

---

Modelos de Sobrevivência de  
Longa-Duração: Uma Abordagem  
Unificada

*Mateus Rodrigues Iritani*

---



# Modelos de Sobrevivência de Longa-Duração: Uma Abordagem Unificada

*Mateus Rodrigues Iritani*

Orientador: *Prof. Dr. Josemar Rodrigues*

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

Junho - 2008

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

1689ms

Iritani, Mateus Rodrigues.

Modelos de sobrevivência de longa-duração : uma abordagem unificada / Mateus Rodrigues Iritani. -- São Carlos : UFSCar, 2008.

65 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2008.

1. Análise de sobrevivência. 2. Distribuição de Poisson. 3. Distribuição binomial. 4. Função geradora. 5. Risco competitivo. I. Título.

CDD: 519.4 (20<sup>a</sup>)



*Aos meus pais,  
Geraldo e  
Maria.*



# Resumo

Em análise de sobrevivência, determinados estudos caracterizam-se por apresentar uma fração significativa de sobreviventes, ou seja, pacientes em tratamento que não apresentaram o evento de interesse, mesmo após um longo período de acompanhamento. Assim considerar modelos de sobrevivência usuais, que assumem que a função de sobrevivência converge para zero quando a variável tempo tende a infinito, pode não ser adequado. Nesse trabalho é apresentado uma extensão do modelo proposto por Chen, Ibrahim e Sinha (1999), usando a função geradora de uma sequência de números reais introduzida por Feller (1967). Essa extensão possibilitou o desenvolvimento de uma teoria unificada para os modelos de sobrevivência de longa-duração, Rodrigues *et al.* (2008). Mostra-se que modelos já existentes na literatura são considerados casos particulares da teoria unificada, por exemplo, o modelo de Berkson & Gage (1952). Também tem-se em Rodrigues *et al.* (2008), que a função geradora de longa-duração satisfaz a propriedade de risco proporcional se, e somente se, o número de causas competitivas relacionadas a ocorrência do evento de interesse segue uma distribuição de Poisson. Como ilustração utiliza-se um conjunto de dados reais.

**Palavras-chave:** Longa-Duração, Poisson, Bernoulli, Risco Competitivo, Função Geradora, Função de Risco Proporcional.



# Abstract

In survival analysis some studies show a meaningful cure rate after treatment follow-up, so considering standard survival models can not be appropriate. In this work is extended the long-term survival model proposed by Chen, Ibrahim and Sinha (1999) via generating function of a real sequence introduced by Feller (1967). This new formulation is the unification of the long-term survival models proposed by Rodrigues *el al.* (2008). Also, as in Rodrigues *el al.* (2008) it is shown that the long-term survival generating function satisfies the proportional hazard property if only if the number of competing causes related to the occurrence of a event of interest follows a Poisson distribution. A real data set is considered to illustrate this approach.

**Keywords:** Long-Term Survival, Poisson, Bernolli, Competing Risks, Generating Function, Proportional Hazards Functions.

# Sumário

---

<b>Introdução</b>	<b>i</b>
<b>1 Conceitos Básicos</b>	<b>3</b>
1.1 Introdução . . . . .	3
1.2 Conceitos da Distribuição do Tempo de Vida . . . . .	4
1.3 Estimador de Kaplan-Meier . . . . .	5
1.4 Teorema de Bayes . . . . .	6
1.5 Métodos MCMC . . . . .	7
1.5.1 Cadeias de Markov . . . . .	7
1.5.2 Algoritmo Metropolis-Hastings . . . . .	7
1.5.3 Amostrados de Gibbs . . . . .	9
1.5.4 Diagnósticos de Convergência . . . . .	10
<b>2 Teoria Unificada para Modelos de Longa-Duração</b>	<b>13</b>
2.1 Introdução . . . . .	13
2.2 Função Geradora de Sobrevivência de Longa- Duração . . . . .	14
2.3 Função de Verossimilhança . . . . .	18
2.3.1 Função de Verossimilhança Unificada . . . . .	19
2.4 Inferência Clássica . . . . .	20
2.5 Inferência Bayesiana . . . . .	21
<b>3 Modelo de Fração de Cura com Poisson</b>	<b>23</b>
3.1 Introdução . . . . .	23
3.2 Caso particular: Poisson . . . . .	24
3.3 Função de Verossimilhança . . . . .	25
3.4 Estimação dos Parâmetros do Modelo . . . . .	27
3.5 Modelo Paramétrico com Distribuição Weibull . . . . .	28
3.6 Abordagem Bayesiana . . . . .	29

<b>4</b>	<b>Modelo de Fração de Cura com Bernoulli</b>	<b>31</b>
4.1	Introdução . . . . .	31
4.2	Caso particular: Bernoulli . . . . .	31
4.3	Função de Verossimilhança . . . . .	33
4.4	Estimação dos Parâmetros . . . . .	33
4.5	Abordagem Bayesiana . . . . .	36
<b>5</b>	<b>Aplicação</b>	<b>39</b>
5.1	Descrição dos Dados . . . . .	39
5.2	Análise Clássica . . . . .	40
5.3	Análise Bayesiana . . . . .	44
5.4	Comparação dos Modelos . . . . .	46
5.5	Estudo de Simulação . . . . .	47
5.6	Testes de Hipótese . . . . .	49
<b>6</b>	<b>Considerações Finais</b>	<b>53</b>
6.1	Conclusão . . . . .	53
6.2	Estudos Propostos . . . . .	54
<b>A</b>	<b>Função de Verossimilhança</b>	<b>55</b>
<b>B</b>	<b>Função de Verossimilhança Unificada</b>	<b>57</b>
<b>C</b>	<b>Distribuição Condicional de <math>N</math> e <math>W</math></b>	<b>59</b>
<b>D</b>	<b>Condição para Teorema 4</b>	<b>61</b>
	<b>Referências Bibliográficas</b>	<b>63</b>

# Introdução

Em análise de sobrevivência é usual assumir a possibilidade de todos os indivíduos em estudo terem experimentado o evento de interesse. Este evento é associado a uma variável resposta ( $y$ ) que é caracterizada como o tempo até a ocorrência do evento (dependendo do estudo). Por sua vez, esta variável é relacionada a uma função de sobrevivência que mede a probabilidade do indivíduo sobreviver a um tempo  $y$  e é denotada por

$$S(y) = P(Y > y).$$

Na análise de sobrevivência padrão a função de sobrevivência converge para zero quando o tempo tende a infinito, ou seja, todos os indivíduos em estudos são suscetíveis ao evento de interesse. Para exemplificar, a Figura 1 mostra o gráfico do estimador de Kaplan-Meier para o estudo realizado com camundongos para verificar a eficácia da imunização pela malária. O gráfico é referente ao Grupo 2 de camundongos (os dados estão em Colosimo & Giolo 2006 pg.14), em que todos sofreram o evento de interesse definido como o tempo desde a infecção pela malária até a morte.

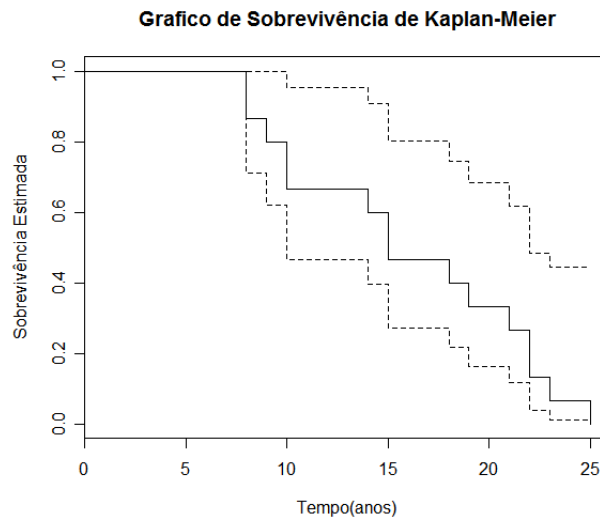


Figura 1: Estimativas Kaplan-Meier para dados de Malária

O estimador não paramétrico de Kaplan-Meier foi proposto por Kaplan & Meier (1958) para estimar a função de sobrevivência e também é chamado por estimador limite-produto.

Entretanto, em alguns estudos clínicos os dados de tempo de falha de uma população são constituídos de dois grupos: os indivíduos suscetíveis e os não suscetíveis ao evento de interesse. O primeiro pode eventualmente sofrer do evento, enquanto que o segundo normalmente aparenta estar livre de sinais ou situações do mesmo, podendo ser

chamado de fração de curados, imunes ou sobreviventes de longa-duração. Essa característica aparece em estudos em que os dados têm uma grande quantidade de censuras no final do período de observação, como por exemplo, reincidência de câncer, desenvolvimento de AIDS em pacientes HIV positivo e durabilidade de componentes eletrônicos.

A suposição de alguns pacientes nunca experimentarem o evento de interesse é baseado em considerações científicas ou empíricas, como a presença de um grande número de sobreviventes de longa-duração (fração de “curados”). O estimador de Kaplan-Meier é uma boa forma de evidenciar essa presença, uma vez que um grande número de censura pode ser observado na calda, ou seja, é plausível aceitar 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 intervalo de tempo considerável. Como exemplo, a Figura 2 abaixo apresenta o gráfico da distribuição estimada de Kaplan-Meier para o tempo de sobrevivência de um conjunto de dados de reincidência de câncer Melanoma Fase III. O Melanoma é um câncer de pele maligno caracterizado por cinco fases: na fase zero, como a menos grave, as células encontram-se apenas na camadas exterior da pele, e na fase mais crítica, as células malignas metastizaram para outros órgãos (fonte: Google). Os dados foram disponibilizados por Ibrahim *et al.* (2001) e foram analisados no capítulo 5.

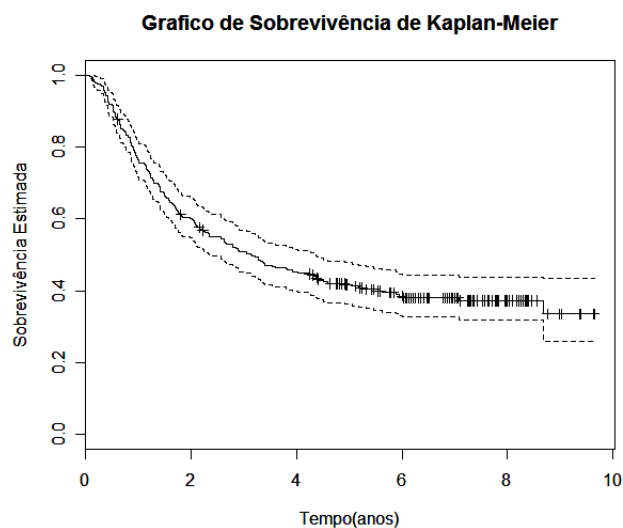


Figura 2: Estimativas Kaplan-Meier para dados de Melanoma

Dessa forma, considerar uma análise de sobrevivência padrão pode ser inapropriada.

Muitos modelos tem surgido na literatura para modelar conjuntos de dados com sobreviventes de longa-duração com os seguintes objetivos: estudar a fração de curados, o efeito do método de tratamento sobre essa fração, a distribuição da função de sobrevivência e o efeito de possíveis covariáveis sobre os indivíduos.

Uma forma de modelar é considerar uma função de sobrevivência como uma mistura de pacientes curados e não curados, onde a função de sobrevivência dos não curados é própria (ou seja, a integral no intervalo  $[0, \infty)$  da função de densidade associada a esta função de sobrevivência tem valor igual a 1, logo a função de sobrevivência tende a um valor nulo). Nesse contexto, o modelo mais conhecido é o de Berkson & Gage (1952) que introduziram um modelo paramétrico assumindo que uma fração  $\pi$  da população total é "curada". Esse modelo foi baseado na idéia de Boag (1949). A função de sobrevivência da população é expressa como

$$S_1(t) = \pi + (1 - \pi)S^*(t), \quad t \geq 0, \quad (1)$$

tal que  $S^*(t)$  é a função de sobrevivência própria associada aos indivíduos não curados. Este modelo tem sido estudado por vários autores que usaram diferentes métodos para ajustar a razão de cura. Por exemplo, Farewell (1982) utilizou risco proporcional de Cox, e Kuk & Chen (1992) combinou a formulação logística e risco proporcional propondo uma generalização semiparamétrica para modelo de Farewell (1982). Entre outros temos, Farewell (1986), Goldman (1984 e 1991), Greenhouse & Wolfe (1984), Taylor (1995) e Maller & Zhou (1996).

Alternativamente, existem modelos de sobrevivência baseados em estrutura de risco competitivos. Neste caso, é definida uma variável aleatória como o número de causas competindo para a ocorrência do evento de interesse que por sua vez, é associada a variável resposta do modelo. Por exemplo, em estudos de câncer as células cancerígenas competem entre si para dar origem ao tumor visível e isto é relacionada a uma variável que marca o tempo até a origem do tumor. Assim, a variável resposta é definida como o menor dentre estes tempos. Baseada nesta definição, Yakovlev *et al.* (1993) determinou a função de sobrevivência como sendo

$$S(t) = \exp\{-\theta F(t)\}, \quad (2)$$

onde considerou o número de causas com distribuição de Poisson( $\theta$ ).

Na literatura, a equação (2) é chamado modelo de sobrevivência para tempo de ativação e um fato importante acontece quando o parâmetro  $\theta$  é definido como uma função de covariáveis, isto caracteriza o modelo (2) com uma estrutura de risco proporcional, enquanto que no modelo (1) isso não ocorre.

Estudos baseados no modelo (2) foram realizados por vários autores, como por exemplo, Yakovlev & Tsodikov (1996), Hoggart & Griffin (2000), Ibrahim *et al.* (2001) e Chen,

Ibrahim & Sinha (1999). Estes últimos apresentam uma abordagem Bayesiana como uma alternativa aos métodos tradicionais utilizados para o modelo de Berkson & Gage, que é apresentado em detalhes no próximo capítulo.

Em estudos mais recentes, Yin & Ibrahim (2005) introduziram uma nova classe de modelos de fração de cura formulada através da transformação (não usual) Box-Cox sobre uma desconhecida função de sobrevivência da população.

Neste trabalho os modelos de sobrevivência de longa-duração é apresentado considerando uma abordagem unificada proposta por Rodrigues *et al.* (2008), que consiste em uma generalização que possibilita a obtenção da função de sobrevivência de longa-duração da população de qualquer modelo baseado em uma distribuição genérica do número de causas de ocorrência do evento. Isso é possível através do uso da função geradora de sequência de números reais definida por Feller (1967).

Os objetivos desse trabalho são: apresentar a teoria unificada proposta por Rodrigues *et al.* (2008); reproduzir a aplicação do modelo de Chen, Ibrahim & Sinha (1999); desenvolver testes de hipóteses para os parâmetros de regressão; comparar os modelos de Berkson & Gage (1952) e Chen, Ibrahim & Sinha (1999) do ponto de vista clássico e bayesiano; e por fim desenvolver um estudo de simulação para verificar o comportamento dos modelos em diferentes situações .

Assim o trabalho está organizado da seguinte forma. No Capítulo 2, é apresentado a função geradora de longa-duração e as suas propriedades. No Capítulo 3 e 4, é apresentado dois casos particulares como uma ilustração do modelo unificado, considerando uma distribuição de Poisson e uma distribuição de Bernoulli, respectivamente. Em seguida, no Capítulo 5, é apresentado os resultados da aplicação com dados de câncer para os modelos apresentados nos Capítulos 3 e 4. E finalmente, no Capítulo 6 a conclusão.





# Capítulo 1

---

## Conceitos Básicos

---

Neste capítulo é apresentado conceitos básicos e ferramentas usados na teoria e nas análises deste trabalho.

### 1.1 Introdução

Em análise de sobrevivência, a variável resposta é normalmente definida como o tempo até a ocorrência de um evento de interesse, e é denominado como **tempo de falha**. Como exemplo, pode ser definido como o tempo até a recorrência, o tempo até a morte do indivíduo ou até a cura de uma doença.

A principal característica de estudos com dados de sobrevivência é a presença de censuras, que é a informação parcial da resposta para alguns indivíduos. Situações assim ocorrem quando o acompanhamento do paciente foi interrompido, o estudo terminou para a análise dos dados ou, o paciente morreu de causa diferente da estudada. Em situações como estas, é considerado que toda informação referente à resposta é resumida no tempo de falha ser maior que o tempo observado. Desta forma, os conjuntos de dados de sobrevivência são caracterizados por tempo de falha e, frequentemente, por censuras.

Existem três tipos de censuras. A censura Tipo I ocorre quando o estudo será terminado após um período pré-estabelecido de tempo. A censura Tipo II é aquela em que o estudo será terminado após ocorrer o evento de interesse em um número pré-estabelecido de indivíduos. E a censura do Tipo aleatória que ocorre quando um indivíduo é retirado do estudo sem ter ocorrido a falha.

A censura aleatória é representado por duas variáveis aleatórias. Considere  $Y$  uma variável aleatória representando o tempo de falha e  $C$  uma outra variável aleatória independente de  $Y$ , representando o tempo de censura. Então o tempo observado é dado por

$$y = \min(Y, C)$$

e

$$\delta = \begin{cases} 1 & \text{se } T \leq C \\ 0 & \text{se } T > C \end{cases}$$

Assim, os dados de sobrevivência para o indivíduo  $i$ ,  $i = 1, \dots, m$ , é representado por  $(y_i, \delta_i)$ . Observe que a censura do Tipo I é caso particular da censura aleatória quando todo  $C_i = C$ .

## 1.2 Conceitos da Distribuição do Tempo de Vida

A variável aleatória não-negativa  $Y$ , que representa o tempo de falha, é especificada pela função de sobrevivência ou pela função de risco (ou taxa de falha). A seguir, apresenta-se as definições e as relações destas funções (Lawless, 1982).

Seja  $f(y)$  a função densidade de probabilidade de  $Y$  e a função de distribuição acumulada

$$F(y) = P(Y \leq y) = \int_0^y f(x) dx. \quad (1.1)$$

A função de sobrevivência é definida como a probabilidade de uma observação sobreviver ao tempo  $y$ , ou seja,

$$S(y) = P(Y \geq y) = \int_y^\infty f(x) dx = 1 - F(y). \quad (1.2)$$

A função de risco de  $Y$  é definida como

$$h(y) = \lim_{\Delta y \rightarrow 0} \frac{P(y \leq Y < y + \Delta y | Y \geq y)}{\Delta y} = \frac{f(y)}{S(y)}. \quad (1.3)$$

Esta função representa a razão instantânea no tempo  $y$  dado que o indivíduo sobreviveu até o tempo  $y$ . Esta função ajuda à descrever a distribuição do tempo de vida, ou seja, descreve a forma que a taxa instantânea de falha muda com o tempo.

As funções  $f(y)$ ,  $F(y)$ ,  $S(y)$  e  $h(y)$  são matematicamente equivalentes, ou seja, qualquer uma delas pode ser obtida a partir do conhecimento de pelo menos uma delas. Assim tem-se as seguintes relações:

$$h(y) = \frac{f(y)}{S(y)} = -\frac{d}{dy} \log S(y) \quad (1.4)$$

implica em

$$H(y) = \int_0^y h(x) dx = -\log S(y) \quad (1.5)$$

e

$$S(y) = \exp\left(-\int_0^y h(x) dx\right) = \exp(-H(y)). \quad (1.6)$$

Pode-se observar que por definição tem-se  $S(0) = 1$  e  $S(\infty) = \lim_{y \rightarrow \infty} S(y) = 0$  então  $H(\infty) = \lim_{y \rightarrow \infty} H(y) = \infty$ .

Finalmente pode-se obter a função densidade de probabilidade por

$$f(y) = h(y) \exp\left(-\int_0^y h(x) dx\right). \quad (1.7)$$

### 1.3 Estimador de Kaplan-Meier

O passo inicial em qualquer análise estatística consiste em uma descrição dos dados. Em conjunto de dados de sobrevivência a presença de censuras se torna um problema para as técnicas convencionais de análise descritivas, envolvendo média, desvio padrão e técnicas gráficas, como box-plot, histogramas e outros. Como ilustração considere o interesse de construir um histograma, na presença de censuras não é possível conhecer a frequência exata associada a cada intervalo. Dessa forma, o principal componente da análise descritiva em dados de sobrevivência é a função de sobrevivência.

O processo inicial é encontrar uma estimativa da função de sobrevivência e a partir dela encontrar as estimativas do tempo médio ou mediano, alguns percentis ou frações de falha em tempos fixos de acompanhamento.

O mais conhecido e usado estimador da função de sobrevivência é o estimador não-paramétrico de Kaplan-Meier, proposto por Kaplan & Meier (1958), também chamado de estimador limite-produto. Na ausência de censura, é definido como

$$\widehat{S}(y) = \frac{\text{número de observações que não falharam até o tempo } y}{\text{número total de observações no estudo}}. \quad (1.8)$$

Considere as seguintes definições:

- $y_1 < y_2 < \dots < y_k$ ,  $k$  tempos distintos e ordenados de falha
- $d_j$  o número de falhas em  $y_j$ ,  $j = 1, \dots, k$
- $n_j$  o número de indivíduos sob risco em  $y_j$ , ou seja, os indivíduos que não foram censurados até o instante imediatamente anterior a  $y_j$ .

O estimador de Kaplan-Meier é definido como

$$\widehat{S}(y) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j}\right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right). \quad (1.9)$$

A função de verossimilhança pode ser escrita como

$$L(S(\cdot)) = \prod_{j=0}^k \left\{ [S(y_j) - S(y_{j+})]^{d_j} \prod_{l=1}^{m_j} S(y_{jl+}) \right\} \quad (1.10)$$

onde  $S(y_{j+}) = \lim_{\Delta y \rightarrow 0^+} S(y_j + \Delta y)$ . O termo  $S(y_j) - S(y_{j+})$  representa a probabilidade de falha no tempo  $y_j$ . E o termo  $S(y_{jl+})$  é a contribuição do tempo de sobrevivência censurado em  $y_{jl}$ , para  $l = 1, \dots, m_j$ , onde  $m_j$  é o número de observações censuradas no intervalo  $[y_j, y_{j+1})$ .

Pode-se mostrar que a função de sobrevivência  $S(y)$  que maximiza  $L(S(\cdot))$  é a expressão (1.9), Kaplan-Meier (1958).

## 1.4 Teorema de Bayes

Considere uma quantidade de interesse desconhecida  $\theta$ . A informação sobre  $\theta$ , resumida probabilisticamente através de  $\pi(\theta)$ , pode ser aumentada observando uma quantidade aleatória  $X$  relacionada com  $\theta$ . A distribuição amostral  $\pi(x|\theta)$  define esta relação. A idéia de que após observar  $X = x$ , a quantidade de informação sobre  $\theta$  aumenta, é bastante intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação,

$$\pi(\theta|x) = \frac{\pi(\theta, x)}{\pi(x)} = \frac{\pi(x|\theta) \pi(\theta)}{\pi(x)} = \frac{\pi(x|\theta) \pi(\theta)}{\int \pi(\theta, x) d\theta}. \quad (1.11)$$

O termo  $1/\pi(x)$ , não depende de  $\theta$  e funciona como uma constante normalizadora de  $\pi(\theta|x)$ .

A função  $l(\theta; x) = \pi(x|\theta)$ , para um valor fixo de  $x$ , fornece a função de verossimilhança de cada dos possíveis valores de  $\theta$ , enquanto  $\pi(\theta)$  é chamada distribuição a *priori* de  $\theta$ . Estas duas fontes de informação, priori e verossimilhança, são combinadas levando à distribuição a *posteriori*. Assim, a forma usual do teorema de Bayes é

$$\pi(\theta|x) \propto l(\theta; x) \pi(\theta) \quad (1.12)$$

Em outras palavras tem-se

Distribuição a posteriori  $\propto$  Verossimilhança  $\times$  Distribuição a priori.

Ao omitir o termo  $\pi(x)$ , a igualdade em (1.11) é substituída por uma proporcionalidade.

## 1.5 Métodos MCMC

Os métodos de Monte Carlo via Cadeias de Markov (MCMC) são uma alternativa aos métodos não interativos em problemas complexos. A idéia é obter uma amostra da distribuição a posteriori e calcular estimativas amostrais de características desta distribuição (Ehlers, 2001).

### 1.5.1 Cadeias de Markov

A cadeia de Markov é um processo estocástico  $\{X_0, X_1, \dots\}$  tal que a distribuição de  $X_t$ , dados todos os valores anteriores  $X_0, \dots, X_{t-1}$ , depende apenas de  $X_{t-1}$ . Matematicamente,

$$P(X_t \in A | X_0, \dots, X_{t-1}) = P(X_t \in A | X_{t-1})$$

para qualquer subconjunto  $A$ .

Os métodos MCMC requerem que a cadeia seja,

- homogênia, isto é, as probabilidades de transição de um estado para outro são invariantes;
- irredutível, isto é, cada estado pode ser atingido a partir de qualquer outro em um número infinito de interações e
- aperiódica, isto é, não haja estados absorventes.

Uma questão importante de ordem prática é como os valores iniciais influenciam o comportamento da cadeia. Conforme o número de interações aumenta, a cadeia gradualmente converge para uma distribuição de equilíbrio. Assim, em aplicações práticas é comum que as interações iniciais sejam descartadas, estes valores são chamados de *burnin*.

### 1.5.2 Algoritmo Metropolis-Hastings

A idéia dos algoritmos de Metropolis-Hastings é gerar um valor de uma distribuição auxiliar e aceitar este valor com uma dada probabilidade. Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio, que do ponto de vista *Bayesiano* é a distribuição a *posteriori*.

Suponha que a cadeia esteja no estado  $\theta$  e um valor  $\theta^*$  é gerado de uma distribuição proposta  $q(\cdot|\theta)$ . A distribuição proposta pode depender do estado atual da cadeia, por exemplo  $q(\cdot|\theta)$  poderia ser uma distribuição normal centrada em  $\theta$ . O novo valor  $\theta^*$  é aceito com probabilidade

$$\alpha(\theta, \theta^*) = \min \left( 1, \frac{\pi(\theta^*) q(\theta|\theta^*)}{\pi(\theta) q(\theta^*|\theta)} \right), \quad (1.13)$$

onde  $\pi$  é a distribuição de interesse.

Uma característica importante é que o conhecimento sobre  $\pi$  pode ser parcial, isto é, a menos de uma constante já neste caso a probabilidade (1.13) não se altera. Isto é fundamental em aplicações Bayesianas onde não se conhece completamente a posteriori. Outra observação é que a cadeia pode permanecer no mesmo estado por muitas interações e na prática costuma-se monitorar isto calculando a porcentagem média de interações para as quais novos valores são aceitos. Em termos práticos, o algoritmo de Metropolis-Hastings pode ser especificado pelo seguintes passos:

1. Inicializar o contador de interações  $j = 0$  e especificar um valor inicial  $\theta^{(0)}$
2. Gerar um novo valor  $\theta^*$  da distribuição  $q(\cdot|\theta)$
3. Calcular a probabilidade de aceitação  $\alpha(\theta, \theta^*)$  e gerar  $u \sim U(0, 1)$
4. Se  $u \leq \alpha$  então aceitar o novo valor e faça  $\theta^{(j+1)} = \theta^*$ , caso contrário rejeitar e fazer  $\theta^{(j+1)} = \theta$ .
5. Incrementar o contador de  $j$  para  $j + 1$  e voltar ao passo 2.

Um caso particular é quando a distribuição proposta não depende do estado atual da cadeia, isto é,  $q(\theta^*|\theta) = q(\theta^*)$ . Em geral,  $q(\cdot)$  deve ser uma boa aproximação de  $\pi(\cdot)$ , mas é mais seguro se  $q(\cdot)$  tiver caudas mais pesadas do que  $\pi(\cdot)$ .

Outro caso é chamado algoritmo de Metropolis e considera apenas propostas simétricas, isto é,  $q(\theta^*|\theta) = q(\theta|\theta^*)$  para todos os valores de  $\theta$  e  $\theta^*$ . Neste caso, a probabilidade de aceitação se reduz para

$$\alpha(\theta, \theta^*) = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta)} \right).$$

Um algoritmo de Metropolis muito utilizado é baseado em um passeio aleatório de modo que a probabilidade da cadeia mover-se de  $\theta$  para  $\theta^*$  depende apenas da distância entre eles, isto é,  $q(\theta^*|\theta) = q(|\theta - \theta^*|)$ . No entanto, no uso de uma distribuição proposta com variância  $\sigma^2$  duas situações extremas podem ocorrer:

- se  $\sigma^2$  for muito pequena os valores gerados estarão próximos do valor atual e quase sempre serão aceitos. Porém levará muitas interações até o algoritmo cobrir todo o espaço do parâmetro;
- valores grandes de  $\sigma^2$  levam a uma taxa de rejeição excessivamente alta e a cadeia se movimenta muito pouco.

Em ambas situações o algoritmo fica ineficiente e na prática deve-se tentar vários valores de  $\sigma^2$ .

### 1.5.3 Amostrados de Gibbs

No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, isto é, não existe mecanismo de aceitação-rejeição. As transições de estado para outro são feitas de acordo com as distribuições condicionais completas  $\pi(\theta_i|\theta_{-i})$ , onde  $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)'$ .

Em geral, cada uma das componentes  $\theta_i$  podem ser uni ou multidimensional. Portanto, a distribuição condicional completa é a distribuição da  $i$ -ésima componente de  $\theta$  condicionada em todas as outras componentes. Ela é obtida a partir da distribuição conjunta como,

$$\pi(\theta_i|\theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}.$$

Em muitas situações, a geração de uma amostra diretamente de  $\pi(\theta)$  pode ser custosa, complicada ou simplesmente impossível. Mas se as distribuições condicionais completas forem completamente conhecidas, então o amostrador de Gibbs é definido pelos seguintes passos:

- Inicializar o contador de interações da cadeia  $j=0$ ;
- Especificar os valores iniciais  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})'$ ;
- Obter um novo valor de  $\theta^{(j)}$  a partir de  $\theta^{(j-1)}$  através da geração sucessiva dos valores

$$\theta_1^{(j)} \sim \pi(\theta_1|\theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$$

$$\theta_2^{(j)} \sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$$

$$\vdots$$

$$\theta_p^{(j)} \sim \pi(\theta_p|\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j)})$$

- Incrementar o contador de  $j$  para  $j + 1$  e retorne ao passo 2 até obter convergência.

Assim cada interação se completa a cada  $p$  movimentos ao longo dos eixos coordenados das componentes de  $\theta$ . Após a convergência, os valores resultantes formam uma amostra de  $\pi(\theta)$ . Vale notar que, mesmo em problemas de grandes dimensões todas as simulações podem ser univariadas, o que em geral é uma vantagem computacional.

Note também que o amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, no qual os elementos de  $\theta$  são atualizados um de cada vez (ou em blocos), tomando a distribuição condicional completa como proposta e probabilidade de aceitação igual a 1.

### 1.5.4 Diagnósticos de Convergência

Para verificar convergência do algoritmo pode-se utilizar:

- (a) Gráficos de séries temporais;

A linha deste gráfico deve estar sempre em torno de uma faixa, não apresentando muitas oscilações.

- (b) Gráficos de autocorrelação;

O gráfico de autocorrelação dos parâmetros somente deve ter a primeira correlação alta sendo as demais próximas de zero, indicando assim que as amostras geradas são não-correlacionadas.

- (c) Índices de convergência, como o índice de Gelman & Rubin (1992).

Gelman e Rubin utilizam procedimentos de análise de variância (ANOVA) para verificar a convergência de cadeias paralelas (com valores iniciais diferentes). A idéia é comparar as variabilidades "entre" e "dentro" das cadeias geradas. O método funciona da seguinte maneira:

Simular  $m \geq 2$  cadeias, cada uma de comprimento  $2n$ , descartando-se as  $n$  primeiras amostras;

Calcular a variabilidade entre cadeias

$$\frac{U}{n} = \sum_{i=1}^m (\bar{\theta}_i - \bar{\theta}_{..})^2 / (m - 1)$$

em que  $\bar{\theta}_i$  é a média das  $n$  amostras geradas na  $i$ -ésima cadeia;  $\bar{\theta}_{..}$  é a média geral.

Define-se, também, a variabilidade dentro de cada cadeia,

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2$$

em que  $s_i^2 = \frac{1}{n-1} \sum_{j=1}^n (\theta_{ij} - \bar{\theta}_i)^2$ ,  $i = 1, \dots, m$ .

Assim, a variância da distribuição estudada (*posteriori*) pode ser estimada por:

$$\hat{\sigma}^2 = \frac{n-1}{n}W + \frac{1}{n}U$$

Gelman e Rubin mostraram que a distribuição de  $\boldsymbol{\theta}$  dado  $y$  segue distribuição  $t$  de Student com centro em  $\hat{\mu} = \bar{\theta}_{..}$ , desvio padrão  $\sqrt{\hat{V}} = \sqrt{\hat{\sigma}^2 + \frac{U}{mn}}$  e  $\nu = \frac{2\hat{V}}{\text{var}(\hat{V})}$  graus de liberdade, em que:

$$\begin{aligned} \text{var}(\hat{V}) &= \left(\frac{n-1}{n}\right)^2 \frac{1}{m} \text{var}(s_i^2) + \left(\frac{m+1}{mn}\right)^2 \frac{2}{m-1} B^2 + \\ &+ 2 \frac{(m-1)(n-1)}{mn^2} \frac{n}{m} [\text{cov}(s_i^2, \bar{u}_i^2) - 2\bar{u}_{..} \text{cov}(s_i^2, \bar{u}_i)] \end{aligned}$$



Estima-se, então, o fator de redução de escala como:

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}}{W} \frac{\nu}{\nu - 2}} \quad (1.14)$$

Quando  $n$  é grande, o fator de redução pode ser simplificado para  $\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}}{W}}$ . Nota-se, também, que  $\sqrt{\hat{R}}$  decresce para 1 quando  $n \rightarrow \infty$ . Em geral, se  $\sqrt{\hat{R}} < 1,2$ , tem-se convergência das cadeias.



## Capítulo 2

---

# Teoria Unificada para Modelos de Longa-Duração

---

### 2.1 Introdução

Os modelos de sobrevivência de longa-duração têm grande importância em análise de sobrevivência e confiabilidade por isso diferentes métodos para ajustar tais modelos têm surgido na literatura. Devido as suas diversificações, esses modelos são aplicados em áreas como biomedicina, finanças, criminologia, confiabilidade industrial, entre outros. No momento, o interesse está em dados biomédicos, cujo evento de interesse pode ser a morte de pacientes, que pode ocorrer devido a diferentes causas competitivas ou a recorrência do tumor devido a presença de um número desconhecido de células cancerígenas. Para mais detalhes ver Yakovlev *et al.* (1993). Em confiabilidade industrial, o evento de interesse pode ser a falha de placas de circuito devido a diferentes fatores de risco ou por desgaste de uso, Meeker & Escobar (1998). Em dados financeiros, o evento pode ser o desligamento do cliente devido a várias causas, Hoggart & Griffin (2000).

Nesse capítulo é apresentado a metodologia proposta por Rodrigues *et al.* (2008) usando a função geradora de uma sequência de números reais definida por Feller (1967). Como consequência temos a unificação dos modelos de sobrevivência de longa-duração proposto por Berkson & Gage (1952), Yin & Ibrahim (2005) e outros. Também é mostrado que a função geradora de longa-duração formulada satisfaz a propriedade de risco proporcional se, e somente se, o número de causas relacionado a ocorrência do evento de interesse segue uma distribuição de Poisson (Rodrigues *et al.* 2008).

Seja  $N$  uma variável aleatória denotando o número de causas competindo para à ocorrência do evento de interesse com distribuição de probabilidade dada por

$$p_n = P[N = n], \quad n = 0, 1, 2, \dots$$

essa variável é não observada com distribuição de probabilidade previamente especificada

$\{p_n\}$ . Seja  $Z_i$  a variável aleatória denotando o tempo de ocorrência do evento de interesse devido a causa  $i$ . Dado  $N = n$ ,  $Z_i, i = 1, \dots, m$  é a variável aleatória independente com função distribuição comum  $F(\cdot) = 1 - S(\cdot)$ .

No entanto, no cenário de competição, apenas o menor tempo de vida  $Z_i$  entre todas as causas é observado (Cox & Oakes, 1984). Dessa forma, na suposição de inclusão de indivíduos não susceptíveis ao evento de interesse, o tempo para a ocorrência do evento de interesse é definido como

$$Y = \min\{Z_0, Z_1, \dots, Z_N\}, \quad (2.1)$$

onde  $P[Z_0 = \infty] = 1$ , proporcionando uma razão  $p_0$  da população não sujeita a ocorrência do evento. Seja  $S(y)$  a função de sobrevivência de  $Y$ .

O capítulo é organizado da seguinte forma. Na Seção 2.2, é mostrado a origem e o desenvolvimento de toda teoria unificada. Na Seção 2.3, tem-se a descrição da função de verossimilhança unificada. Em seguida, na Seção 2.4, é apresentado a abordagem clássica da teoria usando algoritmo EM para obtenção dos estimadores e na Seção 2.5 uma abordagem bayesiana.

## 2.2 Função Geradora de Sobrevivência de Longa-Duração

A função de sobrevivência da variável aleatória  $Y$  definida em 2.1, é a probabilidade de um indivíduo sobreviver a um tempo  $y$  e é representada por

$$S_p(y) = P(Y > y). \quad (2.2)$$

A seguir é apresentado a função geradora de sequência de números reais definida por Feller (1967) que possibilitou o desenvolvimento e unificação da teoria análise de sobrevivência de longa-duração.

**Definição 1.** *Seja  $\{a_n\}$  uma sequência de números reais. Se*

$$A(s) = a_0 + a_1s + a_2s^2 + \dots \quad (2.3)$$

*converge no intervalo  $0 \leq s \leq 1$ , então  $A(s)$  é definida como função geradora da sequência  $\{a_n\}$ .*

A partir desta definição segue os teoremas (Rodrigues *et al.* 2008).

**Teorema 1.** *Dada uma função de sobrevivência própria  $S(y)$ , a função de sobrevivência da variável aleatória  $Y$  dada por (2.2) pode se escrita como*

$$S_p(y) = A_p[S(y)] = \sum_{n=0}^{\infty} p_n [S(y)]^n, \quad (2.4)$$

onde  $A_p(\cdot)$  é a função geradora da sequência  $\{p_n\}$  que converge no intervalo  $0 \leq S(y) \leq 1$ .  $S_p(y)$  é chamada função geradora de longa-duração

**Demonstração:** Podemos obter diretamente esse resultado escrevendo (2.2) da seguinte forma

$$\begin{aligned} S_p(y) &= P(N = 0) + P(Z_1 > y, Z_2 > y, \dots, Z_N > y, N \geq 1) = \\ &= P(N = 0) + P(N \geq 1)P(Z_1 > y|N \geq 1) \dots P(Z_N > y|N \geq 1) = \\ &= p_0 + \sum_{n=1}^{\infty} p_n [S(y)]^n = A_p[S(y)]. \end{aligned} \tag{2.5}$$

■

A função de sobrevivência  $S(y)$  explica como os indivíduos se comportam no tempo. Já  $A_p[S(y)]$ , é um modelo de dois estágios. O primeiro chamado de estágio de iniciação, modela probabilisticamente o número de fatores de risco ou causas. E o segundo estágio, chamado de maturação, modela o tempo de maturação dos fatores de riscos (clonogenes).

A partir deste teorema é possível obter qualquer função geradora de longa-duração apenas pelo conhecimento da função geradora de sequência  $A(\cdot)$  da variável latente  $N$  relacionada a  $S(y)$ . E ainda, se  $a = \{a_n\} = \{p_n\} = p$  então  $A_p(s) = E[s^N] = E[\exp(N \log(s))]$ , ou seja,  $A_p(s)$  é uma função geradora de momentos da variável aleatória  $N$  no ponto  $\log(s)$ .

Como exemplo, considere as seguintes situações (ver Feller, 1967):

- Se  $N \sim \text{Poisson}(\theta)$  temos  $A_p(s) = \exp\{-\theta(1 - s)\}$
- Se  $N \sim \text{Binomial}(r, \theta)$ , para  $0 \leq \theta \leq 1$ , ou seja,  $p_n = \binom{r}{n} \theta^n (1 - \theta)^{r-n}$ , com  $n = 0, 1, \dots, r$ , então  $A_p(s) = [1 - \theta + \theta s]^r$
- Se considerar  $N \sim \text{Binomial Negativa}(r, \theta)$  temos  $A_p(s) = \left( \frac{\theta}{1 - (1 - \theta)s} \right)^r$

Os dois primeiros itens serão apresentados nos capítulos 3 e 4, respectivamente.

Pode-se observar que a função de sobrevivência (2.4) não é própria, isto é,  $\lim_{y \rightarrow \infty} S_p(y) > 0$ , como mostra o teorema a seguir.

**Teorema 2.** Dada uma função de sobrevivência própria  $S(y)$ , então

$$\lim_{y \rightarrow \infty} S_p(y) = P(N = 0) = p_0, \tag{2.6}$$

onde  $p_0$  denota a proporção da não ocorrência do evento na população.

**Demonstração:**

$$\begin{aligned} \lim_{y \rightarrow \infty} S_p(y) &= \lim_{y \rightarrow \infty} \left[ \sum_{n=0}^{\infty} p_n S^n(y) \right] = \lim_{y \rightarrow \infty} \left[ p_0 + \sum_{n=1}^{\infty} p_n S^n(y) \right] \\ &= \lim_{y \rightarrow \infty} p_0 + \lim_{y \rightarrow \infty} \sum_{n=1}^{\infty} p_n S^n(y) = p_0 + \sum_{n=1}^{\infty} p_n \lim_{y \rightarrow \infty} S^n(y) \end{aligned} \quad (2.7)$$

Como  $0 \leq S(y) \leq 1$  e  $S(y)$  é uma função própria ( $\lim_{y \rightarrow \infty} S(y) = 0$ ), segue então que

$$\lim_{y \rightarrow \infty} S_p(y) = P(N = 0) = p_0$$

■

Deste teorema interpreta-se  $p_0$  como uma fração da população que não sofre o evento de interesse e por isso  $p_0$  é definido como a proporção de longa-duração e  $S_p(y)$  como função geradora de longa-duração.

No teorema a seguir é mostrado que a função geradora de longa-duração pode ser escrita na mesma forma do modelo de Berkson & Cage (1952) dado por  $S_1(y) = p_0 + (1 - p_0)S^*(y)$ , onde  $S^*(y) = P(Y > y | N = 1)$  é uma função de sobrevivência própria.

**Teorema 3.** *A função geradora de sobrevivência de longa-duração é dada por*

$$S_p(y) = p_0 + (1 - p_0)S_p^*(y), \quad (2.8)$$

onde

$$S_p^*(y) = \sum_{n=1}^{\infty} p_n^* S^n(y) \quad e \quad p_n^* = \frac{p_n}{1 - p_0}$$

**Demonstração:**

Usando o resultado do Teorema 1 tem-se,

$$S_p(y) = p_0 + \sum_{n=1}^{\infty} p_n S^n(y) = p_0 + (1 - p_0) \frac{\sum_{n=1}^{\infty} p_n S^n(y)}{(1 - p_0)}$$

Pode-se observar que a série  $\sum_{n=1}^{\infty} \frac{p_n S^n(y)}{(1 - p_0)}$  é convergente, logo a função dada por este termo é própria. Seja

$$S_p^*(y) = \sum_{n=1}^{\infty} p_n^* S^n(y),$$

onde  $p_n^* = p_n / (1 - p_0)$ . Portanto, pode-se escrever a seguinte expressão

$$S_p(y) = p_0 + (1 - p_0)S_p^*(y).$$

■

Observa-se que  $S_p^*(y)$  é uma função de sobrevivência própria associada aos indivíduos em risco na população. Além disso,  $S_p^*(y)$  é uma função geradora de distribuição truncada no zero,  $\{p_n^*\}$ .

Uma vez obtido a função geradora de longa-duração por (2.4), é possível obter as funções densidade de probabilidade e a função risco da variável  $Y$ . A função densidade associada a função geradora de longa-duração é dada por

$$f_p(y) = -\frac{dS_p(y)}{dt} = f(y) \left( \frac{dA_p(s)}{ds} \Big|_{s=S(y)} \right) \quad (2.9)$$

onde  $A_p(\cdot)$  é a função geradora de  $\{p_n\}$  e é chamada de sub-função densidade de longa-duração. Assim, a sub-função de risco de longa-duração é

$$h_p(y) = \frac{f_p(y)}{S_p(y)} = f(y) \frac{\frac{dA_p(s)}{ds} \Big|_{s=S(y)}}{S_p(y)}. \quad (2.10)$$

Além disso, é possível obter de (2.8), as funções densidade probabilidade e risco da população em risco, ou seja, associadas a função de sobrevivência própria  $S_p^*(y)$  em (2.8), são dadas por

$$f^*(y) = -\frac{dS_p^*(y)}{dy} = f(y) \frac{\frac{dA_p(s)}{ds} \Big|_{s=S(y)}}{1 - p_0} \quad (2.11)$$

$$h^*(y) = \frac{f^*(y)}{S_p^*(y)} = \frac{S_p(y)}{S_p(y) - p_0} h_p(y), \quad (2.12)$$

respectivamente.

Pode-se observar que a sub-função de risco dada em (2.10) não apresenta propriedade de risco proporcional na presença de covariáveis, uma vez que o termo

$$\frac{\frac{dA_p(s)}{ds} \Big|_{s=S(y)}}{S_p(y)} \quad (2.13)$$

depende de  $y$ . Uma exceção acontece quando a variável aleatória  $N$  assume uma distribuição de Poisson com parâmetro  $\theta$  (Chen, Ibrahim & Sinha, 1999). A caracterização de Poisson em termos de risco proporcional é apresentado no seguinte teorema.

**Teorema 4.** *A sub-função de risco de longa-duração  $h_p(y)$ , tem propriedade de risco proporcional se, e somente se,  $p_n = \frac{\theta^n e^{-\theta}}{n!}$ , para  $n = 0, 1, 2, \dots$ , ou seja,  $N \sim \text{Poisson}(\theta)$ .*

**Demonstração:** Verifica-se que  $\{p_n\}$  satisfaz risco proporcional se, e somente se,

$$p_n = \frac{(n+1)p_{n+1}}{E(N)}, \quad \text{para } n=0,1,2,\dots \quad (2.14)$$

(ver Apêndice D).

Assim, se  $p_n$  é uma função de probabilidade de Poisson, ou seja,  $p_n = \frac{\theta^n e^{-\theta}}{n!}$ , tem-se que a sub-função de risco de longa-duração,  $h_p(y)$ , satisfaz a expressão (2.14) e

$$p_n = \frac{E(N)^n p_0}{n!} \quad \text{para } n=0,1,2,\dots$$

e  $p_0 = \exp(-E(N))$ , ou seja,  $N$  tem distribuição de Poisson com parâmetro  $\theta = E(N)$ . ■

No contexto de risco competitivo, esse teorema mostra que a sub-função de risco assume, na presença de covariáveis, a importante propriedade de risco proporcional no caso da variável aleatória  $N$  assumir distribuição de Poisson.

## 2.3 Função de Verossimilhança

Para a formulação da função de verossimilhança considera-se as seguintes notações. Seja  $N_i$ , o número de causas ou riscos relacionado a ocorrência do evento de interesse, no  $i$ -ésimo indivíduo,  $i = 1, 2, \dots, m$ , variável aleatória independente não observadas com distribuição de probabilidade  $p_{\theta_i}(n_i) = P_{\theta_i}(N_i = n_i)$ , onde  $E(N_i) = \theta_i > 0$ .

Sejam  $Z_{i1}, Z_{i2}, \dots, Z_{iN_i}$  variáveis aleatórias independentes representando o tempo de ocorrência do evento de interesse para as  $N_i$  causas no  $i$ -ésimo elemento da amostra. Assume-se que os  $Z'_i$ 's são variáveis aleatórias identicamente distribuídas com função de distribuição acumulada  $F(\cdot|\psi) = 1 - S(\cdot|\psi)$ , onde  $\psi$  é o vetor de parâmetros.

O tempo de ocorrência do evento de interesse no  $i$ -ésimo indivíduo denota-se por  $Y_i$  e é definido como

$$Y_i = \min\{Z_{i0}, Z_{i1}, Z_{i2}, \dots, Z_{iN_i}\}.$$

Essa variável é observada com censura à direta e cuja indicação é dado pela variável

$$\delta_i = \begin{cases} 1 & \text{se indivíduo é falha} \\ 0 & \text{se indivíduo é censura} \end{cases}$$

Além disso, no caso em que considera-se  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  o vetor de covariáveis associada a cada indivíduo  $i$  da amostra, então a média de  $N_i$  é definida por uma função de ligação  $g(\theta_i) = x'_i \beta$ , onde  $\beta = (\beta_1, \dots, \beta_p)'$  é o vetor de coeficientes de regressão.

Os dados observados são denotados por  $D_{obs} = (m, \mathbf{X}, \mathbf{y}, \delta)$ , onde  $\mathbf{X}$  é a matriz de covariáveis  $m \times p$ . E os dados ampliados por  $D = (m, \mathbf{X}, \mathbf{y}, \delta, \mathbf{N})$ , onde  $\mathbf{N}$  é o vetor de variáveis latente.



Seja  $\phi = (\theta, \psi)$  o vetor de parâmetros associados ao modelo, onde  $\theta$  é o parâmetro da variável  $N$ . A função de verossimilhança para  $\phi$  é dada por

$$L(\phi; D) = \prod_{i=1}^m S(y_i|\psi)^{N_i - \delta_i} (N_i f(y_i|\psi))^{\delta_i} p_{\theta_i}(n_i). \quad (2.15)$$

Aplicando o logaritmo, tem-se

$$l(\phi; D) = \sum_{i=1}^m \left\{ N_i \log(S(y_i|\psi)) + \delta_i \log(N_i) + \delta_i \log\left(\frac{f(y_i|\psi)}{S(y_i|\psi)}\right) + \log(p_{\theta_i}(n_i)) \right\}. \quad (2.16)$$

(ver Apêndice A).

A vantagem dessa função de verossimilhança é a facilidade de obtenção de várias expressões da mesma para diferentes distribuições de  $N$ .

Além disso, foi obtida uma função de verossimilhança unificada cuja expressão é semelhante a função de verossimilhança usual da Análise de Sobrevivência na presença de censura.

### 2.3.1 Função de Verossimilhança Unificada

A função de verossimilhança unificada é definida como sendo

$$L_u(\phi; D) = \sum_{\mathbf{n}} \prod_{i=1}^m S(y_i|\psi)^{n_i - \delta_i} (n_i f(y_i|\psi))^{\delta_i} p_{\theta_i}(n_i), \quad (2.17)$$

ou seja, pode-se dizer que  $L_u(\phi; D)$  é uma verossimilhança marginal de  $L(\phi; D)$ .

Pode-se escrevê-la de uma forma mais simples (Rodrigues *et al.* 2008), como

$$\begin{aligned} L_u(\phi; D_{obs}) &= \prod_{i=1}^m [f_p(y_i|\phi)]^{\delta_i} [S_p(y_i|\phi)]^{1-\delta_i} \\ &= \prod_{i \in \bar{C}} f_p(y_i|\phi) \prod_{i \in C} S_p(y_i|\phi), \end{aligned} \quad (2.18)$$

onde  $\bar{C}$  é o conjunto dos indivíduos não censurado e  $C$  o conjunto do indivíduos censurado. A prova desse resultado está no Apêndice B.

Considerando  $S_p(y|\phi)$  em (2.8) e substituindo em (2.18) tem-se,

$$L_u(\phi; D_{obs}) = \prod_{i \in \bar{C}} f_p(y_i|\phi) \prod_{i \in C} [p_0 + (1 - p_0)S_p^*(y_i|\phi)]. \quad (2.19)$$

Logo, utiliza-se a teoria de dados ampliados introduzindo a variável latente  $W_i \sim \text{Bernoulli}(p_i)$ , dada abaixo, para simplificar a função de verossimilhança (2.19) e facilitar a estimação dos parâmetros.

$$W_i = \begin{cases} 1 & \text{com } p_i = \frac{p_0}{p_0 + (1-p_0)S_p^*(y_i|\phi)} \\ 0 & \text{com } 1 - p_i. \end{cases}$$

Observe que a variável  $W_i$  está associada apenas ao indivíduo censurado, ou seja, os pacientes são diferenciados em censurados por cura ou censurados por algum motivo desconhecido. Assim, pode-se escrever (2.19) como

$$\begin{aligned} L_A(\phi; D) &= \prod_{i=1}^m [f_p(y_i|\phi)]^{\delta_i} \left( p_0^{W_i} [(1-p_0)S_p^*(y_i|\phi)]^{1-W_i} \right)^{1-\delta_i} \\ &= p_0^W (1-p_0)^{m-r-W} \prod_{i=1}^m [f_p(y_i|\phi)]^{\delta_i} [S_p^*(y_i|\phi)^{(1-W_i)}]^{1-\delta_i}, \end{aligned} \quad (2.20)$$

onde  $W = \sum_{i \in C} W_i$  e  $D = (\mathbf{W}, \mathbf{y}, \delta, m, \mathbf{X})$ . Com isso pode-se utilizá-la tanto na Inferência Bayesiana, com algoritmo MCMC, quanto na Clássica com algoritmo EM.

Essa nova verossimilhança, do ponto de vista computacional, é uma alternativa da expressão (2.15) com o intuito de facilitar a implementação de algoritmos.

## 2.4 Inferência Clássica

No contexto clássico, a estimação dos parâmetros mais comum é realizada maximizando a função log-verossimilhança.

No entanto, para obter inferência dos parâmetros neste trabalho, utiliza-se o algoritmo EM a partir do logaritmo da função de verossimilhança (2.20),

$$\begin{aligned} l_A(\phi; D) &= \sum_{i=1}^m \left\{ \delta_i \log(f_p(y_i|\phi)) + (1-\delta_i)(1-W_i) \log(S_p^*(y_i|\phi)) + \right. \\ &\quad \left. + \sum_{i=1}^m (1-\delta_i)(1-W_i) \log(1-p_0) + (1-\delta_i)W_i \log(p_0) \right\}. \end{aligned} \quad (2.21)$$

No passo E do algoritmo deve-se calcular a esperança condicional da log-verossimilhança dos dados  $D$ ,

$$E [l_A(\phi; D) | \phi^{(k)}], \quad (2.22)$$

onde  $\phi^{(k)}$  são as estimativas do vetor de parâmetros na  $k$ -ésima interação do algoritmo.

Observa-se que para calcular a esperança dada por (2.22), basta calcular a esperança condicional de  $W_i$ . Calculando a distribuição condicional de  $W_i$  dado  $D$  (ver Apêndice C), tem-se

$$E [W_i | D; \phi^{(k)}] = \frac{p_0^{(1-\delta_i)}}{p_0^{(1-\delta_i)} + [(1-p_0)S_p^*(y_i|\phi)]^{(1-\delta_i)}}. \quad (2.23)$$

Assim, denota-se

$$W_i^{(k+1)} = E [W_i | D; \phi^{(k)}]$$

e na  $(k + 1)$ -ésima interação do algoritmo temos os seguintes passos:

**Passo E:** Cálculo da esperança condicional da log-verossimilhança dos dados  $D$ ,

$$\begin{aligned} E [l_A(\phi, D) | \phi^{(k)}] &= \sum_{i=1}^m \left\{ \delta_i \log(f_p(y_i | \phi^{(k)})) + (1 - \delta_i)(1 - W_i^{(k+1)}) \log(S_p^*(y_i | \phi^{(k)})) + \right. \\ &\quad \left. + (1 - \delta_i)(1 - W_i^{(k+1)}) \log(1 - p_0) + (1 - \delta_i)W_i^{(k+1)} \log(p_0) \right\} \end{aligned}$$

**Passo M:** Maximização das expressões

$$Q_1(\psi | p_0^{(k)}, \psi^{(k)}) \equiv Q_1(y, w, \psi)$$

e

$$Q_2(p_0 | p_0^{(k)}, \psi^{(k)}) \equiv Q_2(y, w, p_0),$$

sendo que  $Q_1$  e  $Q_2$  geram  $\psi^{(k+1)}$  e  $p_0^{(k+1)}$ , respectivamente. Observe que a expressão em  $Q_1$  contém apenas termos com o vetor de parâmetros  $\psi$  enquanto  $Q_2$ , contém apenas termos com parâmetros relacionados a fração de cura  $p_0$ .

Portanto, as estimativas de máxima verossimilhança  $\hat{\phi} = (\hat{\psi}, \hat{p}_0)$  são obtidas pela convergência do algoritmo considerando o critério de parada dado por

$$|\phi^{(k+1)} - \phi^{(k)}| < \epsilon.$$

**Observação:** O termo  $p_0$  em  $Q_2$  é usado como ilustração para representar o parâmetro relacionado a fração de cura. Essa mesma observação é válida para interpretar as idéias da seção seguinte.

## 2.5 Inferência Bayesiana

A inferência do vetor de parâmetros  $\phi = (\psi, p_0)$  do ponto de vista bayesiano é realizada utilizando a função de verossimilhança  $L_A(\phi; D)$  dada em (2.20).

Para isso considera-se a distribuição conjunta a *priori*, com  $\psi$  e  $p_0$  independentes, dada por

$$\pi(\psi, p_0) \propto \pi(p_0)\pi(\psi),$$

onde  $\psi$  é o vetor de parâmetros da distribuição  $f(y|\psi)$ .

Dessa forma, a distribuição a *posteriori* conjunta de  $(\psi, p_0, W)$  baseada nos dados  $D = (m, \mathbf{y}, \delta, \mathbf{W})$  é dada por

$$\begin{aligned} \pi(\psi, p_0, W|D) &\propto L_A(\psi, p_0; D) \pi(\psi, p_0) \\ &\propto \prod_{i=1}^m f_p(y_i|\phi)^{\delta_i} [p_0^{W_i}(1-p_0)^{1-W_i} S_p^*(y_i|\phi)^{1-W_i}]^{1-\delta_i} \pi(\psi, p_0) \end{aligned} \quad (2.24)$$

As estimativas a *posteriori* do vetor de parâmetros  $\hat{\phi} = (\hat{\psi}, \hat{p}_0)$  são obtidas pelo algoritmo MCMC via distribuições condicionais a *posteriori* obtidas a partir de (2.24). Algoritmos como por exemplo, Gibbs sampler ou Metropolis-Hastings podem ser usados.

Para um melhor entendimento das idéias clássicas e bayesianas apresentadas, são ilustrados nos capítulos seguintes casos particulares da teoria unificada.

## Capítulo 3

---

# Modelo de Fração de Cura com Poisson

---

### 3.1 Introdução

É normal em análise de dados envolvendo tempo de sobrevivência (ou falha), assumir que o evento de interesse ocorrerá em algum instante, para qualquer indivíduo da população em estudo. Isso é possível se o tempo de acompanhamento é suficientemente grande, ou seja, o tempo de sobrevivência é representado por uma variável aleatória não negativa  $Y$ , caracterizada por uma função de sobrevivência  $S(y) = P(Y \geq y)$  e em geral é tal que

$$S(\infty) = \lim_{y \rightarrow \infty} S(y) = 0.$$

Como consequência, a função de risco acumulado  $H(y) = -\log(S(y))$  é não limitada, isto é,

$$H(\infty) = \lim_{y \rightarrow \infty} H(y) = \infty.$$

No entanto, para um determinado grupo da população em estudo, o evento de interesse pode nunca ocorrer. Esse grupo é chamado de imune ou curado. Em tais casos, um dos interesses do pesquisador é a estimação da proporção de indivíduos curados, o que pode auxiliar na escolha de tratamentos a serem utilizados.

O interesse de alguns estudos médicos é o tempo de recorrência de uma determinada doença em pacientes em tratamento, mas parte desses pacientes podem ser considerados curados após o tratamento. Por exemplo, em alguns tratamentos de câncer a não recorrência da doença em um intervalo de 5 a 10 anos é considerado um indicativo de cura do indivíduo.

A existência de indivíduos curados na população é caracterizada pelo fato da função de sobrevivência convergir para um número positivo quando o tempo aumenta. Inicialmente, uma maneira de averiguar a existência da fração de cura em um conjunto de dados é

fazendo o gráfico da função de sobrevivência estimada de Kaplan-Meier e neste caso, a cauda à direita deverá apresentar um nível constante superior a zero em um período suficientemente grande.

Uma forma de modelar conjuntos de dados em que há a possibilidade de cura é considerar uma função de sobrevivência imprópria  $S_p$  para a população total e uma função de sobrevivência própria  $S_p^*$  para a parte da população de não curados. Isto é, considerar  $S_p(\infty) = \lim_{y \rightarrow \infty} S_p(y) > 0$  e  $S_p^*(\infty) = \lim_{y \rightarrow \infty} S_p^*(y) = 0$ .

Nesse contexto, o modelo desenvolvido matematicamente por Chen, Ibrahim & Sinha (1999), cuja motivação foi Yakovlev *et al.* (1993), é abordado neste capítulo.

O capítulo segue com as seguintes seções. Na Seção (3.3), é apresentado a função de verossimilhança obtida a partir da teoria unificada desenvolvida no primeiro capítulo. Na Seção (3.4), apresenta-se o algoritmo EM para estimar os parâmetros do modelo. Na seção seguinte, (3.5), apresenta-se a matriz de informação de Fisher e testes de hipóteses supondo a distribuição Weibull para variável tempo. Na Seção (3.6) apresenta-se uma abordagem Bayesiana.

## 3.2 Caso particular: Poisson

Yakovlev *et al.* (1993) introduz uma estrutura de riscos competitivos no modelo considerando:  $N$  uma variável aleatória com distribuição de Poisson com média  $\theta$ ;  $Z_1, \dots, Z_N$  variáveis aleatórias independentes e identicamente distribuídas denotando o tempo até o evento devido a causa  $v$ , com  $v = 1, \dots, N$ , e são independentes de  $N$  com função de distribuição acumulada e função de sobrevivência  $F(\cdot)$  e  $S(\cdot) = 1 - F(\cdot)$ , respectivamente;  $T$  o tempo de ocorrência do evento de interesse definido como  $T = \min\{Z_0, Z_1, \dots, Z_N\}$ , em que  $P(Z_0 = \infty) = 1$ .

Nessas circunstâncias, o resultado do Teorema 1 é usado para obter a função de sobrevivência da população total

$$S_p(t) = A_p(S(t)) = \sum_{n=0}^{\infty} \frac{\theta^n e^{-\theta}}{n!} S(t)^n = \exp\{-\theta F(t)\} \quad (3.1)$$

Observe que (3.1) converge para  $\exp(-\theta)$  quando o tempo aumenta e portanto,  $\exp(-\theta)$  corresponde a fração de cura da população.

De (3.1) obtém-se a função densidade de probabilidade e a função de risco dadas, respectivamente, por

$$f_p(t) = \theta f(t) \exp(-\theta F(t)), \quad t \geq 0$$

e

$$h_p(t) = \theta f(t).$$

Como a função de sobrevivência (3.1) é imprópria, verifica-se que a função densidade  $f_p(t)$  também é imprópria. Supor que uma função de  $\theta$ , denotada por  $g(\theta)$ , é igual a uma combinação linear de  $x'\beta$ , onde  $\mathbf{x}$  é um vetor de  $p$  covariáveis, ou seja,  $g(\theta) = x'\beta$ , tem-se que  $h_p(t) = g^{-1}(x'\beta) f(t)$ . Como exemplo, pode-se considerar  $g(\theta) = \log(\theta)$ , então tem-se  $\log(\theta) = x'\beta$ , logo  $\theta = \exp(x'\beta)$  e  $h_p(t) = \exp(x'\beta) f(t)$  caracteriza um modelo de risco proporcional.

Um modelo é caracterizado com risco proporcional quando considerando funções de risco de dois indivíduos quaisquer do estudo, a razão destas funções devem ser constantes, ou seja, não depende do tempo (Cox, 1972).

Também obtém-se a função de sobrevivência própria para a população não curada,

$$S^*(t) = P(T > t | N \geq 1) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}. \quad (3.2)$$

Conseqüentemente, a função densidade de probabilidade e a função de risco são

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

e

$$h^*(t) = \frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)} h_p(t),$$

respectivamente.

Para todos esses resultados, assume-se que a população é homogênea e os tempos de falha absolutamente contínuos. No caso em que é possível assumir heterogeneidades populacionais, deve-se incluir covariáveis no modelo através do parâmetro  $\theta$ . Uma maneira de introduzir covariáveis é através da relação  $\theta = \exp(x'\beta)$  (Chen, Ibrahim & Sinha, 1999), onde  $\beta$  é um vetor  $p$ -dimensional com os coeficientes de regressão associados a  $x$ . Assim, a fração de cura é relacionada as covariáveis pela expressão

$$S_p(\infty) = \exp(-\exp(x'\beta)).$$

Observe que as covariáveis influenciam na função de sobrevivência (3.2) e na função de risco da população dos não curados.

### 3.3 Função de Verossimilhança

Suponha  $m$  indivíduos e seja  $N_i$  uma variável aleatória independente e identicamente distribuída não observada com distribuição de Poisson ( $\theta$ ),  $i = 1, \dots, m$ , ou seja,  $p_{\theta_i}(n_i) = \theta^{N_i} e^{-\theta} / N_i!$ . Além disso, supõe-se  $Z_{i1}, \dots, Z_{iN_i}$  variáveis aleatórias independentes

e identicamente distribuídas, também não observáveis, com função de distribuição acumulada  $F(\cdot|\psi)$ . Seja  $T_i$  o tempo de falha para o indivíduo  $i$ , onde  $T_i$  pode ser censurado à direita. Assim,  $y_i$  é o tempo observado dado por  $y_i = \min\{T_i, C_i\}$ , com  $T_i = \min\{Z_{i0}, Z_{i1}, \dots, Z_{iN_i}\}$  e  $C_i$  o tempo de censura. Considerar  $\delta_i$  a variável indicadora de censura, como definida no primeiro capítulo. Seja  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  o vetor de covariáveis introduzidas no modelo pela relação  $\theta_i = \exp(x'_i\beta)$ , com  $\beta = (\beta_1, \dots, \beta_p)$  o vetor de coeficientes de regressão.

Os dados completos são denotado por

$$D_c = (m, \mathbf{y}, \mathbf{X}, \delta, \mathbf{N}),$$

onde

$$\begin{aligned} \mathbf{y} &= (y_1, \dots, y_m) \\ \delta &= (\delta_1, \dots, \delta_m) \\ \mathbf{N} &= (N_1, \dots, N_m) \\ \mathbf{X} &= \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{pmatrix}_{m \times p} \end{aligned}$$

Uma maneira de expressar a função de verossimilhança para o vetor de parâmetros  $\phi = (\psi, \beta)$ , é considerar (2.15) com a inclusão de covariáveis. Dessa forma, obtém-se

$$L(\phi; D_c) = \prod_{i=1}^m \left\{ S(y_i|\psi)^{N_i - \delta_i} (N_i f(y_i|\psi))^{\delta_i} \exp\{[N_i x'_i \beta - \log(N_i!) - \exp(x'_i \beta)]\} \right\}. \quad (3.3)$$

No entanto, essa função de verossimilhança não é observada pois depende da variável latente  $N$ . Por isso, uma forma observável é dada pela função de verossimilhança unificada (2.18), que substituindo a função densidade de probabilidade e a função geradora de longaduração mais adequadas obtém-se,

$$L_u(\phi; D) = \prod_{i=1}^m \left\{ [\exp(x'_i \beta) f(y_i|\psi)]^{\delta_i} \exp(-\exp(x'_i \beta) F(y_i|\psi)) \right\}. \quad (3.4)$$

A função log-verossimilhança de (3.3) é dada por (2.16) substituindo  $p_{\theta_i}(n_i)$  por uma função de probabilidade de Poisson ( $\theta$ ),

$$l(\phi, D_c) = \sum_{i=1}^m \left\{ N_i \log(S(y_i|\psi)) + \delta_i \log(N_i) + \delta_i \log\left(\frac{f(y_i|\psi)}{S(y_i|\psi)}\right) + [N_i x'_i \beta - \log(N_i!) - \exp(x'_i \beta)] \right\}. \quad (3.5)$$



A função log-verossimilhança de (3.4) é

$$l_u(\phi, D) = \sum_{i=1}^m \left\{ \delta_i x'_i \beta + \delta_i \log(f(y_i|\psi)) - \exp(x'_i \beta) F(y_i|\psi) \right\}. \quad (3.6)$$

### 3.4 Estimação dos Parâmetros do Modelo

Na função de verossimilhança unificada em (3.4), observa-se que não há necessidade do uso da teoria de dados ampliados por não apresentar um termo que dificulte a estimação dos parâmetros. Isso ocorre particularmente quando  $N$  tem distribuição de Poisson. Assim, é possível realizar a estimação por algum método numérico, por exemplo Newton-Raphson. Em casos que assume-se outra distribuição de probabilidade para  $N$  (Bernoulli, Binomial Negativa, etc), recomenda-se usar o método de estimação EM com dados ampliados apresentado na seção (2.4).

Uma alternativa de estimação seria usando a função de verossimilhança (3.3) que apresenta a variável latente  $N$ . Neste caso, o algoritmo EM é desenvolvido de uma forma similar apresentada na seção (2.4), Miozi (2004).

A partir do cálculo da distribuição condicional de  $N_i$  dado  $D$  (Apêndice C), obtém-se a distribuição de  $H_i + \delta_i$ , onde  $H_i$  é uma variável Poisson com  $E[H_i] = S(y_i|\psi) \exp(x'_i \beta)$ . Então,

$$E[N_i|D_c; \beta^{(k)}, \psi^{(k)}] = S(y_i|\psi^{(k)}) \exp(x'_i \beta^{(k)}) + \delta_i$$

e essa esperança condicional é denotada por  $N_i^{(k+1)}$ .

Portanto, o algoritmo EM é formado pelos passos:

**Passo E:** Cálculo da esperança condicional da log-verossimilhança (3.5)

$$\begin{aligned} E[l(\phi, D_c)|\beta^{(k)}, \psi^{(k)}] &= \sum_{i=1}^m \left\{ N_i^{(k+1)} x'_i \beta - \exp(x'_i \beta) \right\} \\ &+ \sum_{i=1}^m \left\{ N_i^{(k+1)} \log(S(y_i|\psi)) + \delta_i \log(h(y_i|\psi)) \right\} \end{aligned}$$

onde  $h(y_i|\psi) = \frac{f(y_i|\psi)}{S(y_i|\psi)}$ .

**Passo M:** Maximização das expressões:

$$Q_1(\beta|\beta^{(k)}, \psi^{(k)}) \equiv \sum_{i=1}^m \left\{ N_i^{(k+1)} x'_i \beta - \exp(x'_i \beta) \right\}$$

e

$$Q_2(\psi|\beta^{(k)}, \psi^{(k)}) \equiv \sum_{i=1}^m \left\{ N_i^{(k+1)} \log(S(y_i|\psi)) + \delta_i \log(h(y_i|\psi)) \right\}.$$

As estimativas de máxima verossimilhança  $\hat{\phi} = (\hat{\beta}, \hat{\psi})$  são obtidas pela convergência do algoritmo, considerando o critério de parada

$$|\phi^{(k+1)} - \phi^{(k)}| \leq \epsilon.$$

### 3.5 Modelo Paramétrico com Distribuição Weibull

Suponha que  $F$  seja uma função de distribuição acumulada Weibull com vetor de parâmetros  $\psi = (\alpha, \lambda)$ , ou seja,  $Z_{ik} \sim Weibull(\alpha, \lambda)$  com  $k = 1, \dots, N_i, i = 1, \dots, m$ . Como  $Y_i = \min(Z_{i1}, \dots, Z_{iN_i})$ , então tem-se a função densidade de probabilidade  $f(y|\alpha, \lambda) = \alpha y^{\alpha-1} \exp(\lambda - \exp(\lambda)y^\alpha)$  e função de sobrevivência  $S(y|\alpha, \lambda) = \exp(-y^\alpha \exp(\lambda))$ . Assim, a função log-verossimilhança (3.5) pode ser escrita como,

$$\begin{aligned} l(\phi; D_c) &= \sum_{i=1}^m \left\{ -N_i y_i^\alpha \exp(\lambda) + \delta_i \log(N_i \alpha y_i^{\alpha-1} \exp(\lambda)) \right\} \\ &+ \sum_{i=1}^m \left\{ N_i x'_i \beta - \log(N_i!) - \exp(x'_i \beta) \right\} \end{aligned} \quad (3.7)$$

e a função log-verossimilhança (3.6) é dada por

$$\begin{aligned} l_u(\phi; D) &= \sum_{i=1}^m \left\{ \delta_i [x'_i \beta + \lambda + \log(\alpha y_i^{\alpha-1}) - y_i^\alpha \exp(\lambda)] \right\} \\ &- \sum_{i=1}^m \left\{ \exp(x'_i \beta) [1 - \exp(-y_i^\alpha \exp(\lambda))] \right\}. \end{aligned} \quad (3.8)$$

As estimativas de máxima verossimilhança  $\hat{\phi} = (\hat{\beta}, \hat{\psi})$  podem ser obtidas utilizando o algoritmo EM apresentado na seção anterior.

Além disso, usar o fato que  $\hat{\phi}$  tem distribuição assintótica normal multivariada com média  $\phi$  e matriz de variância-covariância  $I^{-1}(\phi)$  é uma boa forma de calcular estimativas para a variância de  $\hat{\phi}$  e construir testes de hipóteses para os parâmetros com a matriz de informação de Fisher

$$I(\phi) = -E \left[ \frac{\partial^2 l(\phi; D)}{\partial \phi_i \partial \phi_j} \right]_{i,j=1,\dots,p+2}.$$

Pode-se observar que o cálculo da matriz de informação de Fisher  $I(\phi)$ , não é possível devido à presença de censuras. Como alternativa utiliza-se a matriz de informação observada de Fisher  $I(\hat{\phi})$ , que é uma estimativa de  $I(\phi)$ .

Logo, para o modelo (3.8) a matriz de Informação observada de Fisher toma a seguinte forma

$$I(\phi) = \begin{pmatrix} \frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_j^2} & \frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_j \partial \alpha} & \frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_j \partial \lambda} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 l_u(\phi; D_c)}{\partial \alpha^2} & \frac{\partial^2 l_u(\phi; D_c)}{\partial \alpha \partial \lambda} & \\ & & \frac{\partial^2 l_u(\phi; D_c)}{\partial \lambda^2} \end{pmatrix},$$

onde os elementos da matriz são dados pelas expressões (Miozi,2004),

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_k \partial \beta_j} = - \sum_{i=1}^m \left\{ x_{ij} x_{ik} e^{x'_i \beta} [1 - \exp(-y_i^\alpha e^\lambda)] \right\} \quad k, j = 1, \dots, p$$

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \alpha^2} = - \sum_{i=1}^m \left\{ \delta_i \left[ \frac{1}{\alpha} + e^\lambda y_i^\alpha (\log y_i)^2 \right] + \exp(x'_i \beta + \lambda - y_i^\alpha e^\lambda) y_i^\alpha (\log y_i)^2 [1 - y_i^\alpha e^\lambda] \right\}$$

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \lambda^2} = - \sum_{i=1}^m \left\{ y_i^\alpha \delta_i e^\lambda + \exp(x'_i \beta - y_i^\alpha e^\lambda) y_i^\alpha e^\lambda [1 - y_i^\alpha e^\lambda] \right\}$$

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_j \partial \alpha} = - \sum_{i=1}^m \left\{ x_{ij} y_i^\alpha \log(y_i) \exp(x'_i \beta + \lambda - y_i^\alpha e^\lambda) \right\} \quad j = 1, \dots, p$$

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \beta_j \partial \lambda} = - \sum_{i=1}^m \left\{ x_{ij} y_i^\alpha \exp(x'_i \beta + \lambda - y_i^\alpha e^\lambda) \right\} \quad j = 1, \dots, p$$

$$\frac{\partial^2 l_u(\phi; D_c)}{\partial \alpha \partial \lambda} = - \sum_{i=1}^m \left\{ \delta_i e^\lambda y_i^\alpha \log(y_i) + y_i^\alpha \log(y_i) \exp(x'_i \beta + \lambda - y_i^\alpha e^\lambda) [1 - y_i^\alpha e^\lambda] \right\}$$

Com isso, pode-se realizar testes estatísticos para o vetor de parâmetros  $\beta$ . Suponha que é de interesse testar  $H_0 : \beta = \beta_0$ . Seja  $\widehat{\psi}(\beta_0)$  a estimativa de máxima verossimilhança de  $\psi$  aplicada em  $\beta = \beta_0$ . Então, a estatística razão de verossimilhança para testar  $H_0$  é dada por

$$TRV = -2 \left[ l_u(\beta_0, \widehat{\psi}(\beta_0); D) - l_u(\widehat{\phi}; D) \right].$$

Essa estatística tem distribuição assintótica  $\chi_{(p)}^2$ , sob a hipótese  $H_0$ , onde  $p$  é o número de parâmetros a serem testados.

### 3.6 Abordagem Bayesiana

Considera-se que  $Z_k$ ,  $k = 1, \dots, N_i$ , tenha função distribuição Weibull com mesma função densidade de probabilidade e função de sobrevivência da seção (3.5). Além disso, as covariáveis são incluídas no modelo pelo parâmetro de cura  $\theta_i = \exp(x'_i \beta)$ .

Assume-se que a distribuição conjunta *a priori* para os parâmetros  $(\psi, \beta)$ , com  $\psi = (\alpha, \lambda)$ , é

$$\pi(\psi, \beta) \propto \pi(\psi) \equiv \pi(\alpha|\delta_0, \tau_0)\pi(\lambda), \quad (3.9)$$

com uma distribuição *priori* imprópria uniforme para  $\beta$ , ou seja,  $\pi(\beta) \propto 1$ . Para  $\alpha$  toma-se uma distribuição gama com hiperparâmetros  $\delta_0$  e  $\tau_0$  e para  $\lambda$  uma distribuição normal com média zero e variância  $c$ .

Assim, considerando a função de verossimilhança dada por (3.3) e a distribuição *a priori* acima, tem-se que a distribuição conjunta *a posteriori* de  $(\lambda, \alpha, \beta, \mathbf{N})$  é

$$\begin{aligned} \pi(\lambda, \alpha, \beta, N|D) &\propto \prod_{i=1}^m S(y_i|\psi)^{N_i-\delta_i} (N_i f(y_i|\psi))^{\delta_i} \\ &\times \exp \left\{ \sum_{i=1}^m N_i x'_i \beta - \log(N_i!) - \exp(x'_i \beta) \right\} \pi(\psi, \beta). \end{aligned} \quad (3.10)$$

Para obter as estimativas dos parâmetros aplica-se algoritmo MCMC, Gibbs com Metropolis-Hastings, utilizando as distribuições condicionais *a posteriori* dadas abaixo,

$$\begin{aligned} \pi(\alpha|\beta, \lambda, N, D) &\propto \exp \left\{ \sum_{i=1}^m -N_i y_i^\alpha \exp(\lambda) - \delta_i \log(\alpha) + \delta_i (\alpha - 1) \log(y_i) \right\} \\ &\times \pi(\alpha|\delta_0, \tau_0) \end{aligned} \quad (3.11)$$

$$\begin{aligned} \pi(\lambda|\alpha, \beta, N, D) &\propto \exp \left\{ \sum_{i=1}^m -N_i y_i^\alpha \exp(\lambda) + \delta_i \lambda \right\} \pi(\lambda|c) \end{aligned} \quad (3.12)$$

Em ambas condicionais *a posteriori* tem-se termos conhecidos, facilitando a obtenção da distribuição de transição do algoritmo Metropolis.

$$\begin{aligned} \pi(\beta|\alpha, \lambda, D) &\propto \exp \left\{ \sum_{i=1}^m \delta_i x'_i \beta - \exp(x'_i \beta) + \exp(x'_i \beta + \lambda - y_i^\alpha e^\lambda) \right\} \\ &\times \pi(\beta) \end{aligned} \quad (3.13)$$

Para amostrar  $\beta$ , por não apresentar uma forma fechada ou simples, usa-se passeio aleatório.

Além disso, tem-se que

$$N_i \sim \text{Poisson} \left( S(y_i|\psi) \exp(x'_i \beta) \right) + \delta_i$$

(ver Apêndice C).

## Capítulo 4

---

# Modelo de Fração de Cura com Bernoulli

---

### 4.1 Introdução

Muitos cientistas por um longo tempo usavam uma razão de sobrevivência como um índice para verificar a eficiência de tratamentos. A idéia era que se um indivíduo sobrevivesse um periodo maior que 5 anos poderia ser considerado curado. No entanto, essa idéia tem suas divergências.

Baseado em Boag (1949) e após acompanhar estudos de câncer de estômago, Berkson & Gage (1952) propôs uma simples função em termos de dois parâmetros fisicamente significativos usados para comparar a mortalidade de dois grupos, diferença entre tratamentos, tipos de câncer, entre outros. Um termo representa a proporção da população sujeita apenas a uma razão de morte natural e o outro esta relacionado a razão de morte por câncer. A idéia surgiu após observar que os grupos divididos pelo grau de gravidade do câncer, depois de um longo período de acompanhamento tiveram uma razão de morte aproximadamente de uma população normal.

Neste capítulo mostra-se que o modelo proposto por Berkson & Gage (1952) é um caso particular da teoria apresentada no primeiro capítulo.

### 4.2 Caso particular: Bernoulli

Considerar as mesmas propriedades descritas no Capítulo 3 assumindo que a variável aleatória latente  $N_i$ , neste caso, tenha distribuição de Bernoulli com parâmetro  $1 - \theta$ , ou seja,

$$N_i = \begin{cases} 1 & \text{indivíduo tem a causa do evento de interesse,} \\ 0 & \text{indivíduo não tem a causa do evento de interesse.} \end{cases}$$

A função geradora de sequência (Feller, 1967) no caso de assumir uma distribuição Binomial( $r, 1 - \theta$ ) é

$$A_p(s) = [\theta + (1 - \theta)s]^r.$$

Assim, para  $r = 1$  obtém-se a função geradora de longa-duração da população usando o Teorema 1,

$$S_p(y) = A_p[S(y)] = \theta + (1 - \theta)S(y), \quad (4.1)$$

que corresponde exatamente a função obtida por Berkson & Cage (1952). Quando o tempo aumenta a função  $S_p(y)$  converge para  $\theta$ , mostrando ser esta a fração de cura da população.

A sub-função de risco e densidade são obtidas de (4.1) e dadas por

$$f_p(y) = (1 - \theta)f(y)$$

e

$$h_p(y) = f(y) \frac{1 - \theta}{\theta + (1 - \theta)S(y)},$$

respectivamente.

Observa-se que  $h_p(y)$  na presença de covariáveis não satisfaz a propriedade de risco proporcional, uma vez que essa função sempre dependerá do tempo.

A função de sobrevivência própria, densidade e de risco para a população não curada são

$$S_p^*(y) = S(y), \quad f^*(y) = f(y)$$

e

$$h_p^*(y) = \frac{f(y)}{S(y)},$$

respectivamente.

O modelo (4.1) tem sido usado por muitos pesquisadores, no entanto, apresenta algumas desvantagens, (Chen, Ibrahim & Sinha, 1999). Primeiro, na presença de covariáveis não satisfaz a propriedade de risco proporcional que é desejável em modelos de sobrevivência, especialmente do ponto de vista frequentista pois muitos resultados assintóticos e computacionais requerem essa propriedade. E segundo, introduzindo covariáveis pelo parâmetro  $\theta$ , via modelo de regressão binomial, a expressão (4.1) produz distribuição *posteriori* imprópria para vários tipos de distribuição *priori* imprópria não informativa. Por esses motivos as covariáveis não são usadas no modelo (4.1).

### 4.3 Função de Verossimilhança

Devido a facilidade de cálculos e de estimação computacional, utiliza-se a função de verossimilhança (2.20), em que a variável latente considerada é  $W_i$ . Substituindo os termos devidamente tem-se,

$$L_A(\theta, \psi; D) = \prod_{i=1}^m \left[ (1 - \theta) f(y|\psi) \right]^{\delta_i} \left[ \theta^{W_i} (1 - \theta)^{(1-W_i)} S(y|\psi)^{(1-W_i)} \right]^{1-\delta_i}. \quad (4.2)$$

E a função log-verossimilhança de (4.2) é

$$\begin{aligned} l_A(\theta, \psi; D) &= \sum_{i=1}^m \left\{ \delta_i \log(f(y|\psi)) + (1 - \delta_i)(1 - W_i) \log(S(y|\psi)) \right\} + \\ &+ \sum_{i=1}^m \left\{ (1 - \delta_i) W_i \log(\theta) + (1 - W_i + \delta_i W_i) \log(1 - \theta) \right\} \end{aligned} \quad (4.3)$$

### 4.4 Estimação dos Parâmetros

Diferente do que foi feito na seção 2.3, utiliza-se a função de verossimilhança (4.2) contendo a variável latente  $W_i$ .

Suponha que  $F$  seja uma função de distribuição acumulada Weibull com vetor de parâmetros  $\psi = (\alpha, \lambda)$ , ou seja,  $Z_{ik} \sim Weibull(\alpha, \lambda)$  com  $k = 0, 1$  e  $i = 1, \dots, m$ . Assin, tem-se a função densidade de probabilidade  $f(y|\alpha, \lambda) = \alpha y^{\alpha-1} \exp(\lambda - \exp(\lambda)y^\alpha)$  e função de sobrevivência  $S(y|\alpha, \lambda) = \exp(-y^\alpha \exp(\lambda))$ . Assim, a função log-verossimilhança (4.3) pode ser escrita como

$$\begin{aligned} l_A(\theta, \psi; D) &= \sum_{i=1}^m \left\{ \delta_i [\log(\alpha) + (\alpha - 1) \log(y_i) + \lambda - \exp(\lambda)y_i^\alpha] - (1 - \delta_i)(1 - W_i)y_i^\alpha \exp(\lambda) \right\} \\ &+ \sum_{i=1}^m \left\{ (1 - \delta_i) W_i \log(\theta) + (1 - W_i + \delta_i W_i) \log(1 - \theta) \right\}. \end{aligned} \quad (4.4)$$

Assim, do ponto de vista clássico, será usado o algoritmo EM apresentado no Capítulo 2.

A partir do cálculo da distribuição condicional de  $W_i$  dado  $D$  (Apêndice C), tem-se que a esperança condicional (2.23) é expresso por,

$$E [W_i|D; \phi^{(k)}] = \frac{\theta^{(1-\delta_i)}}{\theta^{(1-\delta_i)} + [(1-\theta)S(y_i|\psi)]^{(1-\delta_i)}} \quad (4.5)$$

e denotada por  $W_i^{(k+1)}$ . A variável latente  $W$  é considerada apenas para indivíduos censurados, ou seja, quando  $\delta_i = 0$ .

Portanto, o algoritmo EM é formado pelos seguintes passos:

**Passo E:** Cálculo da esperança condicional da log-verossimilhança (4.4)

$$\begin{aligned} E[l_A(\theta, \psi|D)|D, \theta^{(k)}, \psi^{(k)}] &= \sum_{i=1}^m \left\{ \delta_i [\log(\alpha) + (\alpha - 1) \log(y) + \lambda - \exp(\lambda)y^\alpha] \right. \\ &\quad \left. - (1 - \delta_i)y^\alpha \exp(\lambda) + (1 - \delta_i)W_i^{(k+1)}y^\alpha \exp(\lambda) \right\} \\ &\quad + \sum_{i=1}^m \left\{ (1 - \delta_i)W_i^{(k+1)} \log(\theta) + \log(1 - \theta) \right. \\ &\quad \left. - W_i^{(k+1)}(1 - \delta_i) \log(1 - \theta) \right\}. \end{aligned}$$

**Passo M:** Maximização das expressões

$$\begin{aligned} Q_1(\theta|\theta^{(k)}, \psi^{(k)}) &\equiv \sum_{i=1}^m \left\{ (1 - \delta_i)W_i^{(k+1)} \log(\theta) + \log(1 - \theta) \right. \\ &\quad \left. - W_i^{(k+1)}(1 - \delta_i) \log(1 - \theta) \right\} \\ &\quad \text{e} \end{aligned}$$

$$\begin{aligned} Q_2(\psi|\theta^{(k)}, \psi^{(k)}) &\equiv \sum_{i=1}^m \left\{ \delta_i [\log(\alpha) + (\alpha - 1) \log(y) + \lambda - \exp(\lambda)y^\alpha] \right. \\ &\quad \left. - (1 - \delta_i)y^\alpha \exp(\lambda) + (1 - \delta_i)W_i^{(k+1)}y^\alpha \exp(\lambda) \right\}. \end{aligned}$$

As estimativas de máxima verossimilhança  $\hat{\phi} = (\hat{\theta}, \hat{\psi})$  são obtidas pela convergência do algoritmo, considerando o critério de parada

$$|\phi^{(k+1)} - \phi^{(k)}| \leq \epsilon.$$

Da mesma forma que no Capítulo 3, usa-se o fato que  $\hat{\phi}$  tem distribuição assintótica normal multivariada com média  $\phi$  e matriz de variância-covariância  $I^{-1}(\phi)$  e dessa forma calcular estimativas para a variância de  $\hat{\phi}$  e construir testes de hipóteses para os parâmetros usando a matriz de Informação de Fisher  $I(\phi)$

$$I(\phi) = -E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \phi_i \partial \phi_j} \right]_{i,j=1,\dots,p+2}$$



$$I(\phi) = \begin{pmatrix} E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta^2} \right] & E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta \partial \alpha} \right] & E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta \partial \lambda} \right] \\ & E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \alpha^2} \right] & E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \alpha \partial \lambda} \right] \\ & & E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \lambda^2} \right] \end{pmatrix},$$

Os elementos dessa matriz são dados pelas expressões,

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \alpha^2} \right] = - \sum_{i=1}^m \left\{ \delta_i \left[ \frac{1}{\alpha^2} \right] + \left( 1 - E [W_i | D; \phi^{(k)}] + \delta_i E [W_i | D; \phi^{(k)}] \right) \exp(\lambda) \log^2(y_i) y_i^\alpha \right\}$$

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \lambda^2} \right] = - \sum_{i=1}^m \left\{ \left( 1 - E [W_i | D; \phi^{(k)}] + \delta_i E [W_i | D; \phi^{(k)}] \right) y_i^\alpha \exp(\lambda) \right\}$$

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta^2} \right] = - \sum_{i=1}^m \left\{ \frac{(1 - \delta_i) E [W_i | D; \phi^{(k)}]}{\theta^2} + \frac{(1 - E [W_i | D; \phi^{(k)}] + \delta_i E [W_i | D; \phi^{(k)}])}{(1 - \theta)^2} \right\}$$

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta \partial \alpha} \right] = 0$$

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \theta \partial \lambda} \right] = 0$$

$$E \left[ \frac{\partial^2 l_A(\phi; D)}{\partial \alpha \partial \lambda} \right] = - \sum_{i=1}^m \left\{ y_i^\alpha e^\lambda \log(y_i) + (1 - \delta_i) (1 - E [W_i | D; \phi^{(k)}]) e^\lambda y_i^\alpha \log(y_i) \right\},$$

onde os valores do termo  $E [W_i | D; \phi^{(k)}]$ , dado por (4.5), são obtidos na convergência do método de estimação EM.

## 4.5 Abordagem Bayesiana

Considerando ainda o fato de  $Z_k \sim Weibull(\alpha, \lambda)$ , com  $k = 0, 1$ , assume-se que a distribuição conjunta a *priori* para os parâmetros  $(\alpha, \lambda, \theta)$ , seja

$$\pi(\alpha, \lambda, \theta) \propto \pi(\alpha, \lambda)\pi(\theta) \equiv \pi(\alpha|\delta_0, \tau_0)\pi(\lambda)\pi(\theta), \quad (4.6)$$

ou seja,  $\alpha$  e  $\lambda$  são independentes. Para  $\alpha$  toma-se uma distribuição gama de hiperparâmetros  $\delta_0$  e  $\tau_0$ , para  $\lambda$  uma distribuição normal com média zero e variância  $c$  e para  $\theta$  considera-se uma distribuição *priori* conjugada Beta  $(a, b)$ . Neste caso, a distribuição conjunta a *posteriori* de  $(\alpha, \lambda, \theta, W)$  é

$$\begin{aligned} \pi(\theta, \alpha, \lambda, W|D) &\propto \prod_{i=1}^m [1 - \theta]^{\delta_i} [\theta^{W_i}(1 - \theta)^{(1-W_i)}]^{(1-\delta_i)} \\ &\times [\alpha y_i^{\alpha-1} \exp(\lambda - \exp(\lambda)y_i^\alpha)]^{\delta_i} [\exp(-y_i^\alpha \exp(\lambda))]^{(1-W_i)(1-\delta_i)} \\ &\times \pi(\theta, \alpha, \lambda). \end{aligned} \quad (4.7)$$

Para obter estimativas dos parâmetros aplica-se novamente algoritmo MCMC, Gibbs com Metropolis-Hastings, utilizando as distribuições condicionais a *posteriori* apresentadas abaixo.

$$\pi(\lambda|\alpha, W, \theta, D) \propto \exp \left\{ \sum_{i=1}^m \left[ -(1 - W_i(1 - \delta_i)) \exp(\lambda)y_i^\alpha + \delta_i \lambda \right] \right\} \pi(\lambda) \quad (4.8)$$

$$\begin{aligned} \pi(\alpha|\lambda, W, \theta, D) &\propto \exp \left\{ \sum_{i=1}^m \left[ \delta_i \log(\alpha) + \delta_i(\alpha - 1) \log(y_i) - (1 - W_i(1 - \delta_i)) \exp(\lambda)y_i^\alpha \right] \right\} \\ &\times \pi(\alpha|\delta_0, \tau_0), \end{aligned} \quad (4.9)$$

$$\begin{aligned} \pi(\theta|\alpha, \lambda, W, D) &\propto \prod_{i=1}^m [(1 - \theta)]^{\delta_i} [\theta^{W_i}(1 - \theta)^{(1-W_i)}]^{(1-\delta_i)} \\ &\times \pi(\theta, \psi). \end{aligned} \quad (4.10)$$

Considerando  $\pi(\theta) \sim Beta(a, b)$  tem-se

$$\pi(\theta|\alpha, \lambda, W, D) \propto \theta^{[a+\sum W_i(1-\delta_i)]-1} (1 - \theta)^{[b+m-\sum W_i(1-\delta_i)]-1},$$

ou seja,

$$\theta|\alpha, \lambda, W, D \sim \text{Beta} \left( a + \sum W_i(1 - \delta_i), b + m - \sum W_i(1 - \delta_i) \right)$$

Além disso, considera-se  $W_i \sim \text{bernoulli}(p_i)$  com

$$p_i = \frac{\theta}{\theta + (1 - \theta) \exp(-\exp(\lambda)y_i^\alpha)}.$$



## Capítulo 5

---

# Aplicação

---

Neste capítulo, a aplicação realizada por Chen, Ibrahim & Sinha (1999) é reproduzida em um conjunto de dados clínicos. Também é apresentado os resultados obtidos com o uso do modelo do Capítulo 3. Além disso, apresenta-se os resultados obtidos usando o modelo do Capítulo 4. Um estudo de simulação é apresentado para o modelo Poisson e métodos de comparação de modelos são apresentados no ponto de vista clássico e Bayesiano. A implementação dos programas computacionais foi realizada usando o *software* R.

### 5.1 Descrição dos Dados

Os dados utilizados na aplicação estão relacionados a um ensaio clínico sobre melanoma cutâneo fase III (câncer de pele maligno), para a avaliação do desempenho dos pacientes pós-cirurgia mediante o tratamento de quimioterapia com alta dose de interferon alpha-2b (IFN), para combater a reincidência do câncer. O registro dos pacientes ocorreram de 1984 até 1990 e o acompanhamento foi realizado até 1993. Os dados foram obtidos de Ibrahim *et al.* (2001)(<http://merlot.stat.uconn.edu/mhchen/survbook/>) e a variável resposta é a sobrevivência global definida como o tempo da aleatorização até a morte (para maiores informações ver Kirkwood *et al.*(1996)).

As variáveis  $Z_1, \dots, Z_N$  representam os tempos de ativação de cada uma das  $N$  células cancerígenas do paciente.

A amostra tem um tamanho original de  $m = 286$  pacientes, sendo que 2 não apresentaram informações completas. Retirando estes casos, obtém-se uma amostra de  $m = 284$  pacientes. O tempo médio de vida em anos foi de 3,71. A porcentagem de observações censuradas foi de aproximadamente 39%. A estimativa de Kaplan-Meier do tempo de sobrevivência mediano foi aproximadamente de 3,15 anos, ou seja, pode-se dizer que 50% dos indivíduos sobrevivem pelo menos 3,15 anos.

Para cada paciente  $i$ ,  $i = 1, \dots, m$ , tem-se associados as seguintes variáveis:

- $y_i$ : tempo de sobrevivência observado em anos;
- $\delta_i$ : indicador de censura (0=censura, 1=falha);
- $x_1$ : idade em anos;
- $x_2$ : sexo (0=masculino, 1=feminino) e
- $x_3$ : "performance status"(p.s. - escala de capacidade funcional do paciente em suas atividades diárias, 0=ativo, 1=outros)

Durante toda a análise a covariável idade ( $x_1$ ) foi padronizada para estabilizar a computação *posteriori*.

A Figura 2, apresentado na Introdução, mostra o gráfico da função de sobrevivência estimada de Kaplan-Meier que apresenta uma estabilidade na cauda indicando, dessa forma, o uso de modelos de fração de cura.

## 5.2 Análise Clássica

Inicialmente, o efeito da covariável idade e sexo nos pacientes foram analisadas, usando a função de sobrevivência estimada de Kaplan-Meier. A seguir tem-se os gráficos dessa função.

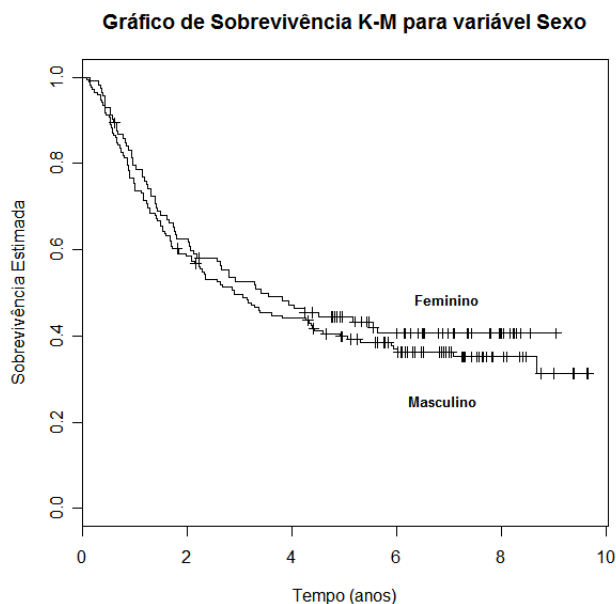


Figura 5.1: Função Kaplan-Meier para covariável Sexo

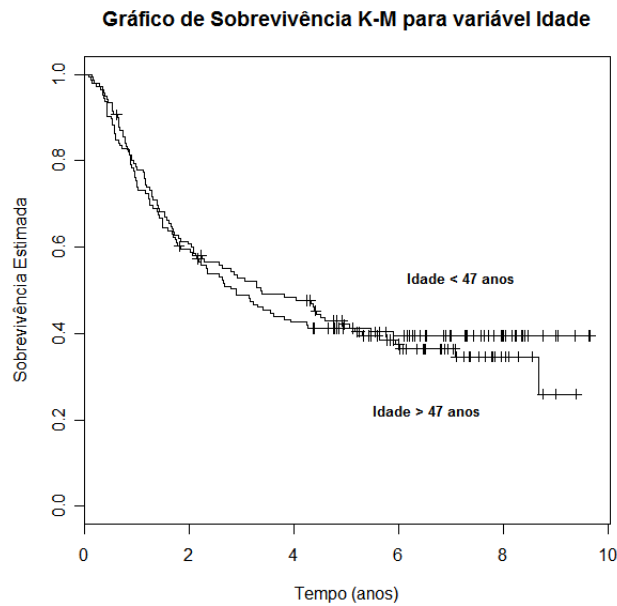


Figura 5.2: Função Kaplan-Meier para covariável Idade

De acordo com o primeiro gráfico, pacientes de sexo feminino tendem a uma probabilidade de sobreviver ao câncer maior do que os de sexo masculino. Já no segundo gráfico, os de idade menor que 47 anos (idade média de vida dos pacientes), têm maior probabilidade de sobreviver à doença, ou seja, pode-se pensar que pacientes mais jovens têm menor risco de sofrer morte ou reincidência do câncer. Isso é confirmado pelos parâmetros estimados. Além disso, pode-se observar que as curvas não indicam violação da suposição de risco proporcional. A situação extrema de violação é caracterizada por curvas que se cruzam.

### I. Modelo Poisson

Ajustando o modelo paramétrico Weibull apresentado na Seção 3.5, obtém-se as estimativas da tabela abaixo, usando a função `nlm` do software R para maximizar as expressões do passo M do algoritmo EM.

Tabela 5.1: EMV para dados de melanoma com  $N \sim \text{Poisson}$

Parâmetro	Estimativa	SD	IC 95%
$\beta_0$ (Intercepto)	0,091	0,0773	(-0,0605 , 0,2425)
$\beta_1$ (Idade)	0,091	0,0718	(-0,0497 , 0,2317)
$\beta_2$ (Sexo)	-0,1213	0,1284	(-0,3729 , 0,1304)
$\beta_3$ (P.S)	-0,189	0,2476	(-0,6743 , 0,2963)
$\alpha$	1,314	0,0758	(1,1654 , 1,4626)
$\lambda$	-1,337	0,1097	(-1,5520 , -1,1219)

Assim, a função geradora de longa-duração (3.1) é expressa por

$$S_p(y) = \exp \left\{ -1,0302 \left( 1 - \exp \left( - \exp(-1,337) y^{1,314} \right) \right) \right\}, \quad (5.1)$$

Para cada indivíduo  $i$  calcula-se o valor de  $\theta_i$  e o valor médio obtido foi  $\bar{\theta} = -1,0302$ .

Pela Tabela 5.1, pode-se observar que a variável idade tem influência na sobrevivência dos indivíduos com câncer, pois quanto maior a idade menor será a seu tempo de vida, ou seja, maior será a chance do paciente vir a morrer da doença ou ocorrer reincidência da mesma. Da mesma forma, observa-se que pacientes de sexo feminino têm menos chance de ocorrer reincidência ou a morte pela doença. Além disso, as estimativas na Tabela 5.1 são próximas das obtidas pelos autores Chen, Ibrahim e Sinha e os intervalos de confiança para os parâmetros de regressão mostram que os mesmos não são significativos. Um teste de hipótese para estes parâmetros é realizado na seção 5.6 para comprovar suas significâncias.

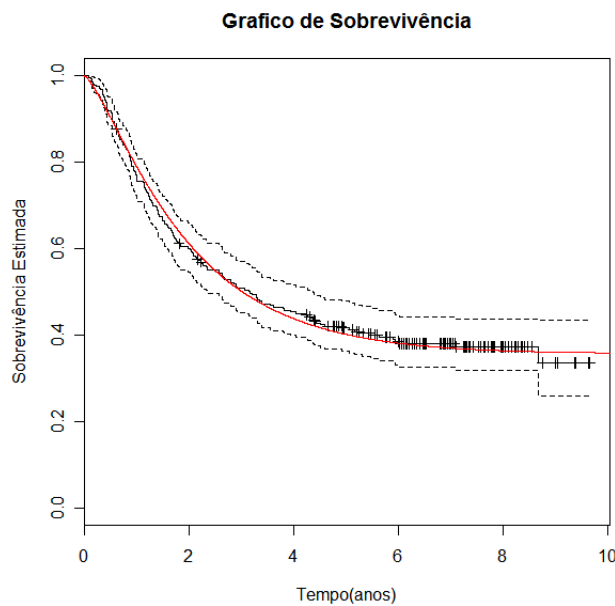


Figura 5.3: Função de Sobrevivência paramétrica Weibull e de Kaplan-Meier

Como pode ser observado na Figura 5.3, a curva da função a (5.1) está aderente a curva de Kaplan-Meier. A fração de cura obtida foi de  $p_0 = 0,357$  (considerando  $\lim_{y \rightarrow \infty} S_p(y)$ ), ou seja, aproximadamente 36% dos pacientes com melanoma fase III sobreviveram a doença com o uso do tratamento de quimioterapia a base de interferon alpha-2b.



## II. Modelo Bernoulli

Nesse caso, ajusta-se o modelo paramétrico considerando  $W \sim \text{Bernoulli}$ , como apresentado na Seção (4.4). As estimativas de máxima verossimilhança dos parâmetros estão apresentados na Tabela 5.2 abaixo.

Tabela 5.2: EMV para dados de melanoma com  $W \sim \text{Bernoulli}$

Parâmetro	Estimativa	SD	IC 95%
$\theta$	0,365	0,0286	(0,308944 , 0,421056)
$\alpha$	1,314	0,0141	(1,286364 , 1,341636)
$\lambda$	-1,337	0,0224	(-1,380904 , -1,293096)

Pode-se observar que as estimativas obtidas de  $\alpha$  e  $\lambda$  são exatamente os mesmos valores obtidos pelo modelo Poisson. Além disso, tem-se que seus intervalos de confiança na Tabela 5.2 são menores do que os intervalos apresentados na Tabela 5.1. A exclusão das covariáveis neste modelo ocorre pelo fato dos coeficientes de regressão não terem sido significativos no modelo de Poisson.

Usando essas estimativas tem-se

$$S_p(y) = 0,365 + 0,635 \exp(-\exp(-1,337) y^{1,314}). \quad (5.2)$$

Observa-se pela tabela acima que a fração de cura é de  $p_0 = 0,365$ . A Figura 5.4 mostra uma boa aderência da função (5.2) em relação a curva de Kaplan-Meier.

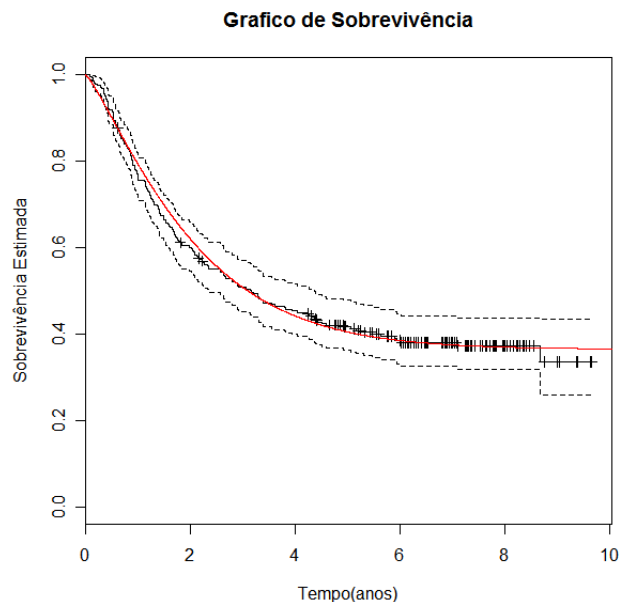


Figura 5.4: Função de Sobrevivência paramétrica Weibull e de Kaplan-Meier

## 5.3 Análise Bayesiana

### I. Modelo Poisson

Para ajustar o modelo paramétrico Weibull apresentado na Seção 3.6 foi considerado uma distribuição *priori* imprópria uniforme para  $\beta$ , uma distribuição gama não informativa para  $\alpha$ , com hiperparâmetros  $\delta_0 = 1$  e  $\tau_0 = 0,01$  e uma distribuição normal com média  $\mu = 0$  e variância  $c = 10000$  (esses valores foram sugeridos por Chen, Ibrahim & Sinha, 1999). Com isso obtém-se as estimativas a *posteriori* dadas na tabela abaixo, usando algoritmo de Metropolis-Hastings com Gibbs.

Tabela 5.3: Estimativas a *posteriori* para dados de melanoma com  $N \sim \text{Poisson}$

Parâmetro	Estimativa	SD	Intervalo de Credibilidade	
			2,5%	97,5%
$\beta_0$ (Intercepto)	0,0881	0,1022	-0,1153	0,2942
$\beta_1$ (Idade)	0,0901	0,0698	-0,0624	0,2153
$\beta_2$ (Sexo)	-0,1159	0,1582	-0,4144	0,1975
$\beta_3$ (P.S)	-0,2504	0,2606	-0,7804	0,2511
$\alpha$	1,2915	0,0813	1,1395	1,4533
$\lambda$	-1,3313	0,1165	-1,5598	-1,0799

Na Tabela 5.3 observa-se que as estimativas, os desvio padrão e os intervalos de credibilidade estão próximos dos obtidos por Chen, Ibrahim & Sinha (1999).

O teste de convergência foi realizado pelo critério de Gelman-Rubin sob duas cadeias. Na tabela abaixo tem-se seu valor para cada covariável.

Tabela 5.4: Critério de Convergência Gelman-Rubin

Parâmetro	Gelman-Rubin
$\beta_0$	1,02
$\beta_1$	1,01
$\beta_2$	1
$\beta_3$	1,1
$\alpha$	1,05
$\lambda$	1,15

A função geradora de longa-duração é expressa por

$$S_p(y) = \exp \left\{ -1,0239 (1 - \exp(-\exp(-1,3313) y^{1,2915})) \right\} \quad (5.3)$$

onde  $\bar{\theta} = -1,0239$  e a fração de sobreviventes é de  $p_0 = 0,359$ .

A Figura 5.5 mostra o modelo paramétrico Weibull comparado a função de sobrevivência estimada de Kaplan-Meier.

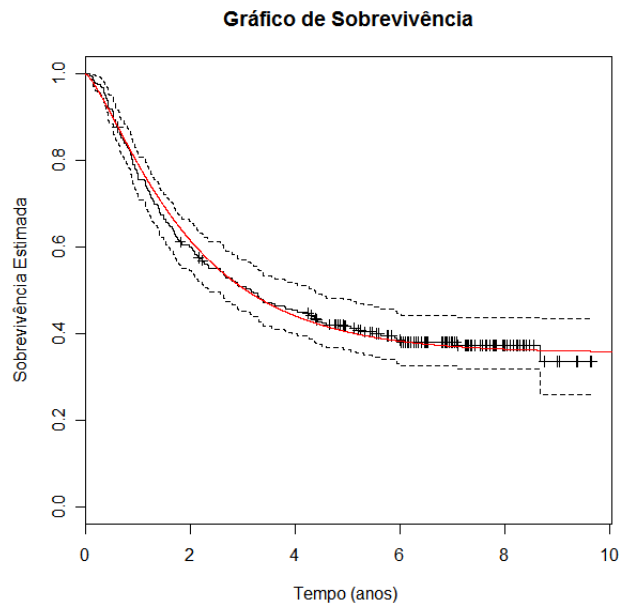


Figura 5.5: Função de Sobrevivência paramétrica Weibull e de Kaplan-Meier

## II. Modelo Bernoulli

Neste caso ajusta-se o modelo considerando uma distribuição  $Beta(1, 1)$  para  $\theta$ , uma distribuição  $Gama(1; 0, 01)$  para  $\alpha$  e uma distribuição  $Normal(0, 10000)$  para  $\lambda$ .

Usando o algoritmo Metropolis-Gibbs nas distribuições condicionais a *posteriori* da Seção 4.5, obtém-se as estimativas dos parâmetros na tabela a seguir.

Tabela 5.5: Estimativas a *posteriori* para dados de melanoma com  $W \sim \text{Bernoulli}$

Parâmetro	Estimativa	SD	Intervalo Credibilidade	
			2,5%	97,5%
$\theta$	0,3619	0,0316	0,3015	0,4239
$\alpha$	1,1679	0,0753	1,0204	1,3289
$\lambda$	-0,9044	0,0995	-1,0743	-0,7494

Neste caso, as covariáveis não são usadas pelo fato de que a distribuição *priori* imprópria para o vetor  $\beta$  produz uma distribuição *posteriori* imprópria (ver teorema Chen, Ibrahim & Sinha, 1999).

Utiliza-se o critério de Gelman-Rubin sob duas cadeias para verificar a convergência do algoritmo. Os valores estão na Tabela 5.6.

Tabela 5.6: Critério de Convergência Gelman-Rubin

Parâmetro	Gelman-Rubin
$\theta$	1
$\alpha$	1,02
$\lambda$	1

Com as estimativas da Tabela 5.5 a função geradora de longa-duração pode ser expressa como

$$S_p(y) = 0,3619 + 0,6381 \exp(-0,4048 y^{1,1679}) \quad (5.4)$$

Assim, a estimativa da fração de cura é de  $p_0 = 0,3619$ , como pode ser observado na figura abaixo.

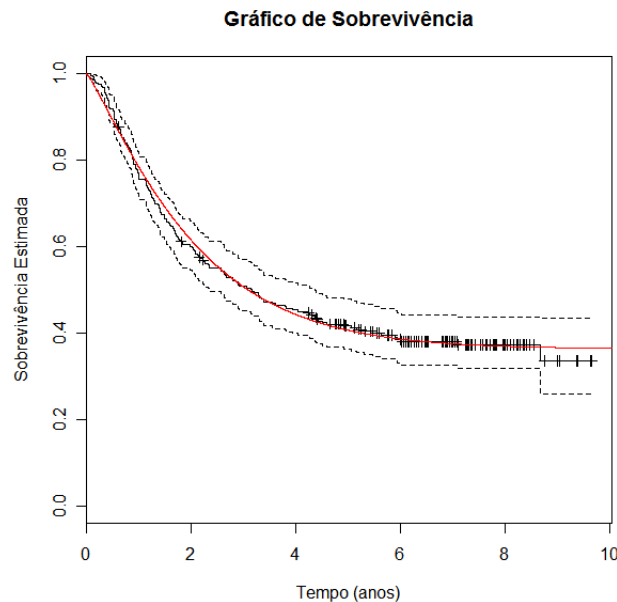


Figura 5.6: Função de Sobrevivência paramétrica Weibull e de Kaplan-Meier

## 5.4 Comparação dos Modelos

Essa seção tem como objetivo comparar os modelos, apresentados anteriormente, através dos métodos AIC (Akaike Information Criterion), Akaike (1973), para a análise clássica e o método BIC (Schwarz's Bayesian criterion), Schwarz (1978), para a análise *Bayesiana*.

Para o cálculo desses métodos foi utilizado a seguinte fórmula:

$$-2 \log(L) + k \text{ npar},$$

onde  $L$  : função de verossimilhança,  $\text{npar}$  = número de parâmetros e para  $k=2$  tem-se o valor para o AIC ou  $k = \log(m)$  tem-se o valor para o BIC. Neste último caso,  $m$  representa o tamanho da amostra.

Para modelar o tempo de sobrevivência foi utilizado a distribuição Weibull e a função de verossimilhança unificada (2.18).

Primeiramente, comparou-se os modelos do ponto de vista clássico. Assim, os valores de AIC encontrados foram:

Tabela 5.7: Valores do método AIC

Modelo	AIC
Poisson	923,45
Bernoulli	979,70

Isso mostra que o modelo de Poisson é mais adequado do que o modelo Bernoulli.

Vale lembrar que o modelo Poisson é regressivo e ainda conseguiu apresentar um desempenho melhor do que o modelo Bernoulli que não é regressivo.

Comparando os modelos do ponto de vista Bayesiano, tem-se os valores do BIC na tabela abaixo.

Tabela 5.8: Valores do método BIC

Modelo	BIC
Poisson	945,48
Bernoulli	969,47

Novamente, o modelo Poisson é mais adequado.

Neste caso, o modelo de Bernoulli não pode ser regressivo, pois segundo Chen, Ibrahim & Sinha (1999) o modelo resultante não produz distribuições a posteriori impróprias.

## 5.5 Estudo de Simulação

Nesta seção, apresenta-se um estudo de simulação com o objetivo de analisar o comportamento dos estimadores dos modelos visto nas seções anteriores, para diversas porcentagens de censura e tamanho de amostra.

Durante o processo, considera-se os riscos competitivos com função de distribuição Exponencial, ou seja, fixa-se  $\alpha = 1$ . Para cada indivíduo foi considerado uma única variável,  $x$ , gerada a partir de uma distribuição Normal.

1.  $N \sim \text{Poisson}(\theta)$ 

Os valores gerados para cada indivíduo  $i$ , foram:

- $x_i \sim N(0, 1)$
- $y_i = \min(t_i, c_i)$
- $\delta_i = I_{\{t_i \leq c_i\}}$
- $T_i = \min(Z_{i0}, \dots, Z_{iN_i})$
- $Z_{il} \sim \text{exponencial}(e^\lambda)$ ,  $l = 1, \dots, N_i$  e  $\lambda = -1, 33$
- $N_i \sim \text{Poisson}(\theta_i)$ ,  $\theta_i = \exp(\beta x_i)$ ,  $\beta = 0, 09$
- $C_i \sim \text{exponencial}(e^\mu)$ , sendo  $\mu = \mu(p_c)$  calculado de acordo com a proporção de censura fixada para a população de não censurados  $p_c$ , pela relação  $p_c = P(T > C)$ . (Mizoi, M. 2004)

As amostras foram geradas com proporções de censura  $p_c = 0\%$ ,  $p_c = 15\%$ ,  $p_c = 30\%$  e  $p_c = 50\%$ , tamanhos de amostras  $m = 50$ ,  $m = 100$ ,  $m = 200$  e  $m = 300$ . Para cada combinação destes valores foram geradas 1000 replicações e tomadas a média das estimativas dos parâmetros e a partir desta calcula-se erro padrão.

Os resultados destas simulações são apresentados nas tabelas a seguir.

Tabela 5.9: Estimativas para dados simulados com  $\beta = 0, 09$ ,  $\lambda = -1, 33$  e 0% de censura

Estimador	n=50		n=100	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0962	0,1983	0,0960	0,1184
$\lambda$	-1,4184	0,2746	-1,3839	0,1823
Estimador	n=200		n=300	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0973	0,0909	0,0922	0,0773
$\lambda$	-1,3633	0,1134	-1,568	0,0909

Tabela 5.10: Estimativas para dados simulados com  $\beta = 0,09$ ,  $\lambda = -1,33$  e 15% de censura

Estimador	n=50		n=100	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,1004	0,2268	0,0884	0,1354
$\lambda$	-1,5129	0,2571	-1,4391	0,1743
Estimador	n=200		n=300	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0925	0,091	0,0908	0,074
$\lambda$	-1,4128	0,1164	-1,4013	0,0952

Tabela 5.11: Estimativas para dados simulados com  $\beta = 0.09$ ,  $\lambda = -1.33$  e 30% de censura

Estimador	n=50		n=100	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0923	0,177	0,0891	0,1437
$\lambda$	-1,6695	0,3098	-1,5717	0,196
Estimador	n=200		n=300	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0896	0,0935	0,0878	0,0772
$\lambda$	-1,5040	0,1301	-1,4692	0,1056

Tabela 5.12: Estimativas para dados simulados com  $\beta = 0.09$ ,  $\lambda = -1.33$  e 50% de censura

Estimador	n=50		n=100	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,1140	0,2251	0,0980	0,152
$\lambda$	-2,1081	0,4037	-1,9319	0,2729
Estimador	n=200		n=300	
	média	Erro Padr.	média	Erro Padr.
$\beta$	0,0894	0,1034	0,0902	0,0778
$\lambda$	-1,8194	0,2202	-1,7204	0,2174

Observa-se das tabelas 5.9 a 5.12 que o erro padrão dos estimadores diminui com o aumento do tamanho da amostra e aumenta quando a proporção de censura cresce.

## 5.6 Testes de Hipótese

Com o intuito de verificar a significância dos parâmetros de regressão usados no modelo de Poisson, foi realizado um teste de hipótese para o vetor  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ .

Os testes usados foram: Wald e Razão de Verossimilhança, e ambos usam estatísticas de qui-quadrado.

A hipótese a ser testada é que o vetor de parâmetros  $\beta$  é um vetor nulo, ou seja,

$$H_0 : \beta = 0$$

vs

$$H_1 : \beta \neq 0$$

- Teste de Wald

A estatística usada neste teste para um escalar é

$$W_i = \frac{\widehat{\beta}_i^2}{\widehat{Var}(\widehat{\beta}_i)}$$

que sob a hipótese nula tem aproximadamente distribuição qui-quadrado com 1 grau de liberdade a um nível  $100\alpha\%$  de significância. Para valores de  $W_i$  superiores ao valor tabelado da distribuição  $\chi_{1,1-\alpha}^2$  indicam rejeição da hipótese nula.

Dessa forma, considerando um teste de 95% de confiança obtém-se os seguintes valores

Tabela 5.13: Valores do teste de Wald

Parâmetros	$W$	p-valor
$\beta_0$	0,7531	0,6145
$\beta_1$	1,5611	0,7885
$\beta_2$	0,5913	0,5580
$\beta_3$	0,5458	0,5399

Comparando esses valores com  $\chi_{1,0,05}^2=3,8415$ , pode-se concluir a não rejeição da hipótese nula, ou seja, a presença dos parâmetros de regressão no modelo não são relevantes.

- Teste de Razão de Verossimilhança (TRV)

A estatística deste teste é dada por

$$TRV = 2 \left[ \log \left( L(\widehat{\beta}_{EMV}) \right) - \log \left( L(\widehat{\beta}_{H_0}) \right) \right],$$

onde  $\widehat{\beta}_{EMV}$  é o estimador de máxima verossimilhança do vetor beta,  $\widehat{\beta}_{H_0}$  é o estimador de máxima verossimilhança sob a hipótese nula e  $L$  é a função de verossimilhança.

A TRV tem aproximadamente distribuição  $\chi_{p,1-\alpha}^2$  com  $p$  graus de liberdade. A hipótese nula é rejeitada a um nível de  $100\alpha\%$  de significância se  $TRV > \chi_{p,1-\alpha}^2$ .



Assim, considerando um teste de 95% de confiança e usando a função de verossimilhança unificada (2.18) tem-se um TRV=3,0718 e um  $\chi_{4,0,05}^2 = 9,4877$ , ou seja, não é rejeitada a hipótese nula.

Os resultados de ambos os teste apresentados mostram que o vetor de parâmetros  $\beta$  não é significativo.



## Capítulo 6

---

# Considerações Finais

---

### 6.1 Conclusão

Nesse trabalho, foi apresentado a teoria unificada de análise de sobrevivência para modelos de longa-duração (Rodrigues *et al.* 2008) possibilitando a obtenção de qualquer função de sobrevivência para o estudo de uma população. Isso é possível devido o uso da função geradora de sequência de Feller (1967). Com essa teoria observa-se que os modelos de Berkson & Gage e de Chen, Ibrahim & Sinha, são considerados casos particulares. Outro fato importante é que o modelo de sobrevivência de longa-duração assume propriedade de risco proporcional se, e somente se, o número de causas relacionado com a ocorrência do evento de interesse assumir uma distribuição de Poisson.

Como ilustração, aplica-se os modelos de sobrevivência apresentados em um conjunto de dados reais. Após análise, foi observado que uma porcentagem de pacientes em tratamento sobreviveram a doença, considerando-os como "curados". Na inclusão de covariáveis, foi possível observar o efeito do tratamento nos pacientes individualmente ou por grupos, melhorando a conclusão sobre a sobrevivência da população. Além disso, os resultados indicam que tratamentos com quimeoterapia a base de interferon alpha-2b, em pacientes com câncer de Melanoma fase III, tem um impacto maior sobre os sobreviventes.

A comparação dos modelos através do método AIC (clássico) e BIC (bayesiano) mostrou que o modelo de Poisson é mais adequado do que o modelo Bernoulli. Além disso, o teste de hipótese revelou que a presença dos parâmetros de regressão não é significativo, dessa forma podendo não ser considerado no modelo.

A simulação mostrou que modelo de Poisson apresenta um bom comportamento nas diferentes quantidades de censura presente nos diversos tamanhos das amostra.

## 6.2 Estudos Propostos

Desenvolver um procedimento bayesiano objetivo no modelo de Berkson & Gage com covariável (pag. 207 Maller & Zhou, 1996), calcular a Matriz de Informação de Fisher via aproximação de Laplace. Obter distribuição *priori* de Jeffrey e de Referência a partir desta matriz de Fisher aproximada, onde a função de ligação é dada por

$$p_i = \frac{\exp(x'_i\beta)}{1 + \exp(x'_i\beta)}.$$

# Apêndice A

---

## Função de Verossimilhança

---

A seguir, mostra-se como foi obtido a função de verossimilhança (2.15).

Tem-se que  $y_i = \min(T_i, C_i)$ , com  $T_i = \min(Z_{i0}, Z_{i1}, \dots, Z_{iN_i})$  e  $C_i$  o tempo de censura observado, onde supõe  $N_i$  com uma distribuição discreta  $f_{N_i}(n_i)$ .

A função de verossimilhança é a distribuição conjunta dos dados, então tem-se,

$$\begin{aligned} f(\mathbf{y}, \delta, \mathbf{n}) &= \prod_{i=1}^m f(y_i, \delta_i, n_i) \\ &= \prod_{i=1}^m f(y_i, \delta_i | n_i) f(n_i) \end{aligned} \quad (\text{A.1})$$

A distribuição de  $(y_i, \delta_i)$  dado  $N_i = n_i$ , para  $i = 1, \dots, m$  é obtida da seguinte forma. Sejam  $f_T$  e  $g$  as funções densidade de probabilidade de  $T_i$  e  $C_i$ , respectivamente, e  $S_T$  e  $G$  as funções de sobrevivência de  $T_i$  e  $C_i$ , respectivamente, para  $i = 1, \dots, m$ .

Logo,

$$\begin{aligned} S_T(t|n_i) &= P(T_i > t | N_i = n_i) = P(\min(Z_{i0}, Z_{i1}, \dots, Z_{iN_i}) > t) \\ &= P(Z_{i0} > t) P(Z_{i1} > t) \cdots P(Z_{iN_i} > t) \\ &= 1 S(t|\psi) \cdots S(t|\psi) \\ &= S(t|\psi)^{n_i} \end{aligned}$$

e, portanto,

$$f_T(t|n_i) = -\frac{d S_T(t|n_i)}{dt} = n_i f(t|\psi) S(t|\psi)^{n_i-1}.$$

Assim, tem-se

$$\begin{aligned}
P(y_i = t, \delta_i = 0 | N_i = n_i) &= P(C_i = t, T_i > C_i | N_i = n_i) \\
&= P(T_i > C_i | C_i = t, N_i = n_i) P(C_i = t) \\
&= S_T(t | n_i) g(t) \\
&= S(t | \psi)^{n_i} g(t)
\end{aligned}$$

e

$$\begin{aligned}
P(y_i = t, \delta_i = 1 | N_i = n_i) &= P(T_i = t, T_i \leq C_i | N_i = n_i) \\
&= P(T_i \leq C_i | T_i = t, N_i = n_i) P(T_i = t | N_i = n_i) \\
&= G(t) f_T(t | n_i) \\
&= G(t) n_i f(t | \psi) S(t | \psi)^{n_i - 1}
\end{aligned}$$

E portanto,

$$\begin{aligned}
f(y_i, \delta_i | n_i) &= [S(y_i | \psi)^{n_i}]^{1 - \delta_i} [n_i f(y_i | \psi) S(y_i | \psi)^{n_i - 1}]^{\delta_i} \\
&= S(y_i)^{n_i - \delta_i} [n_i f(y_i | \psi)]^{\delta_i}
\end{aligned} \tag{A.2}$$

Substituindo (A.2) em (A.1) tem-se a função de verossimilhança dos dados completos dada por

$$L(\theta, \psi; D_c) = \prod_{i=1}^m S(y_i)^{N_i - \delta_i} [N_i f(y_i | \psi)]^{\delta_i} f(N_i). \tag{A.3}$$

## Apêndice B

---

# Função de Verossimilhança Unificada

---

A demonstração da função de verossimilhança (2.18) é simples, considerando as seguintes situações:

- Quando o indivíduo é censurado  $\Rightarrow \delta_i = 0$

Tem-se,

$$\begin{aligned} L_u(\phi; D) &= \sum_{\mathbf{n}} \prod_{i=1}^m S(y_i|\psi)^{n_i} p_{n_i}(\theta) = \prod_{i=1}^m \sum_{n_i=0}^{\infty} S(y_i|\psi)^{n_i} p_{n_i}(\theta) \\ &= \prod_{i=1}^m S_p(y_i) \end{aligned}$$

- Quando o indivíduo é não censurado  $\Rightarrow \delta_i = 1$

Tem-se,

$$\begin{aligned} L_u(\phi; D) &= \sum_{\mathbf{n}} \prod_{i=1}^m S(y_i|\psi)^{n_i-1} (n_i f(y_i|\psi)) p_{n_i}(\theta) \\ &= \prod_{i=1}^m \sum_{n_i=0}^{\infty} S(y_i|\psi)^{n_i-1} n_i f(y_i|\psi) p_{n_i}(\theta) \\ &= \prod_{i=1}^m \sum_{n_i=0}^{\infty} -\frac{d S(y_i|\psi)^{n_i}}{d y_i} p_{n_i}(\theta) = \prod_{i=1}^m -\frac{d \left[ \sum_{n_i=0}^{\infty} S(y_i|\psi)^{n_i} p_{n_i}(\theta) \right]}{d y_i} \\ &= \prod_{i=1}^m -\frac{d S_p(y_i)}{d y_i} = \prod_{i=1}^m f_p(y_i) \end{aligned}$$

Portanto,

$$\begin{aligned} L_u(\phi; D) &= \prod_{i=1}^m [f_p(y_i|\phi)]^{\delta_i} [S_p(y_i|\phi)]^{1-\delta_i} \\ &= \prod_{i \in \bar{C}} f_p(y_i|\phi) \prod_{i \in C} S_p(y_i|\phi) \end{aligned}$$



## Apêndice C

---

# Distribuição Condicional de N e W

---

A seguir, é descrito a obtenção da distribuição condicional de  $N_i$  dado  $D$ , para  $i = 1, \dots, m$ . Analogamente, obtém-se a distribuição para  $W_i|D$ , por isso para esse caso é apresentado apenas o resultado. Considere  $N_i \sim \text{Poisson}(\theta)$  e  $W_i \sim \text{Bernoulli}(1 - \theta)$ .

A densidade condicional de  $N_i|D$  é dada por

$$f(n_i|y_i, \delta_i) = \frac{f(y_i, \delta_i, n_i)}{f(y_i, \delta_i)}. \quad (\text{C.1})$$

O denominador da expressão (C.1) é obtido como segue.

Usando a expressão (A.2) tem-se,

$$\begin{aligned} f(y_i, \delta_i) &= \sum_{n_i=0}^{\infty} f(y_i, \delta_i, n_i) \\ &= \sum_{n_i=0}^{\infty} f(y_i, \delta_i|n_i) f(n_i) \\ &= \sum_{n_i=0}^{\infty} S(y_i|\psi)^{n_i-\delta_i} [n_i f(y_i|\psi)]^{\delta_i} \frac{\theta^{n_i}}{n_i!} e^{-\theta} \\ &= \left\{ \sum_{n_i=0}^{\infty} \frac{n_i^{\delta_i} [\theta S(y_i|\psi)]^{n_i} e^{-\theta S(y_i|\psi)}}{n_i!} \right\} S(y_i|\psi)^{-\delta_i} e^{-\theta} e^{\theta S(y_i|\psi)} f(y_i|\psi)^{\delta_i} \\ &= \theta^{\delta_i} e^{\theta S(y_i|\psi)} f(y_i|\psi)^{\delta_i} e^{-\theta} \end{aligned}$$

Portanto, tem-se

$$\begin{aligned} f(n_i|y_i, \delta_i) &= \frac{n_i^{\delta_i} S(y_i|\psi)^{n_i-\delta_i} f(y_i|\psi)^{\delta_i} \theta^{n_i} e^{-\theta}}{n_i! \theta^{\delta_i} e^{-\theta} e^{\theta S(y_i|\psi)} f(y_i|\psi)^{\delta_i}} \\ &= \frac{\theta^{n_i-\delta_i} e^{-\theta S(y_i|\psi)} S(y_i|\psi)^{n_i-\delta_i}}{n_i! n_i^{-\delta_i}} \end{aligned}$$

Observa-se que

$$\frac{n_i!}{n_i^{\delta_i}} = \begin{cases} n_i! & \text{se } \delta_i = 0 \\ (n_i - 1)! & \text{se } \delta_i = 1 \end{cases}$$

logo,

$$\frac{n_i!}{n_i^{\delta_i}} = (n_i - \delta_i)!.$$

Então,

$$f(n_i|y_i, \delta_i) = \frac{e^{-\theta S(y_i|\psi)} [\theta S(y_i|\psi)]^{n_i - \delta_i}}{(n_i - \delta_i)!}. \quad (\text{C.2})$$

Conclui-se que  $N_i|\mathbf{D}$  possui a mesma distribuição de  $H_i + \delta_i$ , onde  $H_i$  é uma variável Poisson com  $E(H_i) = \theta S(y_i|\psi)$ .

Analogamente, obtém-se a distribuição condicional de  $W_i|\mathbf{D}$  dada por:

$$f(w_i|y_i, \delta_i) = \frac{[p_0^{W_i} (1 - p_0)^{1 - W_i} S_p^*(y_i|\phi)^{1 - W_i}]^{1 - \delta_i}}{p_0^{(1 - \delta_i)} + [(1 - p_0) S_p^*(y_i|\phi)]^{(1 - \delta_i)}}, \quad (\text{C.3})$$

onde  $p_0$  é a fração de cura. A esperança é dada por

$$E(w_i|y_i, \delta_i) = \frac{p_0^{(1 - \delta_i)}}{p_0^{(1 - \delta_i)} + [(1 - p_0) S_p^*(y_i|\phi)]^{(1 - \delta_i)}} \quad (\text{C.4})$$

## Apêndice D

---

### Condição para Teorema 4

---

Neste apêndice é demonstrado a condição usada no Teorema 4.

Pode-se observar que a função dada em (2.10), satisfaz a propriedade de risco proporcional se o termo

$$\frac{-\left.\frac{dA(s)}{ds}\right|_{s=S(y)}}{S_p(y)} \quad (\text{D.1})$$

é constante em relação a  $y$ .

Sabe-se que a função geradora de longa-duração é dada por

$$S_p(y) = A_p[S(y)] = \sum_{n=0}^{\infty} p_n [S(y)]^n \quad (\text{D.2})$$

e sua derivada por

$$S'_p(y) = -f(y) \sum_{n=1}^{\infty} n p_n [S(y)]^{n-1}, \quad (\text{D.3})$$

onde  $f(y)$  é uma função de densidade própria. A série em (D.3) converge e temos  $S'_p(0) = -f(0) E[N]$ .

Logo, para  $y = 0$  tem-se

$$\frac{-\left.\frac{dA(s)}{ds}\right|_{s=S(0)}}{S_p(0)} = -f(0) E[N] \Rightarrow \quad (\text{D.4})$$

$$\begin{aligned}
& \Rightarrow \frac{-\sum_{n=1}^{\infty} f(0) n p_n S(0)^{n-1}}{\sum_{n=0}^{\infty} p_n S(0)^n} = -f(0) E[N] \\
& \Rightarrow \frac{-\sum_{n=0}^{\infty} f(0) (n+1) p_{n+1} S(0)^n}{\sum_{n=0}^{\infty} p_n S(0)^n} = -f(0) E[N] \\
& \Rightarrow \sum_{n=0}^{\infty} (n+1) p_{n+1} S(0)^n = \sum_{n=0}^{\infty} E[N] p_n S(0)^n.
\end{aligned}$$

Sabe-se que duas séries são consideradas iguais quando seus respectivos termos gerais são iguais. Então

$$(n+1) p_{n+1} = E[N] p_n.$$

Portanto tem-se,

$$p_n = \frac{(n+1) p_{n+1}}{E[N]}$$

ou

$$p_{n+1} = \frac{p_n E[N]}{n+1}.$$

# Referências Bibliográficas

---

- [1] Akaike, H. (1973) *Information Theory and an Extension of the Maximum Likelihood Principle*, Proceeding of the Second International Symposium on Information Theory, Petrov, B.N and Caski, F., eds., Akademia Kiado, Budapest, 267-281.
- [2] Berger, J. O.; Liseo, B. (1999) & Wolpert L. Robert *Integrated Likelihood Methods for Eliminating Nuisance Parameters*, Statistical Science Vol14, 1-28.
- [3] Berkson, J. & Gage, R.P. (1952) *Survival Curve For Cancer Patients Following Treatment*, Journal of The American Statistical Association Vol47, 501-515.
- [4] Boag, J. W. (1949) *Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy*, Journal of the Royal Statistical Society Vol.11, 15-53.
- [5] Chen, M-H.; Ibrahim, J.G. & Sinha, D. (1999) *A New Model For Survival Data With a Surviving Fraction*, Journal of The American Statistical Association Vol.94 No.447, 909-919.
- [6] Colosimo, E. A. & Giolo, S. R. (2006) *Análise de Sobrevida Aplicada*, ABE - Projeto Fisher, Editora Edgard Blücher Ltda, Primeira Edição.
- [7] Cox, D. R. (1972) *Regression Models and Life Tables*, Journal Royal Statistical Society, Vol.34 B, 187-220.
- [8] Cox, D. R. & Oakes, D. (1984) *Analysis of Survival Data*, London: Chapman & Hall, Vol.34 B, 187-220.
- [9] Ehlers, R. S. (2001) *Introdução a Inferência Bayesiana*, Apostila.
- [10] Farewell, V. T. (1982) *The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors*, Biometrics Vol.38, 1041-1046.
- [11] Feller, W. (1967) *An Introduction to Probability Theory and its Applications*, Third Edition Vol.1, Wiley.
- [12] Google [www.google.com.br](http://www.google.com.br) procurar por "Câncer Melanoma".

- [13] Goldman, A. I. (1984) *Survivorship Analysis When Cure is a Possibility: A Monte Carlo Study*, Statistics in Medicine, Vol.3, 153-163.
- [14] Greenhouse, J. B. & Wolfe, R. A. (1984) *A Competing Risks Derivation of a Mixture Model for the Analysis of Survival*, Communications in Statistics, Part A - Theory and Methods Vol.13, 3133-3154.
- [15] Hoggart, C.J. & Griffin, J.E. (2000) *A Bayesian Partition Model for Customer Attrition*, Monographs of Official Statistics, Bayesian Methods with Applications to Science, Policy and Official Statistics, 223-232
- [16] Ibrahim, J.G.; Chen, M-H.; & Sinha, D. (2001) *Bayesian Survival Analysis*, Springer-Verlag New York, Inc.
- [17] Kirkwood, J. M.; Strawderman, M. H.; Ernstoff, M. S.; Smith, T. J.; Borden, E. C. & Blum, R. H. (1996) *Interferon Alpha-2b Adjuvant Therapy of High-Risk Resected Cutaneous Melanoma: The Eastern Cooperative Oncology Group Trial EST 1684*, Journal of Clinical Oncology Vol.14, 7-17.
- [18] Kaplan, E.L. & Meier, P. (1958) *Nonparametric Estimation From Incomplete Observation*, Journal of the American Statistical Association Vol.53, 457-81.
- [19] Kuk, A.Y.C & Chen, C-H. (1992) *A Mixture Model Combining Logistic Regression with Proportional Hazards Regression*, Biometria Vol.79, No.3, 531-541.
- [20] Lawless, J. F. (1982) *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.
- [21] Maller, R; Zhou, X. (1996) *Survival Analysis with Long-Term Survivors*, John Wiley & Sons.
- [22] Meeker, W. Q. & Escobar, L. A. (1998) *Statistical Methods for Reliability Data*, John Wiley & Sons, New York.
- [23] Miozi, M.F. (2004) *Influência Local em Modelos de Sobrevida com Fração de Cura*, Tese de Doutorado IME-USP.
- [24] Rodrigues, J.; Cancho, G.V.; Andrade, C.M. & Louzada-Neto, F. (2008) *On The Unification Of The Long-Term Survival Models*
- [25] Schwarz, G. (1978) *Estimating the dimension of a model*, Annals of Statistics Vol.6, 461-464
- [26] Taylor, J.M.G. (1995) *Semi-Parametric Estimation in Failure Time Mixture Models*, Biometrics Vol.51, 899-907.

- [27] Yakovlev, A. Y.; Asselain, B.; Bardou, V. J.; Fourquet, A.; Hoang, T.; Rochefediere, A. & Tsodikov, A. D. (1993) *A Simple Stochastic Model of Tumor Recurrence and Its Applications to Data on Premenopausal Breast Cancer*, Biometrie et Analyses de Donnees Spatio-Temporelles, Vol.12, Société Francaise de Biométrie, ENSA Rennes, France, 66-82.
- [28] Yakovlev, A. Y. & Tsodikov, A. D. (1996) *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. New Jersey: World Scientific.
- [29] Yin, G. & Ibrahim, J.G. (2005) *Cure Rate Models: a unified approach*, The Canadian Journal of Statistics Vol.33, No.4, 559-570.
- Yakovlev & Tsodikov (1996)