*} \{S(t)\}^{n^*}, \quad (2.3)$$

sendo que  $A_p(\cdot)$  é a função geradora da seqüência  $p = \{p_{n^*}\}$ , que converge se  $0 \leq S(t) \leq 1$ .

A função de sobrevivência  $S_{pop}(t)$  dada em (2.2), não é uma função de sobrevivência própria, e o teorema a seguir prova esta afirmação.

**Teorema 2.2.** Dada uma função de sobrevivência própria  $S(t)$ ,

$$\lim_{t \rightarrow \infty} S_{pop}(t) = P[N = 0] = p_0, \quad (2.4)$$

em que  $p_0$  denota a proporção de curados ou imunes que podem estar presentes na população da qual os dados são tomados.

Por outro lado, notemos que a derivada da função de sobrevivência da população (2.3), satisfaz a relação:

$$-\frac{dS_{pop}(t)}{dt} = f(t) \sum_{n^*=1}^{\infty} n^* p_{n^*} \{S(t)\}^{n^*-1}, \quad (2.5)$$

em que  $f(t)$  é uma função de densidade própria. Considerando o Teorema (2.2) e tendo que  $0 \leq S(t) \leq 1$ , a série em (2.5) é, então, convergente, de modo que  $-dS_{pop}/dt|_{t=0} = f(0)E[N]$ .

As funções de densidade e de risco associadas à função de sobrevivência de longa duração são dadas, respectivamente por:

$$f_{pop}(t) = f(t) \frac{dA_p(s)}{ds} \Big|_{s=S(t)} \quad (2.6)$$

e

$$h_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)} = f(t) \frac{\left\{ \frac{dA_p(s)}{ds} \Big|_{s=S(t)} \right\}}{S_{pop}(t)}. \quad (2.7)$$

Notemos que em (2.7) não ocorre a propriedade de riscos proporcionais posto que

$$\frac{\left\{ \frac{dA_p(s)}{ds} \Big|_{s=S(t)} \right\}}{S_{pop}(t)}$$

depende de  $t$ . No entanto, uma exceção acontece quando a variável aleatória  $N$  tem a distribuição de Poisson como parâmetro  $\theta > 0$  (Chen *et al.*, 1999).

A teoria unificada de análise de sobrevivência está caracterizada mediante o seguinte teorema:

**Teorema 2.3.** A função de risco imprópria de longa duração  $h_{pop}(t)$ , tem a propriedade de riscos proporcionais se, e somente se,  $p_{n^*} = \frac{e^{-\theta} \theta^{n^*}}{n^*!}$ ,  $n^* = 0, 1, \dots$ , isto é,  $N \sim \text{Poisson}(\theta)$ .

As provas dos teoremas acima podem ser encontradas em Rodrigues *et al.* (2009a).

Exemplos de funções geradoras de probabilidade, associadas a algumas distribuições de probabilidade (Feller, 1968), para  $0 \leq s \leq 1$ , são:

- Bernoulli:  $p_{n^*} = (1 - \theta)^{n^*} \theta^{1-n^*}$ ,  $0 < \theta < 1$ ,  $n^* = 0, 1$ . Então,

$$A_p(s) = \theta + (1 - \theta)s. \quad (2.8)$$

- Geométrica:  $p_{n^*} = (1 - \theta)^{n^*} \theta$ ,  $0 < \theta < 1$ ,  $n^* = 0, 1, \dots$ . Então,

$$A_p(s) = \frac{\theta}{1 - (1 - \theta)^s}. \quad (2.9)$$

- Poisson:  $p_{n^*} = \frac{e^{-\theta}\theta^{n^*}}{n^*!}$ ,  $\theta > 0$ ,  $n^* = 0, 1, \dots$ . Então,

$$A_p(s) = \exp\{-\theta(1-s)\}. \quad (2.10)$$

- Binomial negativa: Considerando a versão mais geral com parâmetros  $\eta$  e  $\theta$ , a distribuição binomial negativa (Saha & Paul, 2005) é dada por

$$p_{n^*} = \frac{\Gamma(\eta^{-1} + n^*)}{\Gamma(\eta^{-1})n^*!} \left(\frac{\eta\theta}{1 + \eta\theta}\right)^{n^*} (1 + \eta\theta)^{-\frac{1}{\eta}}$$

$\theta > 0$  e  $\eta > -1/\theta$ ,  $n^* = 0, 1, \dots$ . Então,

$$A_p(s) = \{1 + \eta\theta(1-s)\}^{-\frac{1}{\eta}}. \quad (2.11)$$

As distribuições de probabilidade Bernoulli, Poisson e Geométrica, estão relacionados com a distribuição Binomial negativa, das seguintes formas:  $\eta = -1$ , a distribuição Binomial negativa fornece a distribuição Bernoulli com parâmetro  $(1 - \theta)$ ; se  $\eta = 0$ , temos a distribuição de Poisson com parâmetro  $\theta$ , e se  $\eta = 1$ , temos a distribuição geométrica com parâmetro  $\frac{1}{1-\theta}$ .

Apresentamos na Tabela 2.1 as funções de densidade, de sobrevivência e de risco, considerando diferentes distribuições para o número de causas do evento de interesse, junto com a fração de cura para cada uma das distribuições de  $N$  (Rodrigues *et al.*, 2008).

Tabela 2.1: Função de sobrevivência  $S_{pop}(t)$ , função de densidade  $f_{pop}(t)$ , função de risco  $h_{pop}(t)$  e fração de cura para diferentes distribuições do número de causas,  $N$ .

	Distribuição de N		
	Binomial negativa( $\eta, \theta$ )	Poisson( $\theta$ )	Bernoulli( $1 - \theta$ )
$S_{pop}(t)$	$\{1 + \eta\theta F(t)\}^{-1/\eta}$	$\exp(-\theta F(t))$	$1 - (1 - \theta)F(t)$
$f_{pop}(t)$	$\{1 + \eta\theta F(t)\}^{-1/\eta-1} \theta f(t)$	$\theta f(t) \exp(-\theta F(t))$	$(1 - \theta)f(t)$
$h_{pop}(t)$	$\{1 + \eta\theta F(t)\}^{-1} \theta f(t)$	$\theta f(t)$	$\frac{(1-\theta)}{1-(1-\theta)F(t)} f(t)$
$p_0$	$\{1 + \eta\theta\}^{-1/\eta}$	$\exp(-\theta)$	$\theta$

Neste trabalho, consideramos a distribuição de Poisson para a variável latente  $N$ .

### 2.1.2 Função de verossimilhança

Seja  $N_i$  o número de causas ou riscos para a ocorrência do evento de interesse da  $i$ -ésima observação,  $i = 1, \dots, n$ , com distribuição de probabilidade  $p_{n_i^*} = P[N_i = n_i^*]$ , sendo  $E[N_i] = \mu_i > 0$ ,  $i = 1, \dots, n$ . Na presença de covariáveis, seja  $\mathbf{x}_i^\top = (x_{i0}, x_{i1}, \dots, x_{ip})$  o vetor de covariáveis associado a cada observação da amostra, relacionamos as médias a covariáveis ( $\mathbf{x}_i$ ) por meio de uma função de ligação bijetiva na forma  $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ , em que  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$  é o vetor de coeficientes da regressão, de modo que a probabilidade de cura da  $i$ -ésima unidade  $p_{0i} = P[N_i = 0]$ , depende do vetor de parâmetros da distribuição de  $N_i$ , denotado por  $\boldsymbol{\phi}^*$ , composto por  $\boldsymbol{\beta}$  e pelos demais parâmetros da distribuição de probabilidade de  $N_i$ , se existirem.

Sejam  $Z_{i1}, \dots, Z_{iN_i}$ , variáveis aleatórias contínuas e i.i.d que representam o tempo de ocorrência do evento de interesse, para  $N_i$  o número de causas ou riscos no  $i$ -ésimo elemento da amostra, com função de distribuição acumulada  $F(\cdot|\boldsymbol{\gamma}) = 1 - S(\cdot|\boldsymbol{\gamma})$ , em que  $\boldsymbol{\gamma}$  denota o vetor de parâmetros da distribuição.

Sejam  $T_i = \{Z_{i1}, \dots, Z_{iN_i}\}$ , e  $C_i$  o tempo da censura e seja  $Y_i = \min\{T_i, C_i\}$  o tempo observado, e  $\delta_i$  a variável indicadora de censura em que  $\delta_i = 1$  se  $Y_i = T_i$ , e  $\delta_i = 0$ , caso contrário.

Denotamos os dados observados por  $\mathbf{D}_{obs} = (n, \mathbf{X}, \mathbf{y}, \boldsymbol{\delta})$ , onde  $n$  é o número de observações,  $\mathbf{y} = (y_1, \dots, y_n)^\top$ ,  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$  e  $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ , a matriz das covariáveis de ordem  $n \times (p + 1)$ .

Denotamos os dados ampliados por  $\mathbf{D} = (n, \mathbf{X}, \mathbf{y}, \boldsymbol{\delta}, \mathbf{N})$ , sendo  $n, \mathbf{X}, \mathbf{y}, \boldsymbol{\delta}$ , como apresentado nos dados observados ( $\mathbf{D}_{obs}$ ), e  $\mathbf{N} = (N_1, \dots, N_n)$  o vetor de variáveis latentes.

Seja  $\boldsymbol{\varphi} = (\boldsymbol{\gamma}^\top, \boldsymbol{\phi}^{*\top})^\top$  o vetor de parâmetros associado ao modelo.

Desta forma a função de verossimilhança para dados ampliados com censura não-informativa, é dada por

$$L(\boldsymbol{\varphi}; \mathbf{D}) \propto \prod_{i=1}^n \{S(y_i|\boldsymbol{\gamma})\}^{n_i^* - \delta_i} \{n_i^* f(y_i|\boldsymbol{\gamma})\}^{\delta_i} p_{n_i^*}(\boldsymbol{\phi}^*). \quad (2.12)$$

A função de verossimilhança unificada, é definida como:

$$L(\boldsymbol{\varphi}; \mathbf{D}_{obs}) \propto \sum_{\mathbf{n}^*} \prod_{i=1}^n \{S(y_i|\boldsymbol{\gamma})\}^{n_i^* - \delta_i} \{n_i^* f(y_i|\boldsymbol{\gamma})\}^{\delta_i} p_{n_i^*}(\boldsymbol{\phi}^*), \quad (2.13)$$

sendo que o somatório percorre todos os possíveis valores do vetor  $\mathbf{n}^* = (n_1^*, \dots, n_n^*)^\top$ .

No entanto, a função de verossimilhança unificada pode ser escrita como (Rodrigues *et al.*, 2008)

$$L(\boldsymbol{\varphi}, \mathbf{D}_{obs}) \propto \prod_{i=1}^n \{f_{pop}(y_i|\boldsymbol{\varphi})\}^{\delta_i} \{S_{pop}(y_i|\boldsymbol{\varphi})\}^{1-\delta_i} \quad (2.14)$$

$$\propto \prod_{i \in \bar{\mathbf{C}}} f_{pop}(y_i|\boldsymbol{\varphi}) \prod_{i \in \mathbf{C}} S_{pop}(y_i|\boldsymbol{\varphi}), \quad (2.15)$$

em que  $\bar{\mathbf{C}}$  e  $\mathbf{C}$  são os conjuntos de dados não censurados e censurados respectivamente e  $f_{pop}(\cdot)$  e  $S_{pop}(\cdot)$  são funções dadas em (2.6) e (2.3), respectivamente. Para detalhes da construção da função de verossimilhança para dados ampliados e a função de verossimilhança unificada, recomendamos Iritani (2008).

## 2.2 Modelo de mistura padrão

O modelo de mistura padrão introduzido por Boag (1949) e Berkson & Gage (1952), é um dos mais utilizados na literatura estatística, sobretudo na área médica. Baseado em Boag (1949) e, após acompanhar estudos de câncer de estômago, Berkson & Gage (1952) propuseram uma função simples através de dois parâmetros fisicamente significativos, usados para comparar a mortalidade de dois grupos, por exemplo, diferença entre tratamentos, tipos de câncer, entre outros. Um termo representa a proporção da população sujeita apenas à morte natural ( $p_0$ ) e o outro está relacionado a riscos de morte pelo câncer ( $1 - p_0$ ). Verificou-se depois de um período de acompanhamento, que tiveram uma razão de morte próxima a de uma população normal, no sentido que a população não tem a doença do câncer.

A relação da função de sobrevivência de longa duração  $S_{pop}(t)$  com o modelo de Berkson & Gage (1952), é justificada pelo seguinte teorema.

**Teorema 2.4.** A função de longa duração é dada por

$$S_{pop}(t) = p_0 + (1 - p_0)S_{pop}^*(t), \quad (2.16)$$

em que

$$S_{pop}^*(t) = \sum_{n^*=0}^{+\infty} p_{n^*} \{S(t)\}^{n^*} \text{ e } p_{n^*} = \frac{p_{n^*}}{1-p_0}.$$

Notemos que  $S_{pop}^*(t)$  é uma função de sobrevivência própria, associada aos elementos da população que está em risco, isto é, sujeita à ocorrência do evento de interesse.

A função de densidade e risco própria associada a (2.16), são respectivamente:

$$f^*(t) = f(t) \frac{\left\{ \frac{dA_p(s)}{ds} \Big|_{s=S(t)} \right\}}{1 - p_0} \quad (2.17)$$

e

$$h^*(t) = \frac{S_{pop}(t)}{S_{pop}(t) - p_0} h_{pop}(t) \quad (2.18)$$

Como  $\frac{S_{pop}(t)}{S_{pop}(t) - p_0} > 1$  é a função de risco para a população em risco,  $h^*(t)$  é maior que a função de risco da população  $h_{pop}(t)$ . Notamos que (2.18) não tem a propriedade de riscos proporcionais, porque  $\frac{S_{pop}(t)}{S_{pop}(t) - p_0}$  sempre depende de  $t$ .

Quando o número de causas do evento de interesse segue uma distribuição de Bernoulli com parâmetro  $1 - \theta$ , então existe só uma causa para o evento de interesse. Desta forma, dada a função geradora de probabilidades da distribuição Bernoulli  $A_p(s) = \theta + (1 - \theta)s$ , e considerando-se o Teorema 2.1 obtemos as funções de sobrevivência, densidade e de risco para a população representadas, respectivamente, por:

$$S_{pop}(t) = A_p(S(t)) = \theta + (1 - \theta)S(t), \quad (2.19)$$

$$f_{pop}(t) = -\frac{dS_{pop}(t)}{dt} = (1 - \theta)f(t), \quad (2.20)$$

e

$$h_{pop}(t) = f(t) \frac{1 - \theta}{\theta + (1 - \theta)S(t)}, \quad (2.21)$$

em que a fração de cura, é dada por  $p_0 = \lim_{t \rightarrow \infty} S_{pop}(t) = \theta$ .

Portanto, para a população em risco a função de sobrevivência, a função de densidade e a função de risco são dadas por:

$$S_{pop}^*(t) = S(t), \quad f_{pop}^*(t) = f(t), \quad \text{e} \quad h_{pop}^*(t) = \frac{f(t)}{S(t)}. \quad (2.22)$$

Considerando a função de verossimilhança com dados ampliados em (2.12), onde a distribuição de probabilidade para  $N$  (o número de causas do evento de interesse) tem a distribuição de Bernoulli com parâmetro  $\phi^* = 1 - \theta$  e com censura não informativa, obtemos a função de verossimilhança com dados ampliados para o modelo de mistura padrão:

$$L(\boldsymbol{\varphi}; \mathbf{D}) = \prod_n^{i=1} \{S(y_i|\boldsymbol{\gamma})\}^{N_i - \delta_i} \{N_i f(y_i|\boldsymbol{\gamma})\}^{\delta_i} (1 - \theta)^{N_i} \theta^{1 - N_i} \quad (2.23)$$

onde  $\boldsymbol{\varphi} = (\boldsymbol{\gamma}, 1 - \theta)^\top$ ,  $N_i = 0, 1$ ,  $i = 1, \dots, n$ , e  $\delta_i$  é a variável indicadora de censura.

Da mesma forma, considerando a função de verossimilhança para dados observados em (2.14), (2.19) e (2.20), obtemos a função de verossimilhança unificada para o modelo de mistura padrão, dado por:

$$\begin{aligned} L(\boldsymbol{\varphi}; \mathbf{D}_{obs}) &= \prod_{i=1}^n \{(1 - \theta)f(y_i|\boldsymbol{\gamma})\}^{\delta_i} \{\theta + (1 - \theta)S(y_i|\boldsymbol{\gamma})\}^{1 - \delta_i} \\ &= (1 - \theta)^d \prod_{i \in \bar{\mathcal{C}}} f(y_i|\boldsymbol{\gamma}) \prod_{i \in \mathcal{C}} \{\theta + (1 - \theta)S(y_i|\boldsymbol{\gamma})\}, \end{aligned} \quad (2.24)$$

em que  $d = \sum_{i=1}^n \delta_i$ .

O modelo de mistura padrão apresenta algumas desvantagens. Chen *et al.* (1999) discute as desvantagens do modelo considerando uma abordagem bayesiana; uma delas inside no fato de que na presença de covariáveis, através do parâmetro  $\theta$  via o modelo de regressão binomial, a distribuição *a posteriori* é imprópria, mesmo que se tenha uma distribuição *a priori* própria para os coeficientes do modelo de regressão. Neste contexto, apresentamos na próxima seção o modelo de longa duração estudado por Chen *et al.* (1999), como uma alternativa ao modelo de mistura padrão.

## 2.3 Modelo de tempo de promoção

O modelo de longa duração em que o número de causas do evento de interesse tem a distribuição de probabilidade de Poisson, foi estudado por Yakovlev & Tsodikov (1996) e Chen *et al.* (1999). Este modelo de longa duração é conhecido na literatura estatística como modelo de tempo de promoção.

Considerando a função geradora de probabilidade da distribuição de Poisson (2.10) e o Teorema 2.1, obtemos as funções de sobrevivência, densidade e de risco da população, dadas respectivamente por:

$$S_{pop}(t) = A_p(S(t)) = \exp\{-\theta[1 - S(t)]\} = \exp\{-\theta F(t)\}, \quad (2.25)$$

$$f_{pop}(t) = -\frac{dS_{pop}(t)}{dt} = \theta f(t) \exp\{-\theta F(t)\} \quad (2.26)$$

e

$$h_{pop}(t) = \theta f(t). \quad (2.27)$$

Assim, de (2.25) temos a fração de cura dada por  $p_0 = \lim_{t \rightarrow \infty} S_{pop}(t) = \exp(-\theta)$ .

Portanto, para a população em risco, as funções de sobrevivência, densidade e de risco, são dadas respectivamente por:

$$S^*(t) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}, \quad (2.28)$$

$$f^*(t) = \left( \frac{\exp(-\theta F(t))}{1 - \exp(-\theta)} \right) \theta f(t) \quad (2.29)$$

e

$$h^*(t) = \left( \frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)} \right) h_{pop}(t). \quad (2.30)$$

Na presença de covariáveis  $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip})^\top$ , relacionamos a média dos números de causas do evento de interesse  $E[N_i] = \theta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$ ,  $i = 1, \dots, n$ , sendo que  $\boldsymbol{\beta}$  é o vetor de coeficientes da regressão.



Assim, a função de risco da população dada em (2.27), tem a propriedade de riscos proporcionais. Isto é justificado pelo Teorema 2.3.

A relação matemática do modelo de tempo de promoção com o modelo de mistura padrão, é dada por:

$$S_{pop}(t) = \exp(-\theta) + (1 - \exp(-\theta))S^*(t), \quad (2.31)$$

em que  $S^*(t)$  é dada por (2.28). Desse modo,  $S_{pop}(t)$  tem a forma do modelo de mistura padrão com fração de cura  $p_0 = \exp(-\theta)$ .

A função de verossimilhança para os dados ampliados é

$$L(\boldsymbol{\gamma}, \theta | \mathbf{D}) \propto \left\{ \prod_{i=1}^n S(y_i | \boldsymbol{\gamma})^{N_i - \delta_i} (N_i f(y_i | \boldsymbol{\gamma}))^{\delta_i} \right\} \\ \times \exp \left\{ \sum_{i=1}^n [N_i \log(\theta) - \log(N_i!) - n\theta] \right\}. \quad (2.32)$$

Considere a função de ligação de  $\theta$  com as covariáveis dadas por

$$\theta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad (2.33)$$

assim a fração de cura para a  $i$ -ésima observação é

$$\exp(-\theta_i) = \exp(-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})), \quad (2.34)$$

com  $i = 1, \dots, n$ . Então (2.32) pode ser reescrito como,

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta} | \mathbf{D}) \propto \left( \prod_{i=1}^n S(y_i | \boldsymbol{\gamma})^{N_i - \delta_i} (N_i f(y_i | \boldsymbol{\gamma}))^{\delta_i} \right) \\ \times \exp \left\{ \sum_{i=1}^n N_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log(N_i!) - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right\}. \quad (2.35)$$

### 2.3.1 Modelo de tempo de promoção com distribuição Weibull

Suponha que as variáveis não-observáveis  $Z_{i1}, \dots, Z_{iN_i}$ ,  $i = 1, \dots, n$  têm distribuição de probabilidade de Weibull com vetor parâmetros  $\boldsymbol{\gamma} = (\lambda, \alpha)^\top$  como em (1.5). Então a função de verossimilhança em (2.35) considerando dados ampliados  $\mathbf{D} = (n, \mathbf{y}, \mathbf{N}, \mathbf{X}, \boldsymbol{\delta})$  é dada por

$$\begin{aligned}
L(\boldsymbol{\gamma}, \boldsymbol{\beta} | \mathbf{D}) &\propto \prod_{i=1}^n \exp(-\lambda N_i y_i^\alpha) \{N_i \alpha \lambda y_i^{\alpha-1}\}^{\delta_i} \\
&\times \exp \left\{ \sum_{i=1}^n N_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log(N_i!) - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right\} \quad (2.36)
\end{aligned}$$

A função de verossimilhança, considerando dados observados  $\mathbf{D}_{obs}$ , é dada por:

$$\begin{aligned}
L(\boldsymbol{\gamma}, \boldsymbol{\beta} | \mathbf{D}_{obs}) &\propto \sum_{\mathbf{N}} \left( \prod_{i=1}^n \exp(-\lambda N_i y_i^\alpha) \{N_i \alpha \lambda y_i^{\alpha-1}\}^{\delta_i} \right) \\
&\times \exp \left\{ \sum_{i=1}^n N_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log(N_i!) - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right\} \\
&\propto \prod_{i=1}^n (\theta_i \alpha \lambda y_i^{\alpha-1} \exp(-\lambda y_i^\alpha))^{\delta_i} \exp \{-\theta_i (1 - \exp(-\lambda y_i^\alpha))\}, \quad (2.37)
\end{aligned}$$

sendo  $\theta_i$  como em (2.33). O somatório em (2.37), percorre todos os possíveis valores do vetor  $\mathbf{N}$ . As estimativas de máxima verossimilhança (EMV) de  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma} = (\lambda, \alpha)^\top$ , podem ser obtidas por maximização direta da função de verossimilhança (2.37) (Rodrigues *et al.*, 2008). Também as estimativas dos parâmetros podem ser obtidas, aplicando o algoritmo EM (Mizoi, 2004) em (2.36). Entretanto, a proposta deste trabalho é considerar a abordagem bayesiana que será discutida no próximo capítulo.

Assim, neste capítulo foram apresentados os resultados e fatos relevantes da teoria unificada de análise de sobrevivência desenvolvida por Rodrigues *et al.* (2009a). É interessante observar, que todos os modelos de longa duração são unificados por esta teoria fazendo uso da função geradora de probabilidades. Além disso, um importante resultado é a caracterização da propriedade de riscos proporcionais apresentada no Teorema 2.3.

No próximo capítulo, apresentaremos o modelo de partição bayesiano, que é uma extensão desta metodologia de modelos de longa duração, em particular, do modelo de tempo de promoção.

# Capítulo 3

## Modelo de Partição Bayesiano

### 3.1 Introdução

Os modelos de partição são métodos que dividem algum domínio de interesse  $\mathcal{X} \subset \mathbb{R}^p$  em regiões disjuntas, e atribuem uma mesma distribuição de probabilidade para a variável resposta  $Y$  em cada região de  $\mathcal{X}$ .

Seja  $\mathcal{T}$  a estrutura de partição para  $\mathcal{X}$ , isto é, o domínio ou região de interesse onde é feita a partição, e  $R_m$  uma região da partição de  $\mathcal{X}$ ,  $m = 1, \dots, M$ , em que  $M$  representa o número de regiões, defina  $y_{mj}$  a  $j$ -ésima observação de  $\mathbf{Y}_m$ ,  $j = 1, \dots, n_m$ , que está na região  $R_m$ , em que  $n_m$ , é o número de pontos em  $R_m$  e  $\mathbf{Y}_m = (y_{m1}, \dots, y_{mn_m})^\top$ , e  $X_m$  é a matriz de delineamento de pontos, incluindo um termo intercepto, encontrado na  $m$ -ésima região de  $\mathcal{X}$ .

A modelagem por partição não é uma idéia nova, existem várias áreas de aplicações, tais como: geoestatística, genética, finanças, entre outras. Em estatística espacial, por exemplo, um dos problemas é estimar a incidência ou risco de uma certa doença em uma região de interesse. Fazer uma partição desta área e analisar cada sub-região, é uma alternativa na análise deste problema. Vários autores têm estudado modelos de partição bayesiano. Entre eles, podemos citar Yao (1984), Carlin *et al.* (1992), Barry & Hartigan (1993), Stephens (1994),

Green (1995) e Heikkinen (1998).

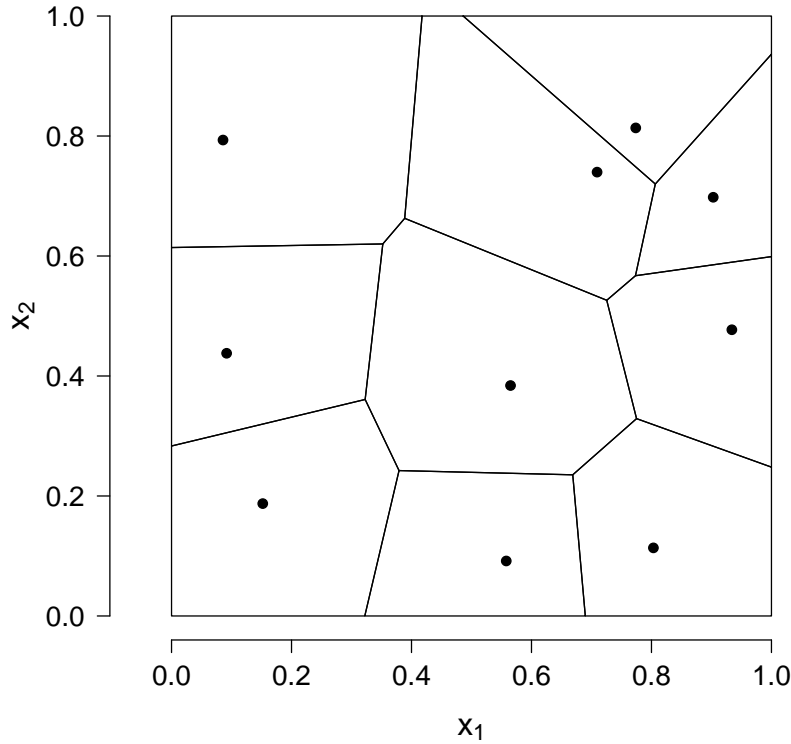
A motivação que fundamenta os modelos de partição é a de que pontos próximos no espaço delineado  $\mathcal{X}$ , têm uma mesma distribuição local, isto é, pontos em uma mesma região têm uma mesma distribuição de probabilidade. A partir desta idéia, são construídas regiões sob  $\mathcal{X}$ , de forma que as regiões são disjuntas entre si e a união delas é  $\mathcal{X}$ . Uma forma, de fazer a partição de  $\mathcal{X}$  é usar uma estrutura de tesselação (Holmes *et al.*, 1999). Uma tesselação é uma coleção de regiões chamadas células, de forma que as células são disjuntas entre si e a união delas é o plano. A construção da tesselação é feita quando a forma geométrica de cada célula é um polígono (ou uma região aberta) em que, não necessariamente, todas as células têm o mesmo número de lados (Green & Sibson, 1978).

A tesselação de Voronoi (Green & Sibson, 1978) é determinada por um número de centros  $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_M)$ , que dividem  $\mathcal{X}$  em  $M$  regiões  $R_1, \dots, R_M$ , com  $\mathcal{S}_m \in \mathbb{R}^p$ ,  $m = 1, \dots, M$ . Os elementos na região  $R_m$  são pontos mais próximos ao centro  $\mathcal{S}_m$ , que de outro centro  $\mathcal{S}_{m'}$ ,  $m \neq m'$ , isto é:

$$R_m = \{x \in \mathcal{X}; \|x - \mathcal{S}_m\| < \|x - \mathcal{S}_{m'}\|, \quad \forall m \neq m'\},$$

em que  $\|(x_1, \dots, x_p)\|^2 = \sum_{r=1}^p w_r^2 x_r^2$ , e  $\mathbf{w}$  é um vetor normal de ponderação, ou seja,  $\|\mathbf{w}\| = 1$ , sendo que  $\mathbf{w}$  indica diferentes ênfases sobre distintas direções. O caso mais usual é quando  $w_r = \frac{1}{\sqrt{p}}$ ,  $r = 1, \dots, p$ , que equivale à distância métrica euclideana. Na Figura 3.1, observamos a partição de um plano em  $\mathbb{R}^2$  feita pela tesselação de Voronoi considerando a distância métrica euclideana  $\mathbf{w} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ , sendo que  $x_1$  e  $x_2$  são variáveis contínuas em  $[0, 1]$ .

Desta forma, a estrutura de tesselação  $\mathcal{T}$  é composta pelo número de centros  $\mathcal{S}$ , número de regiões  $M$  e o vetor unitário de ponderação  $\mathbf{w}$ ,  $\mathcal{T} = (\mathcal{S}, M, \mathbf{w})$ . Além disso, as respostas em cada região de  $\mathcal{X}$  têm uma mesma distribuição de probabilidade  $f$ , indexada pelo vetor de parâmetros  $\phi_m$ ,  $m = 1, \dots, M$ , em cada região, sendo  $\phi = (\phi_1, \dots, \phi_M)$ . Para entender melhor o significado do modelo de partição neste trabalho, fazemos uso da seguinte definição, de Denison *et al.* (2002b):

Figura 3.1: Tesselação de Voronoi em  $\mathbb{R}^2$ 

**Definição 3.1.** Um modelo de partição é composto de um número de regiões disjuntas  $R_1, \dots, R_M$  cuja união é o domínio de interesse  $\mathcal{X}$ , tal que  $R_m \cap R_{m'} = \emptyset$ , para  $m \neq m'$  e  $\bigcup_{m=1}^M R_m = \mathcal{X}$ . As respostas em cada região, dado o vetor de parâmetros relativo a cada região  $\phi = (\phi_1, \dots, \phi_M)$ , são permutáveis e provêm de uma mesma classe de distribuição  $f$ .

Os modelos de partição geralmente envolvem a partição considerando apenas uma covariável, isto é, o espaço preditor  $\mathcal{X}$  tem dimensão 1, como exemplificado em Barry & Hartigan (1993) e em outros mais recentes, Loschi & Cruz (2005). Também nestes trabalhos, assume-se uma dependência entre os parâmetros de regiões próximas, conforme consta em Green & Sibson (1978) e Heikkinen (1998). Holmes *et al.* (1999) descreve os problemas que levam a assumir a dependência na modelagem por partição, que são:

1. Considerando dependência dos parâmetros entre regiões próximas, conse-

quência é a dependência entre regiões próximas em  $\mathcal{X}$ , portanto, a dependência precisa ser especificada. Isto torna o modelo mais complexo, e portanto, com mais parâmetros.

2. A função de verossimilhança marginal para a estrutura de partição, não é analiticamente tratável quando a dependência é adotada e, portanto, a forma analítica para o cálculo dela é difícil.

Neste cenário, Holmes *et al.* (1999) propõe o modelo de partição bayesiano (MPB) assumindo a independência entre os parâmetros e, portanto, a independência entre as regiões. Além disso, a partição do espaço preditor  $\mathcal{X}$  é feita considerando dimensões de  $\mathcal{X}$  maiores que 1 (Denison *et al.*, 2002b). Por outro lado, o modelo de partição bayesiano foi desenvolvido considerando variáveis preditoras contínuas para a variável de resposta. No entanto, extensões a variáveis preditoras categóricas, podem ser encontradas em Giudici *et al.* (2000).

Na próxima seção apresentamos o modelo de partição bayesiano desenvolvido por Holmes *et al.* (1999).

## 3.2 Modelo de partição bayesiano

O modelo de partição bayesiano proposto por Holmes *et al.* (1999), é uma abordagem genérica para o problema de classificação e regressão, em que o espaço preditor  $\mathcal{X}$  é dividido em regiões disjuntas definidas por uma estrutura de tesselação. A idéia intuitiva do modelo de partição bayesiano é dividir o espaço preditor  $\mathcal{X}$  em um número de regiões homogêneas. A partição do espaço preditor  $\mathcal{X}$  é feita pela tesselação de Voronoi. O modelo de partição bayesiano não faz uso de uma estrutura hierárquica como o método de classificação e regressão por árvore (Breiman *et al.*, 1984), porque isto leva a uma forte correlação entre os parâmetros do modelo. Além disso, a simulação da distribuição *a posteriori* é problemática (Denison *et al.*, 1998).

O modelo de partição bayesiano foi proposto originalmente para fazer regressão e classificação. No entanto, aplicações e extensões foram apresentadas

por Denison & Holmes (2001), Hoggart & Griffin (2001), Stephenson *et al.* (2006), Hopcroft *et al.* (2009), entre outros.

Uma das características do modelo de partição bayesiano é que além da independência dos parâmetros entre as regiões, o tratamento da função de verossimilhança marginal é disponível para qualquer estrutura de tesselação. A disponibilidade da função de verossimilhança marginal reduz em grande parte o espaço de modelos, assim como o espaço de parâmetros envolvidos na distribuição dos dados em cada partição. As amostras da distribuição *a posteriori* para a estrutura da tesselação do modelo de partição bayesiano, são geradas pelo método MCMC (Gamerman & Lopes, 2006).

Uma questão importante no modelo de partição bayesiano é a atribuição de distribuições *a priori* conjugadas para os parâmetros  $\phi_m$ ,  $m = 1, \dots, M$ , em cada região de  $\mathcal{X}$ . Isto permite a obtenção de uma expressão analítica da função de verossimilhança marginal para qualquer estrutura de tesselação, integrando todos os parâmetros extras do modelo.

Para ilustrar o modelo de partição bayesiano, consideramos o problema de classificação. Desta forma, a variável de resposta  $y_{mj}$  em  $R_m$  pode tomar só um dos  $k$  possíveis valores,  $k \in \{1, \dots, K\}$ , em que  $k$  representa o rótulo da classe  $\mathcal{C}_k$ . Então, a tarefa é atribuir a probabilidade condicional para cada classe  $\mathcal{C}_k$ ,  $p(y = k|\mathbf{x})$ , para algum valor da variável preditora  $\mathbf{x} \in \mathcal{X}$ .

O modelo de partição bayesiano, no problema de classificação, faz uso da distribuição de probabilidade multinomial para cada região de  $\mathcal{X}$ . A função de verossimilhança na região  $R_m$ ,  $m = 1, \dots, M$ , é dada por:

$$L_m(\phi_m, \mathcal{J}; \mathbf{y}_m) = \prod_{j=1}^{n_m} \left\{ \frac{n_m!}{\prod_{k=1}^K n_{mk}!} \prod_{k=1}^K \phi_{mk}^{n_{mk}} \right\}, \quad (3.1)$$

em que  $\phi_m = (\phi_{m1}, \dots, \phi_{mK})$  e  $n_{mk}$  é o número de observações em  $R_m$  que são da classe  $\mathcal{C}_k$ . Além disso,  $n_m = \sum_{k=1}^K n_{mk}$  e  $\sum_{k=1}^K \phi_{mk} = 1$ . Assim,  $\phi_m$  contém somente  $K - 1$  parâmetros.

A função de verossimilhança para o conjunto de dados  $\mathbf{D}$  posto que

conhecemos a estrutura da tesselação  $\mathcal{J}$  é dada por

$$\begin{aligned} L(\phi, \mathcal{J}; \mathbf{D}) &= \prod_{m=1}^M L_m(\phi_m, \mathcal{J}; \mathbf{y}_m) \\ &= \prod_{m=1}^M \prod_{j=1}^{n_m} \left\{ \frac{n_m!}{\prod_{k=1}^K n_{mk}!} \prod_{k=1}^K \phi_{mk}^{n_{mk}} \right\}. \end{aligned} \quad (3.2)$$

### 3.2.1 Especificação *a priori*

Considerando uma distribuição *a priori* conjugada para o vetor de parâmetros  $\phi_m$  na  $m$ -ésima região, temos que a função de verossimilhança marginal, para qualquer estrutura de partição proposta, é disponível, integrando os parâmetros do modelo, isto é:

$$L(\mathcal{J}; \mathbf{D}) = \int L(\phi, \mathcal{J}; \mathbf{D}) \pi(\phi | \mathcal{J}) d\phi. \quad (3.3)$$

A integral em (3.3) pode ser encontrada explicitamente, desde que adotemos distribuições *a priori* conjugadas para o vetor de parâmetros  $\phi$  condicionado pela estrutura de tesselação  $\mathcal{J}$ . Caso contrário isto conduz a aproximar (3.3) por método numéricos e, portanto, a simulação MCMC se torna complexa, além do custo computacional.

A forma de (3.3) sugere uma estrutura de distribuições *a priori* hierárquicas. Assim atribuímos distribuições *a priori* para  $\pi(\phi | \mathcal{J})$  e  $\pi(\mathcal{J})$ , em que

$$\pi(\phi, \mathcal{J}) = \pi(\phi | \mathcal{J}) \pi(\mathcal{J}). \quad (3.4)$$

Portanto, para o problema de classificação atribuímos distribuição *a priori* independente Dirichlet  $\text{Di}(\phi_0)$  para o vetor de parâmetros  $\phi_m$ , em que o vetor  $\phi_m = (\phi_{m1}, \dots, \phi_{mK})$  contém as probabilidades de cada classe,  $k \in \{1, \dots, K\}$ , na  $m$ -ésima região, de modo que:

$$\pi(\phi_m) = \text{Di}(\phi_m | \phi_0), \quad m = 1, \dots, M, \quad (3.5)$$

e portanto,

$$\pi(\phi | \mathcal{J}) = \prod_{m=1}^M \text{Di}(\phi_m | \phi_0), \quad (3.6)$$



em que  $\boldsymbol{\phi}_0 = (\phi_{01}, \dots, \phi_{0K})^\top$  como vetor de hiperparâmetros,  $\phi_{0k} > 0$ ,  $k = 1, \dots, K$ .

Desta forma, a função de verossimilhança marginal na região  $R_m$ , é dada por:

$$\begin{aligned} L_m(\mathcal{J}; \mathbf{y}_m) &= \int L_m(\phi_m, \mathcal{J}; \mathbf{y}_m) \pi(\phi_m) d\phi_m \\ &= \frac{\Gamma(\sum_{k=1}^K \phi_{0k}) \prod_{k=1}^K \Gamma(n_{mk} + \phi_{0k})}{\prod_{k=1}^K \Gamma(\phi_{0k}) \Gamma(n_m + \sum_{k=1}^K \phi_{0k})}. \end{aligned} \quad (3.7)$$

A função de verossimilhança marginal para a estrutura da tesselação, considerando o conjunto de dados  $\mathbf{D}$ , é:

$$\begin{aligned} L(\mathcal{J}; \mathbf{D}) &= \prod_{m=1}^M L_m(\mathcal{J}; \mathbf{y}_m) \\ &= \left\{ \frac{\Gamma(\sum_{k=1}^K \phi_{0k})}{\prod_{k=1}^K \Gamma(\phi_{0k})} \right\}^M \prod_{m=1}^M \frac{\prod_{k=1}^K \Gamma(n_{mk} + \phi_{0k})}{\Gamma(n_m + \sum_{k=1}^K \phi_{0k})}. \end{aligned} \quad (3.8)$$

Por outro lado, a distribuição *a priori* para para o vetor  $\mathcal{J} = (\mathbf{w}, \mathcal{S}, M)$  que define a estrutura da tesselação é

$$\begin{aligned} \pi(\mathcal{J}) &= \pi(\mathbf{w})\pi(\mathcal{S}, M) \\ &= \pi(\mathbf{w})\pi(\mathcal{S}|M)\pi(M). \end{aligned} \quad (3.9)$$

Adotamos distribuições uniformes próprias, desta forma, para  $\mathbf{w}$  (o vetor normal de ponderação) a distribuição *a priori* é a distribuição de Dirichlet com hiperparâmetros  $w_0 = (w_{01}, \dots, w_{0p})$ ,  $w_{0r} > 0$   $r = 1, \dots, p$ ,

$$\pi(\mathbf{w}) = \text{Di}(\mathbf{w}|w_0). \quad (3.10)$$

Para o número de regiões e, portanto, para o número de centros que definem a tesselação de Voronoi, consideramos a distribuição uniforme em  $\{1, \dots, M_{max}\}$ , em que  $M_{max}$  é o número máximo de regiões escolhido pelo pesquisador. Assim,

$$\pi(M) = \text{U}(M|1, \dots, M_{max}). \quad (3.11)$$

Condicionado a  $M$  regiões, escolhamos centros para a construção das regiões  $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_M)$ , em que  $\mathcal{S}_m = (\mathcal{S}_{m1}, \dots, \mathcal{S}_{mp})$ ,  $m = 1, \dots, M$ .

A distribuição *a priori* para cada coordenada de  $\mathcal{S}_m$  é a distribuição marginal uniforme sobre os valores de cada variável preditora, isto é:

$$\pi(\mathcal{S}_{mr}) = U(\mathcal{S}_{mr} | x_{1r}, \dots, x_{nr}), \quad r = 1, \dots, p, \quad (3.12)$$

e portanto,

$$\pi(\mathcal{S}_m) = \prod_{r=1}^p U(\mathcal{S}_m | x_{1r}, \dots, x_{nr}), \quad m = 1, \dots, M. \quad (3.13)$$

Assim, a distribuição *a priori* para a estrutura da tesselação  $\mathcal{J}$ , é dada por:

$$\begin{aligned} \pi(\mathcal{J}) &= \pi(\mathbf{w})\pi(M)\pi(\mathcal{S}|M) \\ &= \text{Di}(\mathbf{w}|w_0)U(M|1, \dots, M_{max}) \prod_{m=1}^M \prod_{r=1}^p U(\mathcal{S}_{mr} | x_{1r}, \dots, x_{nr}). \end{aligned} \quad (3.14)$$

### 3.2.2 Análise *a posteriori* e estratégia computacional

A distribuição *a posteriori* para a tesselação  $\mathcal{J}$ , é dada por:

$$\pi(\mathcal{J}|\mathbf{D}) \propto L(\mathcal{J}; \mathbf{D})\pi(\mathcal{J}). \quad (3.15)$$

Notemos que em (3.15), a distribuição *a posteriori* de  $\mathcal{J}$  não é analiticamente tratável. Logo, o modelo de partição bayesiano faz uso do método de simulação MCMC para aproximar a distribuição *a posteriori* para  $\mathcal{J}$ .

Como abordamos o problema de classificação e consideramos  $K$  classes, o modelo de partição bayesiano começa com  $K$  regiões para a partição de  $\mathcal{X}$ . Desta forma, a estratégia computacional para a simulação MCMC proposta por Holmes *et al.* (1999), tem os seguintes movimentos:

1. Construir uma nova região após a escolha aleatória de um novo centro retirado uniformemente dos valores preditos.
2. Remover uma região excluindo um centro.

3. Mover um centro, após selecionar sua posição aleatoriamente dos preditores marginais.
4. Alterar o vetor de ponderação  $\mathbf{w}$  usando uma densidade proposta normal durante a fase de aquecimento, para obter uma taxa de aceitação em torno de 30%.

A mudança para uma nova estrutura de tesselação  $\mathcal{T}'$  sobre a tesselação  $\mathcal{T}$ , é aceita com probabilidade:

$$\alpha = \min \left\{ 1, \frac{L(\mathcal{T}'; \mathbf{D})}{L(\mathcal{T}; \mathbf{D})} \right\}. \quad (3.16)$$

Notemos que a fração (3.16) é o fator de Bayes em favor da nova estrutura de tesselação  $\mathcal{T}'$  sobre  $\mathcal{T}$ .

Durante a simulação MCMC, descartamos uma porção inicial das amostras geradas para diminuir a influência dos valores iniciais.

Para verificar a convergência das cadeias na simulação MCMC, Holmes *et al.* (1999) sugere que seja examinadas as estatísticas das amostras, como por exemplo o número de regiões da tesselação e a esperança  $E(y|\mathbf{x}, \mathcal{T}, \phi)$  de forma que se elas começam a estabilizar, então consideramos as amostras geradas pela simulação MCMC.

### 3.2.3 Predição

A principal característica deste modelo é sua performance preditiva, tanto em problemas de regressão e classificação, como em outros problemas, conforme estudos de Hoggart & Griffin (2001), Stephenson *et al.* (2006), Leonte & Nott. (2006), Hopcroft *et al.* (2009), entre outros.

As amostras geradas pela simulação MCMC  $\{\mathcal{T}^1, \dots, \mathcal{T}^B\}$ , são usadas para aproximar a distribuição preditiva *a posteriori*  $\pi(y|\mathbf{x}, \mathbf{D})$ :

$$\pi(y|\mathbf{x}, \mathbf{D}) \approx \frac{1}{B} \sum_{b=1}^B \pi(y|\mathcal{T}^b, \mathbf{D}, \mathbf{x}). \quad (3.17)$$

No problema de classificação, a distribuição de probabilidade condicional preditiva para a variável resposta  $y$ , dado um vetor de covariáveis  $\mathbf{x} \in R_m \subset \mathcal{X}$  (Denison *et al.*, 2002a), é dada por:

$$\pi(y = k | \mathbf{x} \in R_m, \mathcal{J}, \mathbf{D}) = \frac{n_{mk} + 1}{n_m + K}, \quad m = 1, \dots, M \text{ e } k = 1, \dots, K. \quad (3.18)$$

Assim, para efetuar a partição de  $\mathcal{X}$ , fazemos uso da tesselação de Voronoi. Usualmente é adotada a distância métrica euclideana para a construção das regiões, no entanto, pode-se adotar outras distâncias métricas (Denison *et al.*, 2002b).

### 3.2.4 Tesselação por hiperplanos ortogonais

Uma desvantagem ocorre quando fazemos uso da tesselação de Voronoi, caso haja uma grande quantidade de dados e muitas variáveis preditoras. Denison *et al.* (2002b) comprovou que quando  $p \geq 5$ , independentemente da distância métrica adotada, o custo computacional é grande; como exemplo, observe-se em Hopcroft *et al.* (2009) que a simulação MCMC demorou duas semanas.

Para fazer a partição de  $X$ , não necessariamente devemos considerar a tesselação de Voronoi para a construção das regiões no modelo de partição bayesiano. Holmes *et al.* (1999), por exemplo, propõe fazer a partição de  $\mathcal{X}$  por hiperplanos; no entanto, independentemente da estrutura de tesselação  $\mathcal{J}$  adotada, o fundamental é que a função de verossimilhança marginal seja encontrada analiticamente para qualquer estrutura de tesselação.

Uma alternativa é trabalhar com hiperplanos, em particular a tesselação por hiperplanos ortogonais usado por Hoggart & Griffin (2001). A idéia, é construir hiperplanos paralelos aos eixos que determinam hiperplanos ortogonais e, em consequência, obter regiões disjuntas e, portanto, uma partição de  $\mathcal{X}$ .

Para ilustrar esta idéia, observamos a Figura 3.2, na qual apresentamos a tesselação por hiperplanos ortogonais como resultado da interseção de hiperplanos paralelos.

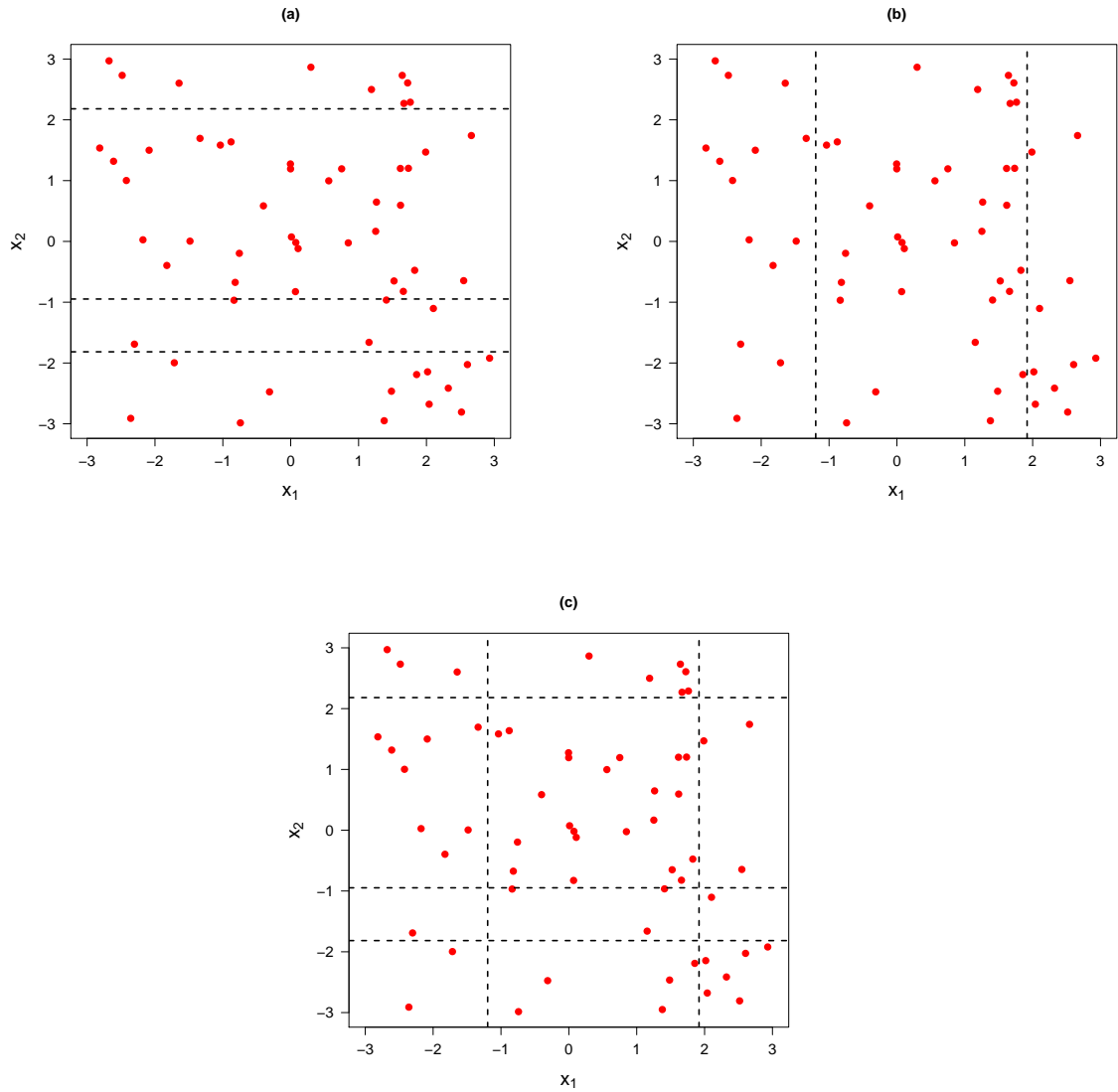


Figura 3.2: (a) Planos paralelos ao eixo  $x_1$ , (b) Planos paralelos ao eixo  $x_2$  e (c) Hiperplanos ortogonais aos eixos  $x_1$  e  $x_2$ .

Uma das vantagens da tesselação usada por Hoggart & Griffin (2001), além da rapidez computacional, é a seleção de variáveis preditoras, isto é, os hiperplanos paralelos dividem ou fazem um corte nos eixos das variáveis preditoras que têm realmente um efeito significativo sobre a variável resposta  $Y$ .

Como foi dito anteriormente, seja  $\mathcal{T}$  a estrutura de tesselação; a tesselação neste caso, é composta pelos hiperplanos paralelos aos eixos que denotamos por  $h$ , e  $M$  é o número de regiões determinadas pelos hiperplanos ortogonais. Assim,  $\mathcal{T} = (h, M)$ . Notemos que nesta estrutura de tesselação, não precisamos do  $w$  (o vetor unitário de ponderação).

Neste caso, também as variáveis preditoras são contínuas, mas no presente trabalho, consideramos uma extensão para variáveis categóricas baseada no artigo de Giudici *et al.* (2000).

Assim, no modelo de partição bayesiano, a tesselação de Voronoi é substituída pela tesselação por hiperplanos ortogonais (Hoggart & Griffin, 2001), porém, a metodologia do modelo de partição bayesiano é a mesma.

Para esta nova estrutura de tesselação  $\mathcal{T}$ , a distribuição *a priori* para os hiperplanos paralelos aos eixos  $\mathbf{h} = (h_1, \dots, h_p)$ , consideramos distribuições de probabilidade uniforme. Assim, a distribuição *a priori* para o hiperplano  $h_r$ ,  $r = 1, \dots, p$ , é a distribuição marginal uniforme sobre os valores preditores  $\{x_{1r}, \dots, x_{nr}\}$ ,  $h_r \sim U(x_{1r}, \dots, x_{nr})$  e, portanto, a distribuição *a priori* para  $h$  é dada por:

$$\pi(\mathbf{h}) = \prod_{r=1}^p U(h_r | x_{1r}, \dots, x_{nr}). \quad (3.19)$$

Para  $M$  (o número de regiões da tesselação), assumimos a distribuição *a priori* geométrica  $M \sim \text{Geo}(\psi)$ ,  $0 < \psi < 1$  assim:

$$\pi(M) = \text{Geo}(M | \psi), \quad 0 < \psi < 1. \quad (3.20)$$

Assim sendo, a distribuição *a posteriori* para  $\mathcal{T}$  é dada como em (3.15), portanto, a distribuição *a posteriori* para  $\mathcal{T}$  não tem uma forma analítica. Então fazemos uso de métodos computacionais de simulação, como o método MCMC, para aproximar a distribuição *a posteriori* de  $\mathcal{T}$ .

Neste cenário, tal como foi dada a estratégia computacional para a tesselação de Voronoi, neste caso, para a tesselação por hiperplanos ortogonais, Hoggart & Griffin (2001) propõem a seguinte estratégia computacional:

### Estratégia computacional

Denotamos por  $\mathcal{J}$ , o conjunto de índices das variáveis preditoras  $\mathcal{J} = \{1, \dots, p\}$ , e sendo  $\mathcal{J}_{\mathcal{T}}$  o conjunto de índices das variáveis preditoras presentes na tesselação  $\mathcal{T}$ , assim se propõem-se três movimentos possíveis, cada um com igual probabilidade de ser escolhido.

- Uma nova partição pode ser adicionada na estrutura da tesselação  $\mathcal{J}$ , escolhendo-se um novo ponto de corte de uma variável preditora em  $\mathcal{J}$ . O ponto de corte é selecionado da distribuição *a priori* para o hiperplano  $h_r \sim U(x_{1r}, \dots, x_{nr})$ .
- Um hiperplano pode ser eliminado, escolhendo-se aleatoriamente uma variável preditora  $r^*$  presente na tesselação,  $r^* \in \mathcal{J}$ .
- Um hiperplano pode ser mudado, selecionando-se um novo ponto de corte da distribuição *a priori* para esse hiperplano,  $h_{r^*}, r^* \in \mathcal{J}$ .

A nova tesselação  $\mathcal{J}'$  proposta, é aceita com probabilidade:

$$\alpha(\mathcal{J}', \mathcal{J}) = \min \left\{ 1, \frac{L(\mathcal{J}'; \mathbf{D})}{L(\mathcal{J}; \mathbf{D})} \right\}. \quad (3.21)$$

A fração em (3.21), é o fator de bayes (Kass & Raftery, 1995) em favor da nova estrutura da tesselação  $\mathcal{J}'$ .

Na próxima seção, apresentamos a extensão local do modelo de partição bayesiano para o modelo de tempo de promoção. A metodologia da modelagem de partição bayesiana é flexibilizada, desde que seja adotada a tesselação por hiperplanos ortogonais quando se trabalha com uma grande quantidade de dados. Esta alternativa de modelagem, foi desenvolvida por Hoggart & Griffin (2001).

### 3.3 Extensão local para modelos de longa duração

Nesta seção, apresentamos a extensão local para os modelos de longa duração proposta por Hoggart & Griffin (2001). Em tal extensão, os efeitos das variáveis predictoras em cada região de  $\mathcal{X}$ , são considerados fazendo uso do modelo de partição bayesiano. Desse modo, a teoria unificada de análise de sobrevivência para modelos de longa duração, é aplicada em cada região do espaço preditor  $\mathcal{X}$ .

Para a construção das regiões de  $\mathcal{X}$ , adotamos a tesselação por hiperplanos ortogonais descrita na Subseção 3.2.3, em que uma das características principais, é a seleção das variáveis predictoras que têm efeito significativo na variável resposta  $Y$ .

### 3.3.1 Função de verossimilhança

A tesselação por hiperplanos ortogonais  $\mathcal{T}$ , define  $M$  regiões  $R_1, \dots, R_M$ . Seja assim  $n_m$  o número de observações na região  $R_m$ .

No cenário de riscos competitivos, seja  $N_{mj}$  (não observável) o número de causas do evento de interesse da  $j$ -ésima observação, com distribuição de probabilidade  $p(N_{mj}|\phi_m)$ ,  $j = 1, \dots, n_m$ , indexado pelo vetor de parâmetro  $\phi_m$  na região  $R_m$ .

Dado  $N_{mj}$ , sejam  $Z_{mj}^1, \dots, Z_{mj}^{N_{mj}}$  os tempos de ocorrência do evento de interesse para  $N_{mj}$  com função de distribuição acumulada  $F(\cdot|\gamma) = 1 - S(\cdot|\gamma)$ , onde  $\gamma$  é o vetor de parâmetros da distribuição.

Seja  $T_{mj}$  como em (2.1) e  $C_{mj}$  o tempo da censura. O tempo observado, é dado por:  $Y_{mj} = \min\{T_{mj}, C_{mj}\}$ . Seja  $\delta_{mj}$  a variável indicadora de censura com  $\delta_{mj} = 1$  se  $Y_{mj} = T_{mj}$ , e  $\delta_{mj} = 0$  caso contrário.

A função de verossimilhança na região  $R_m$ , considerando censura não informativa, é dada por:

$$L_m(\gamma, \phi_m, N_{mj}; \mathbf{y}, \boldsymbol{\delta}) \propto \prod_{j=1}^{n_m} \{S(y_{mj}|\gamma)\}^{N_{mj}-\delta_{mj}} \{N_{mj}f(y_{mj}|\gamma)\}^{\delta_{mj}} p(N_{mj}|\phi_m). \quad (3.22)$$

Dado que conhecemos a estrutura da tesselação  $\mathcal{T}$ , a função de verossimilhança para todos os dados, é dada por:

$$L(\gamma, \mathbf{N}, \mathcal{T}, \boldsymbol{\phi}; \mathbf{y}, \boldsymbol{\delta}) \propto \prod_{m=1}^M \prod_{j=1}^{n_m} \{S(y_{mj}|\gamma)\}^{N_{mj}-\delta_{mj}} \{N_{mj}f(y_{mj}|\gamma)\}^{\delta_{mj}} p(N_{mj}|\phi_m), \quad (3.23)$$

onde  $\mathbf{N} = (N_1, \dots, N_n)^\top$  é um vetor de variáveis latentes,  $\mathbf{y} = (y_1, \dots, y_n)^\top$ ,  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_M)^\top$  e  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$ .

Neste trabalho, consideramos que o número de causas do evento de interesse  $N_{mj}$  tem a distribuição de Poisson com parâmetros  $\phi_m$ ,  $N_{mj} \sim \text{Poisson}(\phi_m)$ , e para os tempos de ocorrência  $Z_{mj}$ , dado  $N_{mj}$ , consideramos a distribuição de Weibull com parâmetros  $\gamma = (\alpha, \lambda)$ ,  $Z_{mj} \sim \text{Weibull}(\alpha, \lambda)$ ,  $j = 1, \dots, n_m$ ,



$m = 1, \dots, M$ , tal como foi proposto no trabalho de Hoggart & Griffin (2001). Então, a função de verossimilhança, é dada por:

$$L(\boldsymbol{\gamma}, \mathbf{N}, \boldsymbol{\mathcal{T}}, \boldsymbol{\phi}; \mathbf{y}, \boldsymbol{\delta}) \propto \prod_{m=1}^M \exp\left(-\lambda \sum_{j=1}^{n_m} y_{mj}^\alpha N_{mj}\right) \prod_{j=1}^{n_m} (N_{mj} \alpha \lambda y_{mj}^{\alpha-1})^{\delta_{mj}} p(N_{mj} | \phi_m), \quad (3.24)$$

em que o número de causas do evento de interesse tem a distribuição de Poisson, portanto:

$$p(\mathbf{N} | \boldsymbol{\phi}, \boldsymbol{\mathcal{T}}) = \prod_{m=1}^M \prod_{j=1}^{n_m} \frac{\exp(-\phi_m) \phi_m^{N_{mj}}}{N_{mj}!}. \quad (3.25)$$

No artigo de Hoggart & Griffin (2001), página 226 foi definida de forma incorreta a distribuição de Weibull; no entanto, os resultados nesse artigo foram coerentes. Neste trabalho, será assumida uma nova parametrização para o parâmetro  $\lambda$ , da distribuição de Weibull.

### 3.3.2 Especificação *a priori* e verossimilhança marginal

Assumindo independência para os parâmetros da distribuição de Weibull  $\boldsymbol{\gamma} = (\alpha, \lambda)$ , consideramos para o parâmetro  $\alpha$  a distribuição Gama, ou seja,  $\alpha \sim \text{Gama}(\alpha_0, \alpha_1)$ , em que  $\alpha_0, \alpha_1 > 0$ , e para o parâmetro  $\lambda$  adotamos uma parametrização de forma que  $\lambda = \exp(\lambda^*)$ ,  $\lambda^* \in \mathbb{R}$ . Levando em conta esta parametrização, assumimos a distribuição normal com media  $\mu$  e variância  $\sigma^2$ , como distribuição *a priori* para  $\lambda^*$ ,  $\lambda^* \sim N(\mu, \sigma^2)$ .

Lembremos que, na metodologia de partição bayesiana, é importante atribuir distribuições *a priori* conjugadas para os parâmetros locais, de modo a obter a verossimilhança marginal de forma explícita e não aproximada por métodos numéricos ou de simulação, devido ao alto custo computacional.

Neste contexto, a distribuição *a priori* conjugada para  $\phi_m$ , é a distribuição gama  $\phi_m \sim \text{Gama}(\vartheta_0, \vartheta_1)$ ,  $\vartheta_0, \vartheta_1 > 0$ ,  $m = 1, \dots, M$ . Distribuição gama para o parâmetro  $\alpha$ , distribuição normal para o parâmetros  $\lambda^*$  e distribuição geométrica para o número de regiões na estrutura de tesselação  $M$ . Para o parâmetro da tesselação  $\boldsymbol{\mathcal{T}} = (\mathbf{h}, M)$ , a distribuição *a priori* foi dada na seção

anterior em (3.20). Resumindo, as distribuições *a priori* para os parâmetros, são:

$$\begin{aligned}\pi(\alpha) &= \text{Gama}(\alpha|\alpha_0, \alpha_1), \\ \pi(\lambda^*) &= \text{N}(\lambda^*|\mu, \sigma^2), \\ \pi(\phi_m) &= \text{Gama}(\phi_m|\vartheta_0, \vartheta_1), \\ &\text{e} \\ \pi(M) &= \text{Geo}(M|\psi).\end{aligned}$$

Dada a estrutura da tesselação  $\mathcal{T}$  por hiperplanos ortogonais com  $M$  regiões, a distribuição *a priori* para os parâmetros em cada região  $\phi_m$ , é dada por:

$$\pi(\phi|\mathcal{T}) = \prod_{m=1}^M \text{Gama}(\phi_m|\vartheta_0, \vartheta_1), \quad (3.26)$$

com  $\phi = (\phi_1, \dots, \phi_M)^\top$ .

Portanto, a função de verossimilhança marginal em cada região  $R_m$ , é dada por:

$$\begin{aligned}p(N_{m1}, \dots, N_{mn_m}|\vartheta_0, \vartheta_1) &= \int \prod_{j=1}^{n_m} p(N_{mj}|\phi_m) \pi(\phi_m) d\phi_m \\ &= \int \prod_{j=1}^{n_m} \frac{\exp(-\phi_m) \phi_m^{N_{mj}}}{N_{mj}!} \frac{\vartheta_1^{\vartheta_0}}{\Gamma(\vartheta_1)} \phi_m^{\vartheta_0-1} \exp(-\phi_m \vartheta_1) d\phi_m \\ &= \frac{1}{\prod_{j=1}^{n_m} N_{mj}!} \frac{\vartheta_1^{\vartheta_0}}{\Gamma(\vartheta_1)} \int \phi_m^{\sum_{j=1}^{n_m} N_{mj} - \vartheta_0} \exp(-\phi_m(n_m + \vartheta_1)) d\phi_m \\ &= \frac{1}{\prod_{j=1}^{n_m} N_{mj}!} \frac{\vartheta_1^{\vartheta_0}}{\Gamma(\vartheta_1)} \frac{\Gamma(\sum_{j=1}^{n_m} N_{mj} + \vartheta_0)}{(n_m + \vartheta_1)^{\sum_{j=1}^{n_m} N_{mj} + \vartheta_0}}.\end{aligned} \quad (3.27)$$

Portanto, a função de verossimilhança para a estrutura de tesselação  $\mathcal{T} = (\mathbf{h}, M)$ , é dada por:

$$P(N|\mathcal{T}) = \prod_{m=1}^M \frac{1}{\prod_{j=1}^{n_m} N_{mj}!} \frac{\vartheta_1^{\vartheta_0}}{\Gamma(\vartheta_1)} \frac{\Gamma(\sum_{j=1}^{n_m} N_{mj} + \vartheta_0)}{(n_m + \vartheta_1)^{\sum_{j=1}^{n_m} N_{mj} + \vartheta_0}}. \quad (3.28)$$

### 3.3.3 Análise a posteriori

Dada a estrutura da tesselação  $\mathcal{T}$ , as distribuições condicionais *a posteriori* para os parâmetros da distribuição Weibull  $\gamma = (\alpha, \lambda^*)$ , são, respectivamente:

$$\begin{aligned} \pi(\lambda^* | \alpha, \mathbf{N}, \mathcal{T}, \mathbf{y}, \boldsymbol{\delta}) &\propto \prod_{m=1}^M \exp\left(-\lambda \sum_{j=1}^{n_m} y_{mj}^\alpha N_{mj}\right) \prod_{j=1}^{n_m} (e^{\lambda^*})^{\delta_{mj}} \pi(\lambda^*) \\ &\propto e^{d\lambda^*} \exp\left(-e^{\lambda^*} \sum_{i=1}^n N_i y_i^\alpha\right) \exp\left(\frac{(\lambda^* - \mu)^2}{2\sigma^2}\right) \end{aligned} \quad (3.29)$$

e

$$\pi(\alpha | \lambda^*, \mathbf{N}, \mathcal{T}, \mathbf{y}, \boldsymbol{\delta}) \propto \alpha^d \left(\prod_{i=1}^n y_i^{\delta_i}\right)^\alpha \exp\left(-e^{\lambda^*} \sum_{i=1}^n N_i y_i^\alpha\right) \pi(\alpha), \quad (3.30)$$

em que  $d = \sum_{i=1}^n \delta_i$ .

A distribuição condicional para o número de causas do evento de interesse, depende da censura para cada observação. Assim, temos:

$$\begin{aligned} p(N_{mj} | \gamma, \phi, \mathcal{T}, \mathbf{y}, \boldsymbol{\delta}) &\propto \exp\left(-e^{\lambda^*} N_{mj} y_{mj}^\alpha\right) N_{mj}^{\delta_{mj}} p(N_{mj} | \phi_m) \\ &\propto \exp\left(-e^{\lambda^*} N_{mj} y_{mj}^\alpha\right) N_{mj}^{\delta_{mj}} \frac{\phi_m^{N_{mj}}}{N_{mj}!}. \end{aligned} \quad (3.31)$$

Se  $\delta_{mj} = 0$ , então:

$$\begin{aligned} p(N_{mj} | \gamma, \phi, \mathcal{T}, \mathbf{y}, \boldsymbol{\delta}) &\propto \exp\left(-e^{\lambda^*} N_{mj} y_{mj}^\alpha\right) \frac{\phi_m^{N_{mj}}}{N_{mj}!} \\ &\propto \frac{\{\phi_m \exp(-e^{\lambda^*} y_{mj}^\alpha)\}^{N_{mj}}}{N_{mj}!} \\ &= \text{Poisson}(N_{mj} | \phi_m \exp(-e^{\lambda^*} y_{mj}^\alpha)). \end{aligned} \quad (3.32)$$

Se  $\delta_{mj} = 1$ , então:

$$\begin{aligned} p(N_{mj} | \gamma, \phi, \mathcal{T}, \mathbf{y}, \boldsymbol{\delta}) &\propto \exp\left(-e^{\lambda^*} N_{mj} y_{mj}^\alpha\right) N_{mj} \frac{\phi_m^{N_{mj}}}{N_{mj}!} \\ &\propto \frac{\{\phi_m \exp(-e^{\lambda^*} y_{mj}^\alpha)\}^{N_{mj}-1}}{(N_{mj}-1)!} \\ &= \text{Poisson}(N_{mj}-1 | \phi_m \exp(-e^{\lambda^*} y_{mj}^\alpha)). \end{aligned} \quad (3.33)$$

A distribuição *a posteriori* para a estrutura da tesselação  $\mathcal{J}$  e para os parâmetros locais  $\phi$ , é dada por:

$$\begin{aligned}\pi(\mathcal{J}, \phi | \mathbf{N}, \mathbf{X}) &\propto p(\mathbf{N} | \phi, \mathcal{J}) \pi(\phi, \mathcal{J}) \\ &\propto p(\mathbf{N} | \phi, \mathcal{J}) \pi(\phi | \mathcal{J}) \pi(\mathcal{J}).\end{aligned}\quad (3.34)$$

Notemos que em (3.34) temos o vetor de parâmetros  $\phi$ . Logo, para obter a distribuição *a posteriori* para a estrutura da tesselação  $\mathcal{J}$ , precisamos integrar sobre o vetor de parâmetros  $\phi$ :

$$\pi(\mathcal{J} | \mathbf{N}, \mathbf{X}) \propto \int p(\mathbf{N} | \phi, \mathcal{J}) \pi(\phi | \mathcal{J}) \pi(\mathcal{J}) d\phi = p(\mathbf{N} | \mathcal{J}) \pi(\mathcal{J}).\quad (3.35)$$

Assim, temos que em (3.35), a função de verossimilhança  $p(\mathbf{N} | \mathcal{J})$  está presente na distribuição *a posteriori* de  $\mathcal{J}$ .

Portanto a distribuição condicional *a posteriori* para o parâmetro  $\phi_m$  da distribuição de Poisson, dada a estrutura da tesselação  $\mathcal{J}$ , é dada por:

$$\begin{aligned}\pi(\phi_m | \mathcal{J}, \mathbf{N}) &\propto \prod_{j=1}^{n_m} p(N_{mj} | \phi_m) \pi(\phi_m) \\ &\propto \exp(-n_m \phi_m) \phi_m^{\sum_{j=1}^{n_m} N_{mj}} \phi_m^{\vartheta_0 - 1} \exp(-\phi_m \vartheta_1).\end{aligned}$$

Reescrevendo a última expressão, obtemos:

$$\begin{aligned}\pi(\phi_m | \mathcal{J}, \mathbf{N}) &\propto \phi_m^{\sum_{j=1}^{n_m} N_{mj} + \vartheta_0 - 1} \exp(-\phi_m(\vartheta_1 + n_m)) \\ &= \text{Gama}(\phi_m | \sum_{j=1}^{n_m} N_{mj} + \vartheta_0, \vartheta_1 + n_m).\end{aligned}\quad (3.36)$$

A estratégia computacional para aproximar a distribuição *a posteriori* de  $\mathcal{J}$ , foi dada na Seção 3.2.3. É importante notar, que todos os parâmetros do modelo, dependem da estrutura da tesselação  $\mathcal{J}$ .

Observamos anteriormente, que no artigo de Hoggart & Griffin (2001), os parâmetros da distribuição condicional *a posteriori* (dada em (3.36)) estão invertidos, isto é, em ordem diferente. Em função disso, a distribuição condicional *a posteriori* demonstrada matematicamente por nós, está em (3.36).

### 3.3.4 Predição

Aplicando o método MCMC, obtemos uma amostra para aproximar a distribuição *a posteriori* de  $\mathcal{T}$  e, portanto, dos parâmetros do modelo.

Uma vez que consideramos o modelo de tempo de promoção, a fração de cura, é dada por:  $p_0 = \exp(-\theta)$ ,  $\theta > 0$ . No entanto, nesta metodologia de partição bayesiana para o modelo de tempo de promoção, a fração de cura é calculada de forma local.

Assim, dada uma observação  $(y', \delta', \mathbf{x}')$  em que  $y'$  é o tempo observado,  $\delta'$  a variável indicadora de censura associada a  $y'$ , e  $\mathbf{x}' = (x'_1, \dots, x'_p)^\top$  o vetor de variáveis preditoras, se  $\mathbf{x}' \in R_m \subset \mathcal{X}$ , dada a estrutura de tesselação  $\mathcal{T}$ , a fração de cura é dada por:

$$p_0(y', \delta' | \mathbf{x}' \in R_m, \mathcal{T}) = \exp(-\phi_m). \quad (3.37)$$

Se considerarmos uma amostra de tesselações de tamanho  $B$ , a fração de cura é dada pela média ponderada de cada fração de cura para diferentes estruturas de tesselação, considerando que cada tesselação têm o mesmo peso, assim

$$p_0(y', \delta' | \mathbf{x}', \mathcal{T}) = \frac{1}{B} \sum_{b=1}^B p_0(y', \delta' | \mathbf{x}', \mathcal{T}^b). \quad (3.38)$$

No próximo capítulo, apresentamos aplicações desta metodologia a dados simulados e dados reais.

# Capítulo 4

## Aplicações

Neste capítulo, ilustramos a teoria descrita no Capítulo 3 através de um estudo de simulação e de um exemplo com dados reais. O estudo de simulação, foi desenvolvido com base em dados simulados, segundo o modelo exponencial. Para o exemplo de dados reais, utilizamos o conjunto de dados estudados por Kirkwood *et al.* (2000), em que consideramos o modelo Weibull para os tempos de ocorrência do evento de interesse. Toda a programação utilizada neste trabalho, foi desenvolvida em linguagem de programação R (R Development Core Team, 2009).

### 4.1 Estudo de simulação

Para o estudo de simulação, geramos uma amostra de tamanho  $n = 200$ , em que a distribuição dos tempos de ocorrência do evento de interesse ( $Z$  na Seção 2.1.1) seguem uma distribuição exponencial com parâmetro  $\lambda = \exp(-2)$ . Neste estudo, consideramos uma variável preditora  $x_1$ , gerada da distribuição normal com média 0,  $\mu^* = 0$ , variância  $\sigma_*^2 = 1$  e coeficientes de regressão  $\beta = (0, 02, -0, 8)^\top$ .

Do modelo de tempo de promoção, temos que o parâmetro  $\theta$  da distribuição de Poisson está relacionado com as variáveis preditoras, pela função de ligação  $\theta_i = \exp(\mathbf{x}_i^\top \beta)$ .

A fração de cura  $p_0$ , é dada por:  $p_0 = \exp(-\exp(\mathbf{x}^\top \boldsymbol{\beta}))$ , e a proporção de observações censuradas, é igual a  $p_c = p_0 + 0,07$ . Assim, neste estudo de simulação para este conjunto de dados, a fração de cura média é  $p_0 = 0,358$ . Os tempos observados e a variável indicadora de censura são gerados, como segue abaixo:

1. Gerar  $u_i \sim U(0, 1)$ .
2. Se  $u_i < p_{0i}$ ,  $y_i^* = +\infty$ , caso contrário,
 
$$y_i^* = F^{-1}\left(\frac{u_i - 1}{\lambda \exp(x_i^\top \boldsymbol{\beta})}\right).$$
 Com  $F^{-1}(t^*) = \frac{-\log(1-t^*)}{\lambda}$
3. Gerar  $c_i \sim \text{Exp}\left(\frac{\lambda(p_{ci} - p_{0i})}{1 - (p_{ci} - p_{0i})}\right)$ .
4.  $y_i = \min\{y_i^*, c_i\}$ .
5. Se  $y_i < c_i$  fazer  $\delta_i = 1$ , caso contrário  $\delta_i = 0$ ,  $i = 1, \dots, n$

Considerando que a estrutura de tesselação por hiperplanos ortogonais seleciona as variáveis preditoras que realmente têm efeito sobre a variável resposta, geramos a variável  $x_2 \sim \text{Gama}(4, 7)$ , que neste estudo de simulação não é significativa, pois geramos os tempos observados e a variável indicadora de censura, levando em conta a variável preditora  $x_1$ .

O objetivo é estimar os parâmetros da regressão  $\boldsymbol{\beta}$  e os parâmetros da distribuição do tempo de sobrevivência  $\lambda$ , como por exemplo, em Chen *et al.* (1999) e Rodrigues *et al.* (2008). No entanto, nesta metodologia proposta por Hoggart & Griffin (2001), em vez de estimar os coeficientes de regressão  $\boldsymbol{\beta}$ , o algoritmo de simulação MCMC gera estruturas de tesselação  $\mathcal{T}$  e, condicionado a  $\mathcal{T}$ , gera-se o parâmetro da distribuição de Poisson  $\phi$  para cada tesselação. Daí estima-se a fração de cura em cada região da tesselação, além de estimar-se os parâmetros da distribuição do tempo de sobrevivência  $\lambda$ .

Atribuímos distribuições *a priori* para os parâmetros do modelo como na Seção 3.3.2. A distribuição *a priori* para o parâmetro  $\lambda^*$  é a distribuição normal  $N(0, 100)$ . Para o parâmetro  $\phi_m$  da distribuição de Poisson em cada região

de  $\mathcal{X}$ , assumimos a distribuição de probabilidade gama com hiperparâmetros  $\vartheta_0 = \vartheta_1 = 0, 1$ ,  $\phi_m \sim \text{Gama}(0, 1, 0, 1)$  e para o número de regiões  $M$ , assumimos a distribuição geométrica com média 10,  $M \sim \text{Geo}(0, 1)$ . Resumindo, temos:

$$\pi(\lambda^*) = N(\lambda^*|0, 100)$$

$$\pi(\phi_m) = \text{Gama}(\phi_m|0, 1, 0, 1)$$

e

$$\pi(M) = \text{Geo}(M|0, 1).$$

Considerando as distribuições *a priori*, foram geradas duas cadeias independentes com 330000 iterações, das quais foram descartadas as primeiras 5000 iterações. Adotamos um espaçamento de tamanho 65 para evitar a correlação das séries, obtendo uma amostra final de tamanho 5000.

Para monitorar a convergência fizemos uso do fator de redução de escala de potencial  $\hat{R}$ , proposto por Gelman & Rubin (1992). Na Tabela 4.1, apresentamos as probabilidades de corte das variáveis preditoras  $x_1$  e  $x_2$ . Note que a probabilidade de corte para  $x_2$  é próximo a zero. Isto era esperado, pois os dados foram gerados a partir da variável  $x_1$ . Então, dado que a tesselação por hiperplanos ortogonais seleciona as variáveis preditoras que têm realmente efeito na variável resposta e, por isso, a probabilidade de corte de  $x_2$  é próxima de zero e a probabilidade de corte para  $x_1$  é 1.

Tabela 4.1: Probabilidades de corte para as variáveis preditoras.

	Covariável	
	$x_1$	$x_2$
Probabilidade de corte	1,00	0,002

Na Figura 4.1, apresentamos a distribuição *a posteriori* para o número de regiões  $M$  da tesselação  $\mathcal{T}$ . Observamos que o número de regiões com maior probabilidade *a posteriori*, ocorre com  $M = 3$ .



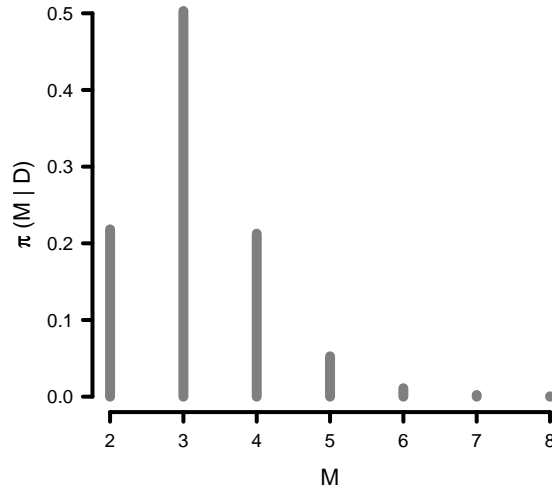


Figura 4.1: Distribuição *a posteriori* de  $M$ .

Na Tabela 4.2, apresentamos medidas-resumo *a posteriori* para o parâmetro da distribuição exponencial  $\lambda^*$ , sendo elas a média, o desvio padrão (DP) e o intervalo de credibilidade de 95% (IC(95%)).

Tabela 4.2: Resumo da distribuição *a posteriori* para o parâmetro da distribuição exponencial do modelo de tempo de promoção

Parâmetro	Média	Desvio padrão	IC(95%)	$\hat{R}$
$\lambda^*$	-1,954	0,1305	(-2, 220; -1, 701)	1,001

Observamos que  $\hat{R}$  é próximo a 1, indicando que o número de iterações foi suficiente para chegar à convergência. A estimativa do parâmetro da distribuição exponencial, é próxima ao verdadeiro parâmetro  $\lambda^* = -2$ . Também observamos que a estimativa do parâmetro  $\lambda^*$  é significativa, uma vez que o valor  $-2$  pertence ao intervalo de credibilidade.

A Figura 4.2 a seguir, mostra o histórico da sequência das iterações para o parâmetro  $\lambda^*$ , no qual observamos a convergência das cadeias. Também na Figura 4.3, apresentamos a função densidade marginal aproximada para  $\lambda^*$ .

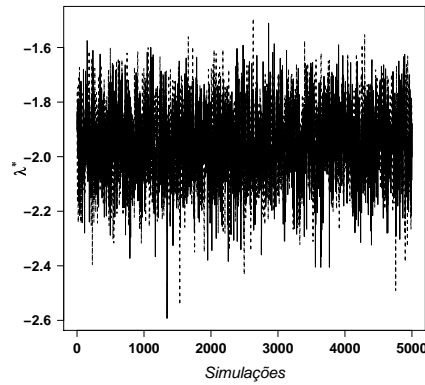


Figura 4.2: Histórico da sequência de iterações para o modelo de tempo de promoção (exponencial).

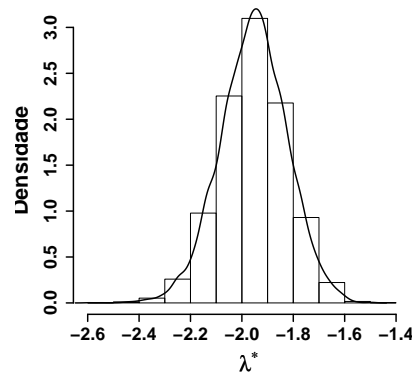


Figura 4.3: Densidade marginal *a posteriori* aproximada para o modelo de tempo de promoção (exponencial).

Dado que temos uma amostra de tamanho 5000, isto significa que temos 5000 estruturas de tesselação. Cada estrutura de tesselação está associada a um vetor de parâmetros  $\phi$ , que contém os parâmetros da distribuição de Poisson para cada região. Desta forma, levando em conta (3.38), a fração de cura média estimada de  $p_0$  é 0,358, a fração de cura média para estudo de simulação é  $p_0 = 0,358$ . Observamos, então, que a estimativa da fração de cura e a fração de cura para o conjunto de dados simulados coincidem.

## 4.2 Exemplo de aplicação

Resumidamente, o conjunto de dados para esta aplicação provém de um estudo de melanoma com o objetivo de avaliar a eficácia da aplicação de uma dosagem alta de interferon alfa-2b como forma de prevenir a recorrência de câncer. Os pacientes foram incluídos no estudo entre 1991 e 1995, tendo sido acompanhados até 1998. A variável resposta  $Y$ , representa o tempo até a morte de paciente ou tempo de censura. Nesta amostra, temos  $n = 417$  pacientes, com 56% de observações censuradas. As variáveis incluem  $y$ : tempo (em anos);  $x_1$ : tipo de tratamento (0: sem tratamento ; 1: interferon) ;  $x_2$ : idade ;  $x_3$ : categoria do nódulo (1,2,3,4) ;  $x_4$ : sexo (0: masculino; 1: feminino);  $x_5$ : capacidade funcional (0: ativo ; 1: outras) e  $x_6$ : espessura do tumor(em mm). Para mais detalhes deste conjunto de dados, Kirkwood *et al.* (2000) e Ibrahim *et al.* (2001) podem ser consultados. A Figura 4.4, mostra a estimativa de Kaplan-Meier da função de sobrevivência para este conjunto de dados, em que observamos que, neste conjunto de dados, há indícios de uma proporção de pacientes curados. Dessa forma, considerar uma análise de sobrevivência padrão pode ser inapropriado.

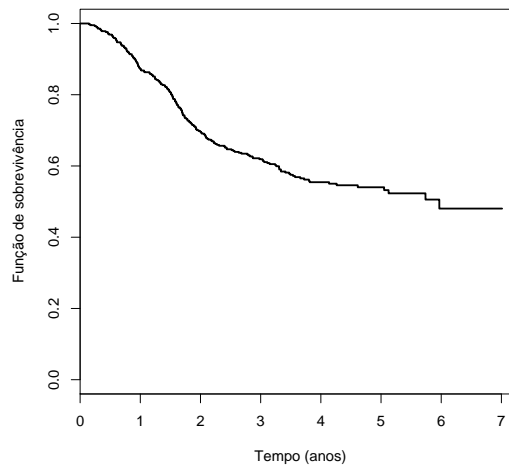


Figura 4.4: Estimativa de Kaplan-Meier da função de sobrevivência.

Neste estudo, consideramos a distribuição de probabilidade Weibull para os tempos de ocorrência do evento de interesse  $Z$ , tal como na Seção 2.1.1 Como

a metodologia é bayesiana, atribuímos distribuições *a priori* para os parâmetros do modelo.

Para os parâmetros da distribuição de Weibull  $\gamma = (\alpha, e^{\lambda^*})$  da Seção 3.3.2, temos que a distribuição *a priori* para  $\alpha$  é a distribuição gama. Neste estudo, assumimos hiperparâmetros  $\alpha_0 = \alpha_1 = 0, 1$ ,  $\pi(\alpha) \sim \text{Gama}(0, 1, 0, 1)$ . Para o parâmetro  $\lambda^*$ , temos a distribuição normal com média  $\mu = 0$  e variância  $\sigma^2 = 100$ ,  $\pi(\lambda^*) \sim \text{N}(0, 100)$ . Para os parâmetros  $\phi_m$ , em cada região  $R_m$ , a distribuição *a priori* é a distribuição gama com hiperparâmetros  $\vartheta_0 = \vartheta_1 = 0, 01$ , e para o número de regiões  $M$ , temos a distribuição geométrica com média 10,  $\pi(M) = \text{Geo}(0, 1)$ . Resumindo as distribuições *a priori*, temos:

$$\pi(\alpha) = \text{Gama}(\alpha | 0, 1, 0, 1),$$

$$\pi(\lambda^*) = \text{N}(\lambda^* | 0, 100),$$

$$\pi(\phi_m) = \text{Gama}(\phi_m | 0, 01, 0, 01)$$

e

$$\pi(M) = \text{Geo}(M | 0, 1).$$

Considerando estas distribuições *a priori*, foram geradas duas cadeias independentes com 2.600.000 iterações, das quais descartamos as primeiras 100.000 iterações e, visando evitar o problema da autocorrelação das séries, consideramos um salto de tamanho 50, conduzindo a uma amostra final de tamanho 5000 para cada cadeia. Para monitorar a convergência das cadeias fizemos uso do fator de redução de escala  $\hat{R}$ , proposto por Gelman & Rubin (1992).

Observamos que no conjunto de dados temos que  $x_1$ ,  $x_4$  e  $x_5$  são variáveis binárias, então, a divisão sobre qualquer destas variáveis, é tal que, se ocorrer, resultará em dois grupos, por exemplo,  $x_4$  (a variável que representa sexo), se for dividida, então os grupos que surgem são o masculino e o feminino. No caso da covariável  $x_3$ , em que temos quatro categorias,  $x_3 \in \{1, 2, 3, 4\}$ , a idéia da divisão em grupos de  $x_3$  foi tomada como no artigo de Giudici *et al.* (2000), só que em nosso exemplo, os grupos formados pelas categorias tem uma ordem. Desta forma, o número de grupos formados pela divisão  $x_3$ , são: 1,2,3 ou 4. Assim, os possíveis grupos representando a natureza ordinal da variável, são dados na

Tabela 4.3.

Tabela 4.3: Possíveis grupos para a covariável  $x_3$ .

Número de grupos	grupos
1	$\{1,2,3,4\}$
2	$\{1\},\{2,3,4\}$
	$\{1,2\},\{3,4\}$
3	$\{1,2,3\},\{4\}$
	$\{1\},\{2\},\{3,4\}$
	$\{1\},\{2,3\},\{4\}$
4	$\{1,2\},\{3\},\{4\}$
	$\{1\},\{2\},\{3\},\{4\}$

Na Tabela 4.3 observamos, por exemplo, que quando o número de grupos é 4, isto é,  $\{1\}, \{2\}, \{3\}, \{4\}$ , temos que as observações são distintas em cada categoria, no entanto, no caso em que temos apenas um grupo  $\{1, 2, 3, 4\}$ , significa que as categorias formam um mesmo grupo, então intuitivamente, as categorias são homogêneas porque fazem parte de um mesmo grupo. As covariáveis  $x_2$  e  $x_6$  são contínuas, assim, não tem restrições como as variáveis binárias ou categóricas, de modo que os hiperplanos (pontos de corte) para essas variáveis, são escolhidos de cada uma das distribuições marginais.

Na Tabela 4.4, apresentamos as probabilidades de corte para cada uma das covariáveis. Na Figura 4.5, observamos o histórico de cada iteração das variáveis preditoras presentes na tesselação  $\mathcal{T}$ , onde o ponto no gráfico significa que a variável está presente ou passou por um corte. Podemos notar, que a covariável  $x_3$  está sempre na estrutura da tesselação, portanto, a tesselação por hiperplanos ortogonais, considera que  $x_3$  tem um efeito significativo na variável resposta. Uma consequência disso, é a probabilidade de corte de  $x_3$  igual a 0,9939 muito próxima a 1. Este resultado não é novo, encontramos também em Mizoi (2004) e Rodrigues *et al.* (2008) obtendo resultados análogos, no sentido de que  $x_3$  é significativa. Para as outras covariáveis, a probabilidade de corte está próxima de zero, o que significa que este conjunto de variáveis preditoras são

não-informativas.

Na Figura 4.6, observamos a distribuição *a posteriori* para o número de regiões  $M$ , onde notamos que o número de regiões com maior probabilidade *a posteriori*, ocorre quando  $M = 2$ .

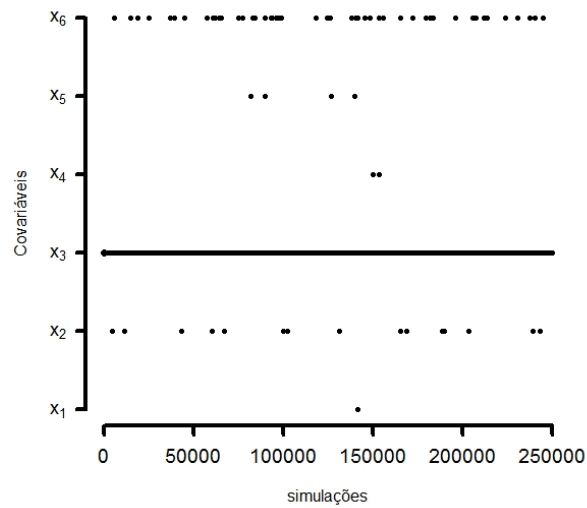


Figura 4.5: Histórico das covariáveis presentes na tesselação.

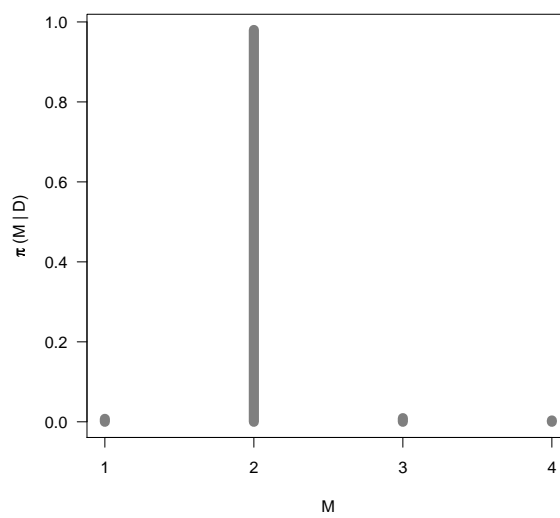


Figura 4.6: Distribuição *a posteriori* de  $M$ .

Tabela 4.4: Probabilidades de corte para cada uma das variáveis preditoras.

	Covariável					
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
Probabilidade de corte	0,0000	0,0013	0,9939	0,0000	0,0001	0,0018

Baseados nos resultados anteriores, apresentamos na Tabela 4.5 a probabilidade *a posteriori* de cada uma das partições para a covariável  $x_3$ .

Tabela 4.5: Probabilidades *a posteriori* para as partições da covariável  $x_3$ 

Partições	Probabilidade <i>a posteriori</i>
$\{1,2,3\},\{4\}$	0,8080
$\{1,2\},\{3,4\}$	0,1744
$\{1,2,3,4\}$	0,0069
$\{1\},\{2,3\},\{4\}$	0,0050
$\{1,2\},\{3\},\{4\}$	0,0040
$\{1\},\{2,3,4\}$	0,0015
$\{1\},\{2\},\{3,4\}$	0,0002
$\{1\},\{2\},\{3\},\{4\}$	0,0000

Observamos que para este conjunto de dados, o modelo de partição bayesiano considera que a partição com maior probabilidade *a posteriori*, é a partição formada por  $\{1, 2, 3\}$  e  $\{4\}$ , isto é, a probabilidade *a posteriori* é 0,8080, enquanto a partição formada pelas categorias  $\{1\}, \{2\}, \{3\}, \{4\}$  tem probabilidade *a posteriori* zero. Isto significa, que considerar esta partição na modelagem não é apropriado.

Na Tabela 4.6, apresentamos as medidas-resumo *a posteriori* dos parâmetros  $\alpha$  e  $\lambda^*$  da distribuição Weibull, tais como a média, desvio padrão e intervalo de 95% de credibilidade IC (95%). Na Figura 4.7, mostramos o histórico das iterações para cada um dos parâmetros. O histórico indica a convergência das cadeias.

Tabela 4.6: Resumos a *posteriori* para os parâmetros da distribuição Weibull do modelo de tempo de promoção

Parâmetro	Média	Desvio padrão	IC (95%)	$\hat{R}$
$\alpha$	1,721	0,116	(1, 495; 1, 947)	1,000
$\lambda^*$	-1,647	0,135	(-1, 920; -1, 388)	1,003

Notamos na Tabela 4.6 que os parâmetros da distribuição de Weibull são significativos e, também, para ambos parâmetros, observamos valores de  $\hat{R}$  próximos a 1, indicando que as iterações foram suficientes para atingir a convergência.

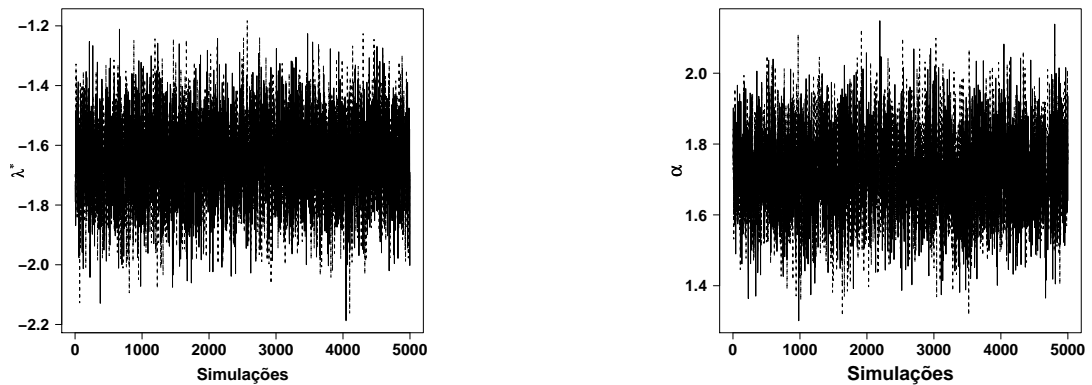


Figura 4.7: Histórico da seqüência de iterações dos parâmetros da distribuição Weibull do modelo de tempo de promoção.

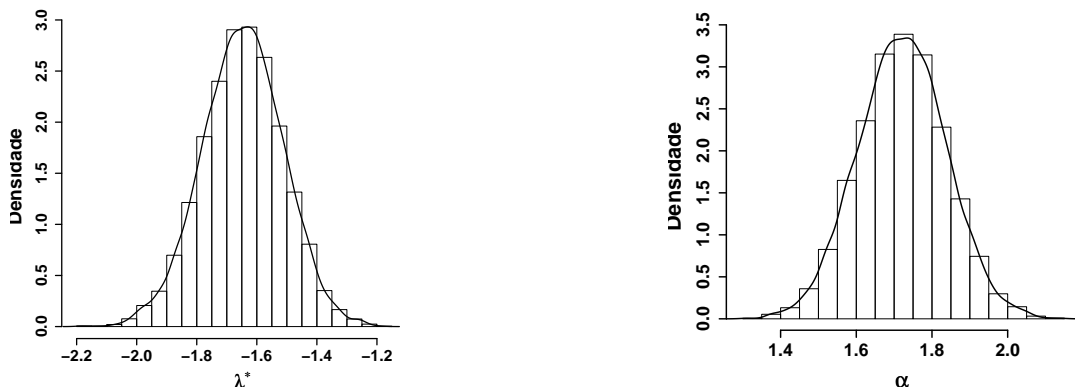


Figura 4.8: Histograma e função de densidades marginais a *posteriori* aproximadas para os parâmetros da distribuição Weibull do modelo de tempo de promoção.



A Figura 4.9 mostra as estimativas de Kaplan - Meier da função de sobrevivência, assim como a estimativa obtida do modelo de longa duração (tempo de promoção) com partição bayesiana, considerando a partição com maior probabilidade para a covariável  $x_3$ , isto é, a partição formada por  $\{1, 2, 3\}$  e  $\{4\}$ , e na Figura 4.10, a segunda partição de maior probabilidade, isto é,  $\{1, 2\}$  e  $\{3, 4\}$ .

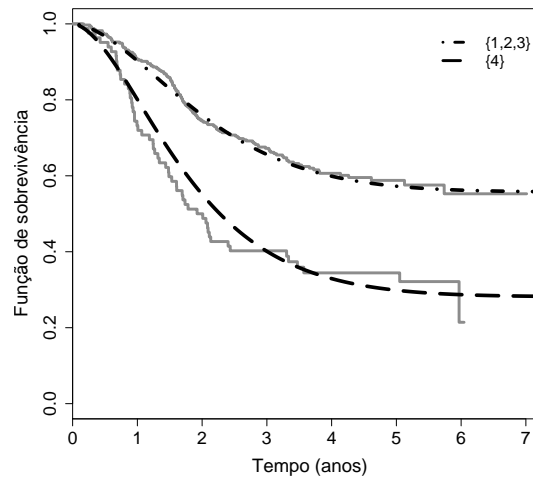


Figura 4.9: Estimativa de Kaplan-Meier e paramétrica da função de sobrevivência de acordo com a partição  $\{1, 2, 3\}$  e  $\{4\}$  de  $x_3$ .

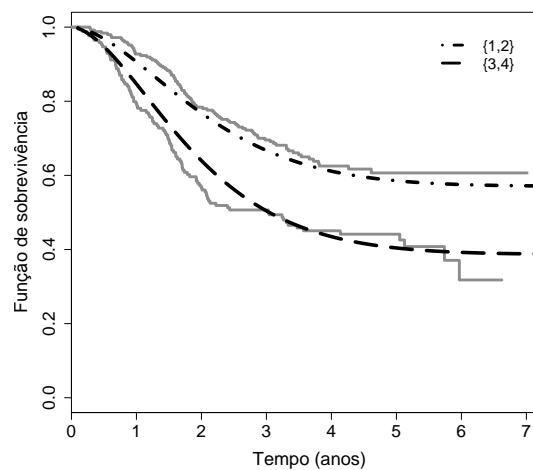


Figura 4.10: Estimativa de Kaplan-Meier e paramétrica da função de sobrevivência de acordo com a partição  $\{1, 2\}$  e  $\{3, 4\}$  de  $x_3$ .

Conforme já vimos, a fração de cura para uma observação pertencente à região  $R_m$ , é dada por  $p_0 = \exp(\phi_m)$ ,  $\phi_m > 0$ . As estimativas foram obtidas considerando (3.38) a fração de cura média estimada para este conjunto de dados, é 0,507. A estimativa anterior, é similar ao obtido por Mizoi (2004), que utilizou o algoritmo EM.

Consideramos também, um pequeno estudo de sensibilidade por meio da escolha de diferentes valores de hiperparâmetros para as distribuições *a priori* como por exemplo,  $\alpha \sim \text{Gama}(0,01,0,01)$  e  $\alpha \sim \text{Gama}(0,001,0,001)$ , em que as diferentes escolhas levaram aos mesmos resultados.

Para avaliar a capacidade preditiva *a posteriori* do modelo de longa duração (tempo de promoção) com partição bayesiana em estudo, em contraste com o modelo de tempo de promoção estudado por Chen *et al.* (1999), dividimos o conjunto de dados em dois grupos, 70% para gerar os parâmetros do modelo e 30% para calcular a probabilidade preditiva *a posteriori* para essas observações, considerando os parâmetros estimados dos dados de treinamento (70%). O grupo que considera 30% dos dados é chamado conjunto de dados teste. Neste contexto, denotemos por  $(\mathbf{y}', \boldsymbol{\delta}')$  o conjunto de dados teste, assim, considerando que temos uma amostra de tamanho  $B$  obtida pelo método MCMC (Gamerman & Lopes, 2006), então, a probabilidade preditiva *a posteriori* é dada por:

$$\pi(\mathbf{y}', \boldsymbol{\delta}' | \mathbf{y}, \boldsymbol{\delta}) = \frac{1}{B} \sum_{b=1}^B L(\mathbf{y}', \boldsymbol{\delta}' | \mathcal{T}^b, \alpha^b, \lambda^b, \phi^b), \quad (4.1)$$

em que

$$L(\mathbf{y}', \boldsymbol{\delta}' | \mathcal{T}^b, \alpha^b, \lambda^b, \phi^b) = \prod_{i'=1}^{n'} \sum_{n^*=0}^{\infty} \{S(y'_{i'} | \gamma)\}^{N_{i'} - \delta_{i'}} \{N_{i'} f(y'_{i'} | \gamma)\}^{\delta_{i'}} p(N_{i'} = n^* | \phi_{i'}).$$

Considerando que os tempos para a ocorrência do evento de interesse  $Z$  na Seção 2.1.1, têm distribuição de probabilidade de Weibull, a expressão anterior fica da seguinte forma:

$$L(\mathbf{y}', \boldsymbol{\delta}' | \mathcal{T}^b, \alpha^b, \lambda^b, \phi^b) = \prod_{i'=1}^{n'} (\phi_{i'} \alpha \lambda y_{i'}^{\alpha-1} \exp(-\lambda y_{i'}^{\alpha-1}))^{\delta_{i'}} \exp\{-\phi_{i'}(1 - \exp(-\lambda y_{i'}^{\alpha}))\}.$$

Na Tabela 4.7, apresentamos o logaritmo da distribuição preditiva *a posteriori* considerando o conjunto de dados teste. Observamos que o modelo

longa duração (tempo de promoção) com partição bayesiana em comparação ao modelo de tempo promoção estudado por Chen *et al.* (1999), tem melhor capacidade preditiva, isto é, tem menor valor para o logaritmo da distribuição preditiva *a posteriori* ( $-166,7548$ ). Resultados análogos com outro conjunto de dados podem ser encontrados em Hoggart & Griffin (2001), onde a capacidade preditiva da extensão do modelo de tempo de promoção também foi considerada com melhor desempenho.

Tabela 4.7: Logaritmo da distribuição preditiva *a posteriori*, para a extensão do modelo de partição bayesiana e o modelo de tempo de promoção

Modelo de partição bayesiana	Modelo tempo de promoção
-166,7548	-170,0967

## Capítulo 5

### Considerações finais e propostas

Neste trabalho, apresentamos os resultados mais relevantes da teoria unificada de longa duração apresentada em Rodrigues *et al.* (2009a). O modelo de mistura padrão (Boag, 1949; Berkson & Gage, 1952) e o modelo de tempo de promoção (Yakovlev & Tsodikov, 1996; Chen *et al.*, 1999), são casos particulares desta teoria. Um resultado importante desta teoria, está relacionado com a propriedade de riscos proporcionais, em que a função de risco da população tem a propriedade de riscos proporcionais se, e somente se, o número de causas para a ocorrência de evento de interesse tem a distribuição de Poisson.

Também apresentamos a metodologia do modelo de partição bayesiano proposto por Holmes *et al.* (1999), cuja característica especial é sua capacidade preditiva, como pode ser visto em Denison & Holmes (2001), Hoggart & Griffin (2001), Stephenson *et al.* (2006), Hopcroft *et al.* (2009), entre outros.

Assim, utilizando a metodologia de partição bayesiana, apresentamos uma extensão do modelo de tempo de promoção proposto por Hoggart & Griffin (2001), em que foram consideradas variáveis preditoras categóricas (Giudici *et al.*, 2000) em um conjunto de dados reais (Kirkwood *et al.*, 2000; Ibrahim *et al.*, 2001). Ademais, comparamos a capacidade preditiva do modelo de partição bayesiano para dados de longa duração com o modelo de tempo de promoção (Yakovlev & Tsodikov, 1996; Chen *et al.*, 1999). Uma característica importante

na modelagem de partição bayesiana com hiperplanos ortogonais, é a natural seleção de covariáveis. Cada hiperplano divide o conjunto de dados em somente uma covariável, e assim, os hiperplanos são incluídos quando a covariável afeta o ajuste do modelo. Em nossa experiência, o modelo de partição bayesiano com hiperplanos ortogonais mostrou-se computacionalmente eficiente, quando somente uma pequena proporção das covariáveis são utilizadas para predição.

As principais contribuições neste trabalho, foram os estudos detalhados da teoria unificada do modelo de longa duração proposto por Rodrigues *et al.* (2009a), o modelo de partição de bayesiano proposto por Holmes *et al.* (1999) e a extensão local dos modelos de longa duração proposto por Hoggart & Griffin (2001), a partir das quais foi possível desenvolver uma estratégia computacional para análise dos dados, considerando variáveis preditoras contínuas e categóricas, em que a estratégia computacional seleciona de forma natural as variáveis preditoras que são significativas no modelo. Além disso foi demonstrada a capacidade preditiva da modelagem de partição bayesiana, o objetivo principal desta metodologia.

A programação da estratégia computacional, está baseada em linguagem de programação R (R Development Core Team, 2009); para detalhes do programa podem ser consultados aos autores do trabalho.

De forma geral, o modelo de partição bayesiano com hiperplanos ortogonais mostrou-se eficiente diante de um grande conjunto de dados e quando utilizamos a distribuição Poisson para o número de causas do evento de interesse. Por outro lado, a distribuição de Poisson não considera a questão de subdispersão e sobredispersão. Para flexibilidade do modelo em estudo, em um trabalho futuro pretendemos considerar a distribuição binomial negativa (Ross & Preece, 1985; Piegorsch, 1990) para modelar o número de causas ou riscos para o evento de interesse, em que a fração de cura é considerada como parâmetro da distribuição binomial negativa. Também pretendemos considerar a distribuição COM-Poisson para o número de causas do evento de interesse  $N$ , o que encontramos em Rodrigues *et al.* (2009b).

# Referências

- Barry, D. & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, **80**(421), 309–319.
- Berkson, J. & Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(259), 501–515.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, **11**(1), 15–53.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, California.
- Carlin, B. P., Gelfand, A. E. & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, **41**(2), 389–405.
- 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**(447), 909–919.
- Chipman, H., George, E. I. & McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association*, **93**(443), 935–960.
- Denison, D. G. T. & Holmes, C. C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics*, **57**(1), 143–149.
- Denison, D. G. T., Mallick, B. K. & Smith, A. F. M. (1998). A Bayesian CART algorithm. *Biometrika*, **85**(2), 363–377.
- Denison, D. G. T., Adams, N. M., Holmes, C. C. & Hand, D. J. (2002a). Bayesian partition modelling. *Computational Statistics & Data Analysis*, **38**(4), 475–485.
- Denison, D. G. T., Holmes, C. C., Mallick, B. K. & Smith, A. F. M. (2002b). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, Chichester.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York.
- Gamerman, D. & Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/ CRC, Boca Raton, FL, second edition.

- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**(4), 457–472.
- Giudici, P., Knorr-Held, L. & Rasser, G. (2000). Modelling categorical covariates in Bayesian disease mapping by partition structures. *Statistics in Medicine*, **19**(17–18), 2579–2593.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.
- Green, P. J. & Sibson, R. (1978). Computing Dirichlet tessellations in the plane. *Statistical Science*, **21**(2), 168–173.
- Heikkinen, J. (1998). Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and Semiparametric Bayesian Statistics, no. 133 in Lecture Notes in Statistics, chap. 14*, pages 255–272. Springer-Verlag.
- Hoggart, C. & Griffin, J. E. (2001). A Bayesian partition model for customer attrition. In E. I. George, editor, *Bayesian Methods with Applications to Science, Policy, and Official Statistics (Selected Papers from ISBA 2000)*, pages 61–70, Creta, Greece. International Society for Bayesian Analysis, Proceedings of the the Sixth World Meeting of the International Society for Bayesian Analysis.
- Holmes, C., Denison, D. & Mallick, B. (1999). Bayesian Partitioning for Classification and Regression. Technical report, Department of Mathematics, Imperial College.
- Hopcroft, P., Gallagher, K. & Pain, C. (2009). A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion. *Geophysical Journal International*, **178**(2), 651–666.
- Ibrahim, J., Chen, M. & Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York, NY.
- Iritani, M. (2008). *Modelos de sobrevivência de longa duração: uma abordagem unificada*. Dissertação de mestrado, DES – Universidade Federal de São Carlos, São Carlos, SP.
- Kaplan, E. & Meir, P. (1958). Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*, **53**, 457–481.
- Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**(430), 773–795.
- Kirkwood, J. M., Ibrahim, J., Sondak, V., Richards, J., Flaherty, L., Ernstoff, M., Smith, T., Rao, U., Steele, M. & Blum, H. (2000). High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. *Journal of Clinical Oncology*, **18**(12), 2444–2458.
- Lawless, J. (2002). *Statistical Models and Methods for Lifetime Data*. Wiley, New York, NY, second edition.

- Leonte, D. & Nott, D. J. (2006). Bayesian spatial modelling of gamma ray count data. *Mathematical Geology*, **38**(2), 135–154.
- Loschi, R. & Cruz, F. (2005). Extension to the product partition model: computing the probability of a change. *Computational Statistics & Data Analysis*, **48**(2), 255–268.
- Maller, R. & Zhou, S. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York, NY.
- Meeker, W. & Escobar, L. (1998). *Statistical Methods for Reliability Data*. Wiley, New York, NY.
- Mizoi, M. (2004). *Influência local em modelos de sobrevivência com fração de cura*. Tese de doutorado, IME – Universidade de São Paulo, São Paulo.
- Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, **46**(3), 863—867.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rodrigues, J., Cancho, V. G. & de Castro, M. (2008). *Teoria Unificada de Análise de Sobrevivência*. Associação Brasileira de Estatística, São Paulo, SP.
- Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009a). On the unification of long-term survival models. *Statistics & Probability Letters*, **79**(6), 753–759.
- Rodrigues, J., de Castro, M., Cancho, V. G. & Balakrishnan, N. (2009b). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, **130**(10), 3605—3611.
- Ross, G. J. S. & Preece, D. A. (1985). The negative binomial distribution. *The Statistician*, **34**(3), 323—336.
- Saha, K. & Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, **61**(1), 179–185.
- Stephens, D. A. (1994). Bayesian retrospective multiple-change-point identification. *Applied Statistics*, **43**(1), 159–178.
- Stephenson, J., Gallagher, K. & Holmes, C. (2006). Low temperature thermochronology and strategies for multiple samples: 2: partition modelling for 2D/3D distributions with discontinuities. *Earth and Planetary Science Letters*, **241**(3–4), 557–570.
- Tsodikov, A., Ibrahim, J. & Yakovlev, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**(464), 1063–1078.
- Yakovlev, A. & Tsodikov, A. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.



---

Yao, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics*, **12**(4), 1434–1447.