



# Modelos de Sobrevivência para Populações com Indivíduos Imunes

Ana Maria Cortesão Pais Figueira da Silva Abreu

Dissertação para obtenção do grau de Doutor em Matemática,  
na especialidade de Probabilidades e Estatística

Orientadora: Cristina Maria Tristão Simões Rocha

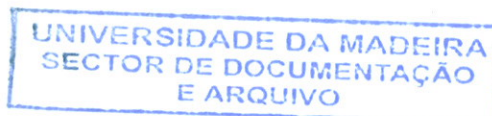
Dezembro de 2004



51  
ABA Mod  
+ C  
F2.2  
T15



# Modelos de Sobrevivência para Populações com Indivíduos Imunes



Ana Maria Cortesão Pais Figueira da Silva Abreu

Dissertação para obtenção do grau de Doutor em Matemática,  
na especialidade de Probabilidades e Estatística

Orientadora: Cristina Maria Tristão Simões Rocha

Dezembro de 2004



# Modelos de Sobrevivência para Populações com Indivíduos Imunes

22 de Março de 2005



## Agradecimentos

Em primeiro lugar gostaria de aqui testemunhar o meu profundo agradecimento à minha orientadora, Professora Doutora Cristina Simões Rocha, cujas qualidades humanas e científicas foram essenciais à realização deste trabalho. A responsabilidade e dedicação que depositou neste projecto, mesmo em momentos pessoalmente difíceis, são um reflexo da sua personalidade.

Também não podia deixar de agradecer ao Departamento de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa, pela forma simpática como sempre me receberam e pelos recursos que me disponibilizaram.

Ao Centro de Ciência e Tecnologia da Madeira, agradeço a bolsa de doutoramento que me concederam.

À Teresa, à Maribel e à Elci, pela amizade.

À Sandra, pela disponibilidade no debate de alguns assuntos.

Ao Sílvio, ao Fermé e à Lisete, pelos esclarecimentos em relação ao  $\LaTeX$ .

Ao Leonel, pela eficiência dos seus comentários informáticos.

À Rita, pela ajuda na preparação das aulas nesta recta final.

Ao Departamento de Matemática e Engenharias da Universidade da Madeira, pela dispensa que me concederam e pelo bom ambiente de trabalho que os seus elementos proporcionam.

Aos meus amigos, por terem compreendido o meu isolamento.

À Céu e à Ana Cristina, pela amizade e pela hospitalidade.

Ao Abreu, pelo companheirismo e pela compreensão.



## Resumo

Esta dissertação visa o estudo de modelos probabilísticos de sobrevivência para populações nas quais se admite a presença de indivíduos curados.

Hoje em dia, à medida que a cura de várias doenças se vai tornando cada vez mais uma realidade, o interesse por este tipo de modelos aumenta.

Assim, vamos estudar modelos adequados para situações nas quais se admite que para alguns indivíduos o acontecimento de interesse nunca ocorre, ainda que o tempo em observação pudesse ser prolongado indefinidamente. Tais indivíduos são considerados indivíduos curados (ou imunes ou não susceptíveis) e, como consequência, a função de sobrevivência que lhes irá corresponder tomará sempre o valor um em qualquer instante. Deste modo, as nossas populações alvo vão ser constituídas por uma mistura de indivíduos imunes e de indivíduos susceptíveis.

Os modelos que vamos considerar, habitualmente designados por modelos de cura, são essencialmente paramétricos, embora façamos referência a alguns não paramétricos. Neste âmbito, introduzimos uma nova distribuição imprópria para um modelo de cura de não-mistura e, num modelo de cura de mistura, sugerimos o uso de uma distribuição ainda não usada neste contexto. Além disso, também dedicamos grande parte do nosso trabalho à questão da estimação dos parâmetros dos modelos, pois neste caso reveste-se de características específicas.

Por outro lado, vamos abordar os modelos de cura do ponto de vista dos riscos competitivos, ou seja, admitindo que existe mais do que uma causa de interesse.

Por fim, vamos estudar modelos com fragilidade no contexto dos modelos de cura. Para tal, definimos uma variável fragilidade, que será nula para os indivíduos imunes e estritamente positiva para os indivíduos susceptíveis.



# Abstract

This dissertation focus on the study of survival probabilistic models for populations where one admits the presence of cured individuals.

Nowadays, as cure becomes more and more a reality in medicine, the interest for survival models which allow for a proportion of cured individuals, naturally grows.

Therefore, we are going to study models in which we allow that for some individuals, the event of interest is never observed, although the follow-up time could be prolonged indefinitely. Those individuals are considered cured (or immune or non susceptible) and, as a consequence, the corresponding survival function will always be one, whatever the instant considered. Then, our target populations will be formed by a mixture of immune individuals and susceptible ones.

The models that we are going to use are essentially parametric models, although we make some references to non parametric ones. In this ambit, we introduce a new improper distribution for a non-mixture cure model and, in a cure mixture model, we suggest the use of a distribution that has not been used yet in this context. We also dedicate a great part of our work to the parameter estimation issue, because in this case there are some specific characteristics.

On the other hand, we will consider survival cure models with competing risks, that is, admitting that there is more than one cause of interest.

Lastly, we are going to study models with frailty in the cure models context. With that purpose, we define a frailty variable, which will be zero for immune individuals and strictly positive for susceptible individuals.



# Índice

<b>Nota Introdutória</b>	<b>i</b>
<b>1 Conceitos preliminares</b>	<b>5</b>
1.1 Introdução . . . . .	5
1.2 Alguns conceitos básicos . . . . .	6
1.3 Função hazard . . . . .	9
1.4 Censura . . . . .	10
1.5 Modelos de regressão . . . . .	12
1.5.1 Modelos com funções hazard proporcionais . . . . .	14
1.5.2 Modelos de tempo de vida acelerado . . . . .	14
1.5.3 Modelos de possibilidades proporcionais . . . . .	15
1.6 Modelos de mistura . . . . .	16
1.7 A função de verosimilhança . . . . .	18
<b>2 Modelos com fragilidade</b>	<b>21</b>
2.1 Introdução . . . . .	21
2.2 Modelo multiplicativo com fragilidade . . . . .	22
2.2.1 Um caso particular . . . . .	25
2.3 Modelo de tempo de vida acelerado com fragilidade . . . . .	31
2.4 Modelo aditivo com fragilidade . . . . .	35
2.4.1 A distribuição $\chi_0^2(\gamma)$ . . . . .	36
<b>3 Análise de Sobrevida com indivíduos imunes: estado da arte</b>	<b>39</b>
3.1 Introdução . . . . .	39
3.2 Descrição histórica: até 1990 . . . . .	41
3.3 Descrição histórica: a partir de 1990 . . . . .	49

<b>4</b>	<b>Modelos de Cura</b>	<b>79</b>
4.1	Introdução . . . . .	79
4.2	Modelos de cura . . . . .	80
4.2.1	Modelo de cura de mistura . . . . .	81
4.2.2	Modelo de cura de não-mistura . . . . .	84
4.2.3	Relação entre os modelos . . . . .	89
4.3	Estimação dos parâmetros . . . . .	90
4.4	Modelo de cura baseado na distribuição log-logística . . . . .	94
4.4.1	Motivação . . . . .	94
4.4.2	Desenvolvimento . . . . .	95
4.4.3	Aplicação do algoritmo EM . . . . .	99
4.5	Um novo modelo de cura paramétrico . . . . .	105
4.5.1	A função de distribuição dos indivíduos susceptíveis . . . . .	105
4.5.2	O novo modelo de cura . . . . .	106
4.5.3	Estimação dos parâmetros . . . . .	109
4.5.4	Aplicações . . . . .	111
<b>5</b>	<b>Modelos de cura na presença de riscos competitivos</b>	<b>115</b>
5.1	Introdução . . . . .	115
5.2	Modelo de riscos competitivos usual . . . . .	116
5.2.1	Tempos de vida potenciais . . . . .	117
5.2.2	Funções específicas de causa . . . . .	119
5.2.3	Modelos de mistura . . . . .	125
5.3	Modelo de cura . . . . .	127
5.3.1	Relação entre os modelos de cura e os riscos competitivos	127
5.3.2	Cura observável . . . . .	128
5.3.3	Cura não observável . . . . .	135
5.3.4	Interligação entre o modelo aditivo com fragilidade, o modelo de cura e o de riscos competitivos . . . . .	141
5.4	Modelo de cura com censura parcialmente informativa . . . . .	142
5.4.1	Motivação . . . . .	142
5.4.2	O modelo de Koziol-Green parcial . . . . .	144
5.4.3	Modelo de cura com censura parcialmente informativa	146
<b>6</b>	<b>Fragilidade no contexto dos modelos de cura</b>	<b>153</b>
6.1	Introdução . . . . .	153
6.2	Modelo com pseudo-fragilidade . . . . .	154
6.3	Modelo de cura com fragilidade . . . . .	158

---

6.3.1	Formulação e resultados gerais . . . . .	158
6.3.2	Distribuição Gaussiana inversa como distribuição para a parte contínua da fragilidade . . . . .	162
6.3.3	Estimação dos parâmetros e aplicação do algoritmo EM	165
6.4	Modelo de tempo de vida acelerado . . . . .	174
<b>A</b>	<b>Algoritmo EM</b>	<b>181</b>
A.1	Introdução . . . . .	181
A.2	Formulação . . . . .	182
A.3	Vantagens e desvantagens . . . . .	183
A.4	O método de Newton-Raphson . . . . .	185
A.5	Algumas variantes do algoritmo EM . . . . .	186
A.5.1	Algoritmo EM generalizado . . . . .	186
A.5.2	Algoritmo EM gradiente . . . . .	187
A.5.3	Um algoritmo EM híbrido . . . . .	188
A.5.4	Algoritmo ECM . . . . .	188
A.5.5	Algoritmo ECM multiciclos . . . . .	189
A.5.6	Outras variantes . . . . .	190
<b>B</b>	<b>Estimação no modelo multiplicativo com fragilidade</b>	<b>191</b>
<b>C</b>	<b>Anexos do capítulo 4</b>	<b>195</b>
C.1	Funções hazard da distribuição Gompertz/Weibull modificada	196
C.2	Estimação no modelo de cura baseado na distribuição log- logística . . . . .	201
C.2.1	Algoritmo EM . . . . .	203
C.2.2	Algoritmo ECM . . . . .	206
C.2.3	Algoritmo híbrido . . . . .	210
C.2.4	Algoritmo EM gradiente . . . . .	212
C.3	Procedimento para a obtenção da variância das estimativas no modelo de cura baseado na distribuição log-logística . . . . .	215
C.3.1	O modelo . . . . .	215
C.4	Funções hazard do modelo de cura baseado na distribuição de Chen . . . . .	222
C.5	Estimação no modelo de cura baseado na distribuição de Chen	225
C.5.1	Algoritmo EM . . . . .	226
C.6	Procedimento para a obtenção da variância das estimativas no modelo de cura baseado na distribuição de Chen . . . . .	229

## Índice

---

C.6.1	O modelo . . . . .	229
<b>D</b>	<b>Estimação no modelo cura com censura parcialmente informativa</b>	<b>235</b>
<b>E</b>	<b>Estimação no modelo de cura com fragilidade</b>	<b>239</b>
E.1	Algoritmo EM . . . . .	240
	<b>Bibliografia</b>	<b>243</b>

# Nota Introdutória

O objectivo desta dissertação é o estudo de modelos de sobrevivência para populações com indivíduos imunes. Estes modelos surgem em contextos variados e pressupõem vários conceitos, daí a variedade de temas que abordámos.

No capítulo 1 fizemos uma pequena introdução à Análise de Sobrevivência, apresentando alguns conceitos preliminares. A variável em estudo é o tempo decorrido desde um instante inicial até à ocorrência de determinado acontecimento. O que caracteriza esta área da Estatística é a existência de censura, isto é, a impossibilidade de observar o acontecimento de interesse para todos os indivíduos durante o período de observação, devido a certas restrições na recolha de dados.

No entanto, há situações em que, de facto, para certos indivíduos o acontecimento de interesse nunca ocorrerá. É neste último contexto que nós nos colocaremos ao longo desta tese.

Uma vez que os modelos com fragilidade se encontram relacionados com os modelos de cura, como veremos no decorrer deste trabalho, introduzimos a Análise de Sobrevivência com fragilidade no capítulo 2. Apresentámos o modelo multiplicativo com fragilidade e uma sua aplicação através da escolha da distribuição Gaussiana inversa para a fragilidade, referimos o modelo de tempo de vida acelerado com fragilidade bem como o modelo aditivo com fragilidade. A variável fragilidade é uma variável não negativa que toma em consideração a heterogeneidade populacional não observável, quer por não existirem covariáveis observáveis no modelo, quer por as consideradas serem insuficientes para explicar a heterogeneidade.

Só no capítulo 3 é que começámos verdadeiramente a abordar o tema desta dissertação. Assim, embora os modelos de cura sejam um tema

relativamente recente, a diversidade de artigos sobre o assunto ou com ele relacionado, levou-nos a fazer uma revisão da bibliografia existente. Sem a pretensão de sermos exaustivos, tendo em conta a acessibilidade dos artigos, o nosso objectivo foi o de dar uma noção das várias formas de abordar este tema.

O capítulo 4 é aquele que melhor revela o propósito do nosso trabalho. Em primeiro lugar apresentámos os dois tipos de modelos de cura (de mistura e de não-mistura), bem como a relação entre eles e os pontos em comum. O que caracteriza qualquer tipo de modelo de cura é o facto de a correspondente função de sobrevivência ser imprópria, isto é,  $S(\infty) > 0$ . Uma outra forma de caracterizar esta propriedade é observar que a função hazard cumulativa é limitada. De uma forma geral, podemos dizer que a primeira caracterização conduz aos modelos de cura de mistura e a segunda aos modelos de cura de não-mistura.

No âmbito dos modelos de não-mistura, propusemos uma nova distribuição imprópria, a que chamamos distribuição de Gompertz/Weibull modificada, por ser baseada na distribuição de Gompertz modificada (função imprópria) e na distribuição de Weibull (função própria). A função hazard desta distribuição revelou-se bastante flexível pois tanto pode ser decrescente como unimodal. Além disso, é perfeitamente consentânea com situações em que existem indivíduos curados uma vez que, para certos valores dos parâmetros, a partir de determinado instante, quase estabiliza num valor próximo de zero.

Nos modelos de cura, importa saber a proporção de indivíduos curados e a distribuição do tempo de vida dos indivíduos não curados (habitualmente designados por susceptíveis). Em geral, a cura é um acontecimento que não é observável; o que acontece é que suspeitamos que um indivíduo que tenha um tempo de vida censurado com valor elevado seja um indivíduo curado. Significa pois que tivemos que trabalhar com dados omissos e, como tal, usar o algoritmo EM no processo de estimação dos parâmetros. Esta foi uma área em que procurámos expor com detalhe, uma vez que a bibliografia existente está bastante dispersa e algo incompleta. Além disso, para tornar a exposição mais clara, exemplificámos com o modelo de cura baseado na distribuição log-logística, aplicando ainda a um exemplo real, de dados já conhecidos da literatura.

No contexto dos modelos de cura de mistura, também propusemos um novo modelo, baseado na distribuição de Chen. Seguimos o mesmo

procedimento que o adoptado para o modelo anterior, inclusivamente no que diz respeito aos dados. O ajustamento efectuado revelou-se significativamente melhor. Ao ajustarmos ainda este modelo a outros dados, os do Kersey *et al.*, também notámos um ajustamento melhor do que o com o modelo de cura baseado na distribuição de Weibull e na distribuição de Burr de tipo XII.

Os riscos competitivos, por si só, constituem uma área recente da Análise de Sobrevivência e, aliados aos modelos de cura, ainda mais recente. Na verdade, a primeira referência que encontramos data de 1998 e, de então para cá, apenas alguns artigos têm surgido. O nosso contributo no capítulo 5 foi essencialmente mostrar o trabalho já feito, alertar para eventuais problemas relativamente à definição de cura neste contexto e revelar uma relação entre os riscos competitivos, os modelos de cura e um caso particular do modelo aditivo com fragilidade. Além disso, também propusemos um modelo de cura de não-mistura com censura parcialmente informativa.

Por fim, no capítulo 6, começámos por desenvolver um modelo de cura de mistura com pseudo-fragilidade, baseados em Sy e Taylor (2000). Em seguida, generalizámos o modelo, considerando que os indivíduos susceptíveis não o eram em igual grau. Deste modo, como para cada indivíduo não é possível quantificar o seu grau de susceptibilidade, definimos uma variável fragilidade que é nula para os indivíduos curados e é estritamente positiva para os indivíduos susceptíveis, ou seja, trata-se de uma variável aleatória mista. Assim, para a parte contínua da variável, isto é, para os indivíduos susceptíveis, consideramos o modelo multiplicativo com fragilidade em que a distribuição da fragilidade é a distribuição Gaussiana inversa. Em seguida, procedemos à estimação dos parâmetros, uma vez mais, através da aplicação do algoritmo EM. Paralelamente ao nosso trabalho, surgiram alguns resultados comuns, embora com bastante menos desenvolvimento.

Na sequência do que já tínhamos feito para os modelos com fragilidade, consideramos ainda o modelo de cura em que, para os indivíduos susceptíveis, adoptamos um modelo de tempo de vida acelerado com fragilidade.



# Capítulo 1

## Conceitos preliminares

### 1.1 Introdução

Análise de Sobrevivência é o termo utilizado para descrever a análise estatística de dados referentes a medições do tempo decorrido desde um instante inicial até à ocorrência de determinado acontecimento, numa população. Assim, as observações são tempos de vida (que podem ser medidos em horas, semanas, meses, ou outros), onde tempo de vida tem um sentido muito vasto pois pode representar nomeadamente, o tempo decorrido até à morte, à falha de componentes mecânicas ou electrónicas, à manifestação de certos sintomas e ao aparecimento de metástases. Este tipo de análise estatística merece uma atenção particular devido a alguns factores, nomeadamente:

1. os dados exibem, em geral, uma assimetria positiva tornando inadequado supor-se que têm distribuição normal;
2. os dados são frequentemente censurados tornando inviável uma análise clássica dos mesmos.

Esta última característica é preponderante na escolha deste tipo de análise e corresponde à impossibilidade de observar o acontecimento de interesse durante o período em que o indivíduo está em observação. No entanto, habitualmente supõe-se que o acontecimento de interesse ocorre sempre para todos os indivíduos; pode é ser observado ou não.

Mais recentemente, admite-se que, para alguns dos indivíduos com tempo de vida censurado, o acontecimento de interesse nunca ocorrerá. Tais

indivíduos são designados por indivíduos curados ou imunes. Este tipo de abordagem será o cerne do nosso trabalho.

Para caracterizar o tempo de vida existem duas funções que merecem destaque especial: a função de sobrevivência e a função hazard.

Neste capítulo, começaremos por introduzir alguns conceitos básicos relacionados com as funções mais usadas nesta área. Em seguida, abordamos o tema da censura, com vista a esclarecer algumas noções que são pouco referidas ou pouco claras na maior parte da bibliografia existente. Os modelos de regressão mais usuais também integram este capítulo, bem como os modelos de mistura finita. Por fim, fazemos referência à construção da função de verosimilhança num modelo paramétrico, em que os tempos de vida estão sujeitos a censura à direita.

## 1.2 Alguns conceitos básicos

O tempo de vida (ou tempo de sobrevivência) de um indivíduo de uma dada população homogénea pode ser encarado como o valor que uma variável aleatória (v.a.)  $T$  toma, sendo essa variável não negativa. A variável  $T$  pode ser discreta ou contínua. No entanto, os resultados apresentados nos capítulos seguintes referem-se apenas a variáveis contínuas.

Define-se a função de sobrevivência (f.s.),  $S(t)$ , da v.a.  $T$  por

$$S(t) = P(T \geq t) = 1 - F(t), \quad t \geq 0, \quad (1.1)$$

a qual representa a probabilidade de um indivíduo sobreviver pelo menos até ao instante  $t$ , onde  $F(t)$  é a função de distribuição correspondente. Das propriedades da função de distribuição, resulta que  $S(t) = 1$  para  $t = 0$  e  $S(t) = 0$  quando  $t \rightarrow +\infty$ ; além disso,  $S(t)$  é uma função monótona não crescente e contínua à esquerda.

Começemos por considerar  $T$ , uma v.a. absolutamente contínua com função densidade de probabilidade (f.d.p.)  $f(t)$ . Uma questão frequente na Análise de Sobrevivência é, sabendo que um indivíduo está vivo num determinado instante qual a probabilidade de morrer no instante seguinte, isto é, qual a sua taxa instantânea de morte ou taxa instantânea de falha. A resposta é-nos dada pela função hazard, que no caso contínuo é definida do seguinte modo

$$h(t) = \lim_{\delta t \rightarrow 0} \left[ \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \right] \quad (1.2)$$

e que verifica as seguintes propriedades:

- $h(t) \geq 0, \forall t \geq 0$ ;
- $\int_0^{+\infty} h(t) dt = +\infty$ .

De (1.1) e (1.2) podemos deduzir várias relações entre  $f(t)$ ,  $S(t)$  e  $h(t)$ ; em particular,

$$h(t) = \frac{f(t)}{S(t)}, \quad (1.3)$$

$$S(t) = \exp \left( - \int_0^t h(u) du \right), \quad (1.4)$$

$$f(t) = h(t) \exp \left( - \int_0^t h(u) du \right). \quad (1.5)$$

Define-se ainda a função hazard cumulativa ou hazard integrada por

$$H(t) = \int_0^t h(u) du,$$

donde, temos ainda  $H(t) = -\log S(t)$ .

Das relações apresentadas podemos concluir que a distribuição de  $T$  fica univocamente determinada a partir de qualquer uma das seguintes funções: função de sobrevivência, função densidade ou função hazard.

Se, quando  $t \rightarrow +\infty$ ,  $S(t) > 0$ , dizemos que a função de sobrevivência é imprópria. Sendo esta uma situação pouco habitual, ela ocorre quando a população em estudo inclui indivíduos nos quais nunca vai ser observado o acontecimento de interesse. Esta alteração na função de sobrevivência dá origem a modificações nas propriedades das funções com ela

relacionadas, nomeadamente, na função hazard cumulativa. Assim, enquanto que anteriormente tínhamos que  $H(\infty) = \infty$ , agora esta função será limitada, isto é,  $H(\infty) = \theta$ , com  $\theta < \infty$ . Deste modo, embora nesta situação não tenhamos funções densidade e hazard no sentido habitual, faremos o abuso de linguagem de as designar pelo mesmo nome.

Vejam os como se definem as funções anteriores quando temos uma variável discreta. Seja então  $T$  uma v.a. discreta, a tomar valores  $t_0, t_1, t_2, \dots, t_k$ ,  $t_0 = 0$ , onde  $k$  pode ser finito ou infinito, assim como  $t_k$ . A função massa de probabilidade (f.m.p.) de  $T$  é definida por

$$f(t_i) = P(T = t_i), \quad i = 1, 2, \dots, k$$

e a função de sobrevivência é dada por

$$S(t) = \sum_{i:t_i \geq t} f(t_i).$$

A função hazard no instante  $t_i$  define-se como

$$h_i = P(T = t_i | T \geq t_i) = \frac{f(t_i)}{S(t_i)} = \frac{S(t_i) - S(t_{i+1})}{S(t_i)}$$

logo, a f.s. também pode ser escrita na forma

$$S(t) = \prod_{i:t_i < t} (1 - h_i).$$

Note-se que  $0 \leq h(t) \leq 1, \forall t$ , onde  $h(t) = 0$  excepto nos instantes  $t_1, t_2, \dots, t_k$ ,  $h(0) = 0$  e  $h(t)$  só é igual a 1 em  $t_k$ .

Para que a relação  $S(t) = \exp[-H(t)]$  se mantenha verdadeira também para o caso discreto, a função hazard cumulativa é definida por

$$H(t) = - \sum_{i:t_i < t} \log(1 - h_i).$$

No entanto, uma definição alternativa é dada por

$$H(t) = \sum_{i:t_i < t} h_i,$$

a qual fornece uma boa aproximação para a função hazard cumulativa se os  $h_i$  forem pequenos, pois nesse caso  $\log(1 - h_i) \approx -h_i$ .

Por último, vejamos o caso em que  $T$  é uma v.a. mista. Admita-se que  $T$  tem massa de probabilidade não nula em  $t_1, t_2, \dots, t_k$ . Designe-se por  $f_c$  e por  $h_c$  a função densidade e a função hazard, respectivamente, correspondente à parte contínua da distribuição. A função de sobrevivência de  $T$  é definida por

$$S(t) = \left[ \prod_{i:t_i < t} (1 - h_i) \right] \exp \left[ - \int_0^t h_c(u) du \right]$$

ou, alternativamente, por

$$S(t) = \sum_{t_i \geq t} P(T = t_i) + \int_t^\infty f_c(s) ds$$

A função hazard escreve-se na forma

$$h(t)dt = \sum_{i \geq 1} h_i \delta(t - t_i) dt + h_c(t) dt,$$

onde  $\delta(\cdot)$  designa a função delta de Dirac, e a função hazard cumulativa correspondente é dada por

$$H(t) = \sum_{i:t_i < t} h_i + \int_0^t h_c(u) du.$$

### 1.3 Função hazard

Quando estamos a lidar com tempos de vida, a função hazard é extremamente útil pois fornece-nos uma interpretação directa da realidade. De facto, por traduzir a evolução ao longo do tempo da probabilidade instantânea de morte de um indivíduo, pode levar-nos a conclusões sobre a evolução do padrão de mortalidade. Além disso, informação qualitativa sobre a função hazard, pode ser-nos útil na escolha de uma família de modelos pois, em particular, esta função pode ser:

- monótona crescente - mais comum; surge quando os indivíduos são observados num período da sua vida em que ocorre um envelhecimento gradual, pois traduz o facto de a proporção de indivíduos que morrem num dado instante, de entre os sobreviventes nesse instante, ir aumentando com o decorrer do tempo;

- monótona decrescente - menos comum; ocorre, por exemplo, após um transplante de órgãos ou um enfarte cardíaco, em que o risco de morte vai diminuindo com o decorrer do tempo; também é adequada se a população em estudo for uma mistura de indivíduos susceptíveis e não susceptíveis ao acontecimento de interesse;
- constante - mais adequada para o tempo de vida de componentes electrónicas; no entanto, os indivíduos cujos únicos riscos de morte sejam acidentes ou doenças raras também podem apresentar este tipo de função hazard;
- *bathhtub-shaped* - adequada quando os indivíduos de uma população são observados desde o nascimento até à morte reais;
- unimodal - característica de situações em que o risco de morte aumenta durante um certo período de tempo e depois decresce. Quando a população é constituída por indivíduos susceptíveis e não susceptíveis ao acontecimento de interesse, é frequente que, a partir de determinado instante, a função hazard se torne praticamente constante.

## 1.4 Censura

Uma característica muito particular dos dados de sobrevivência é a possibilidade de serem censurados. A censura ocorre quando não é possível observar o acontecimento de interesse durante o período em que o indivíduo está em observação. Por exemplo, alguns pacientes podem ainda estar vivos no fim do estudo; o tempo de sobrevivência de tais indivíduos é desconhecido, sabemos apenas que excede determinado valor. Temos então observações censuradas. Um outro exemplo de observações censuradas ocorre quando para alguns dos indivíduos em estudo, não é possível saber se estão vivos ou mortos, isto é, os indivíduos consideram-se perdidos para o *follow-up* e é registada apenas a data do último contacto. Por fim, temos ainda uma outra situação: suponhamos que um indivíduo participa num estudo com vista a analisar o tempo de sobrevivência de indivíduos com cancro do pulmão; se o indivíduo morrer antes do fim do estudo mas devido a uma outra causa, também temos uma observação censurada.

Em qualquer dos casos anteriores, um indivíduo entra no estudo no instante  $t_0$  e morre no instante  $t_0 + t$ , onde  $t$  é desconhecido. Sabe-se apenas

que o indivíduo estava vivo em  $t_0 + c$  ( $c < t$ ) onde  $c$  representa um tempo de sobrevivência censurado. Temos então o que se designa por censura à direita.

Uma outra forma de censura é a censura à esquerda, que se traduz por o tempo de sobrevivência de um indivíduo ser menor que o tempo observado, ou seja, significa que o acontecimento de interesse já ocorreu quando o indivíduo é observado no instante de censura  $c_e$ . Consideremos a situação de mulheres sujeitas a mastectomia devido a cancro da mama e suponhamos que a variável em estudo é o tempo decorrido desde a intervenção cirúrgica até ao aparecimento de metástases; se ao fim de 3 meses essas mulheres forem sujeitas a um rastreio, relativamente às pacientes que apresentam metástases, tudo o que podemos dizer é que o tempo até à ocorrência de metástases é menor que 3 meses.

Temos ainda a censura intervalar, que se caracteriza por se saber apenas que o acontecimento de interesse ocorreu num determinado intervalo. Este tipo de censura é frequente em situações de ensaios clínicos em que os doentes efectuem visitas regulares ao médico.

Note-se que quer a censura à direita quer a censura à esquerda são casos particulares da censura intervalar.

Na Análise de Sobrevivência, a censura à direita é a mais comum; de aqui em diante utilizaremos o termo censura para designar censura à direita, pois é o único caso que iremos considerar.

A maior parte dos estudos em Análise de Sobrevivência com censura à direita, pressupõe a designada censura aleatória clássica ou simples. Neste tipo de censura admite-se que as variáveis que representam os tempos de vida censurados,  $U_1, \dots, U_n$ , são independentes e idênticamente distribuídas e independentes das variáveis  $X_1, \dots, X_n$ , correspondentes aos verdadeiros tempos de vida. A censura aleatória geral é uma generalização deste conceito. Neste caso,  $\mathbf{U} = (U_1, \dots, U_n)$  é independente de  $\mathbf{X} = (X_1, \dots, X_n)$  e, além disso,  $\mathbf{U}$  pode ter uma distribuição arbitrária.

Uma questão crucial que é preciso ter em conta quando existe censura à direita é o facto de a censura poder vir a alterar o risco de ocorrência do acontecimento de interesse. Se, por exemplo, num ensaio clínico os pacientes que se encontram muito doentes ou já bastante saudáveis são retirados do estudo, então os pacientes que permanecem em risco já não são representativos da amostra dos pacientes que teríamos se não tivesse havido censura. Os mecanismos de censura à direita que evitam este tipo de problemas e que, de certa forma, mantêm o conjunto de risco representativo do que seria se

não tivesse havido censura são designados por independentes. Assim, para que os métodos habituais em Análise de Sobrevivência sejam válidos temos que ter um mecanismo de censura independente, ou seja, tem que ser observada a seguinte condição: o tempo de sobrevivência de um indivíduo,  $t$ , é independente de qualquer mecanismo que cause a sua censura no instante  $c$ , onde  $c < t$ . Por outras palavras, se considerarmos um grupo homogêneo de indivíduos, um indivíduo que seja censurado no instante  $c$  deve ser representativo (relativamente ao risco de morte) de todos os indivíduos do grupo que sobreviveram até esse instante.

Suponhamos que estamos a estudar o tempo de vida de doentes com leucemia sujeitos a um tratamento. Definindo o tempo de vida como o tempo desde o tratamento até à morte por leucemia, as mortes que se verificarem devido a outra causa mas relacionada com a leucemia pode ser considerado um exemplo de censura informativa. Dito de outra forma, o tempo de vida dos doentes que morrem devido a outra causa relacionada com a leucemia será considerado um tempo de vida censurado mas que engloba alguma informação acerca da própria doença. Deste modo, a distribuição do tempo de censura também dependerá do parâmetro da distribuição do verdadeiro tempo de vida. Um exemplo desta situação é o modelo de Koziol-Green, proposto por Koziol e Green (1976), no qual as funções hazard das duas distribuições são proporcionais. Embora habitualmente a censura seja não-informativa, existem situações em que não é possível evitar a censura informativa. Nesses casos, é necessário adoptar métodos específicos.

## 1.5 Modelos de regressão

De um modo geral, os indivíduos em estudo constituem um grupo heterogêneo, pois diferem entre si relativamente a factores passíveis de influenciar o seu tempo de vida. Define-se então para cada indivíduo um vector  $\mathbf{z}$  de variáveis chamadas explanatórias, concomitantes ou covariáveis, representando esses factores, ou seja, as componentes de  $\mathbf{z}$  representam várias características capazes de influenciar o tempo de vida do indivíduo, nomeadamente:

- tratamentos (comparação dos efeitos de dois ou mais tratamentos, estudo do efeito de um novo medicamento);

- propriedades intrínsecas dos indivíduos (variáveis demográficas tais como idade ou sexo, ou variáveis que descrevem a história clínica do indivíduo antes da entrada no estudo);
- variáveis exógenas.

Tais variáveis explanatórias podem ser constantes ou dependentes do tempo. Se estivermos interessados em estudar o efeito de um medicamento, a covariável é apenas uma variável indicatriz do grupo de tratamento e é constante. Mas, se à situação anterior acrescentarmos o facto de o medicamento só ser administrado ao fim de um certo período de tempo após o início do estudo, já teremos que definir duas componentes para o vector de covariáveis: a primeira, a variável indicatriz do grupo de tratamento, e a segunda, dependente do tempo, a variável indicatriz do momento de administração do medicamento.

As covariáveis podem ser observadas (idade, sexo, ...) ou não observadas, caso em que se considera a variável fragilidade, de que falaremos mais à frente.

Uma forma de estudar a relação entre as variáveis explanatórias e o tempo de vida é através de um modelo de regressão, no qual a distribuição do tempo de vida depende das covariáveis. Isto implica especificar um modelo para a distribuição de  $T$ , dado  $\mathbf{z}$ , onde  $T$  representa o tempo de vida e  $\mathbf{z} = (z_1, z_2, \dots, z_k)$  é um vector  $1 \times k$  de variáveis explanatórias associadas a um indivíduo. Aparecem assim os modelos de regressão, em que o tempo de vida é a variável dependente e as covariáveis são as variáveis independentes. Os modelos de regressão podem ser classificados como modelos paramétricos, não paramétricos ou semi-paramétricos. Os modelos paramétricos são construídos a partir de distribuições específicas, habitualmente contínuas univariadas, nomeadamente as distribuições exponencial, Weibull, log-normal, log-logística, Gompertz, gama e Gaussiana inversa, entre outras. Os modelos não paramétricos, não se baseiam em qualquer tipo de distribuição paramétrica. Os modelos semi-paramétricos, dos quais se destaca o modelo de regressão de Cox (1972), não supõem uma distribuição paramétrica para o tempo de vida da população subjacente; no entanto, são admitidas conjecturas do tipo paramétrico para caracterizar o efeito das covariáveis no tempo de vida.

Os modelos de regressão mais utilizados em Análise de Sobrevivência são os modelos com funções hazard proporcionais, os modelos de tempo de vida

acelerado e os modelos de possibilidades proporcionais.

### 1.5.1 Modelos com funções hazard proporcionais

Sejam  $\mathbf{z}_1$  e  $\mathbf{z}_2$  os vectores de covariáveis associadas a dois indivíduos. Dize-mos que estamos perante um modelo de hazards proporcionais se a razão  $h(t; \mathbf{z}_1)/h(t; \mathbf{z}_2)$  não depende de  $t$ , isto é, se dados dois indivíduos as suas funções hazard no mesmo instante, são proporcionais. Seja  $\mathbf{z} = \mathbf{0}$  o vector de covariáveis correspondente à situação padrão e designemos por  $h_0(t)$  a função hazard subjacente, ou seja, a função hazard de um indivíduo cujo vector de covariáveis corresponde à situação padrão. Então a função hazard de  $T$ , dado  $\mathbf{z}$ , pode escrever-se na forma

$$h(t; \mathbf{z}) = h_0(t)g(\mathbf{z}),$$

onde  $g(\cdot)$  é tal que  $g(\mathbf{0}) = 1$ . Como a função hazard toma sempre valores não negativos,  $g(\mathbf{z})$  pode, por exemplo, ser parametrizada por  $g(\mathbf{z}) = \exp(\mathbf{z}\boldsymbol{\beta})$  onde  $\mathbf{z}\boldsymbol{\beta} = z_1\beta_1 + z_2\beta_2 + \dots + z_k\beta_k$ , e os coeficientes  $\beta_m$ ,  $m = 1, 2, \dots, k$ , designam-se por coeficientes de regressão. Notemos que as covariáveis têm um efeito multiplicativo na função hazard.

Se o modelo for caracterizado pelas funções de sobrevivência, temos

$$S(t; \mathbf{z}) = S_0(t)^{g(\mathbf{z})},$$

onde  $S_0(t) = \exp\left(-\int_0^t h(u)du\right)$  é a f.s. de um indivíduo para o qual  $\mathbf{z} = \mathbf{0}$ .

O modelo de regressão de Cox é um modelo de hazards proporcionais em que  $h_0(t)$  não é especificada.

### 1.5.2 Modelos de tempo de vida acelerado

Neste modelo, também designado por modelo de localização-escala para  $\log T$  ou modelo log-linear para  $T$ , as variáveis explanatórias têm um efeito multiplicativo em  $T$ , logo influenciam a razão à qual um indivíduo avança ao longo do tempo. Isto significa que o modelo pode ser interpretado em termos da velocidade de progressão da doença, ou seja, tem uma interpretação directa em termos clínicos.

Considerando  $Y = \log T$ , dado  $\mathbf{z}$ ,  $Y$  tem uma distribuição com parâmetro de localização  $\mu(\mathbf{z})$  e parâmetro de escala  $\sigma > 0$ . O modelo pode então ser escrito na forma

$$Y = \mu(\mathbf{z}) + \sigma\varepsilon,$$

onde  $\varepsilon$  tem uma distribuição que é independente de  $\mathbf{z}$ .

A função de sobrevivência de  $T = \exp(Y)$ , verifica a relação

$$S(t; \mathbf{z}) = S_0[t\alpha(\mathbf{z})],$$

onde  $\alpha(\mathbf{z}) = \exp[-\mu(\mathbf{z})]$ . Assim, a razão entre o tempo de sobrevivência de um indivíduo com vectores de covariáveis  $\mathbf{z}$  e o tempo de sobrevivência de um indivíduo com vectores de covariáveis  $\mathbf{z} = \mathbf{0}$  é  $1/\alpha(\mathbf{z})$ . O factor  $\alpha(\mathbf{z})$  é designado por factor de aceleração e, consoante  $0 < \alpha(\mathbf{z}) < 1$  ou  $\alpha(\mathbf{z}) > 1$ , o efeito das covariáveis é travar ou acelerar, respectivamente, o tempo até à ocorrência do acontecimento de interesse.

A função hazard e a f.d.p. são dadas, respectivamente, por

$$h(t; \mathbf{z}) = \alpha(\mathbf{z})h_0(t\alpha(\mathbf{z}))$$

e

$$f(t; \mathbf{z}) = \alpha(\mathbf{z})f_0(t\alpha(\mathbf{z})).$$

O modelo de regressão Weibull e o modelo de regressão baseado na distribuição log-logística constituem exemplos deste tipo de modelos, sendo o primeiro simultaneamente o único exemplo que é também um modelo de hazards proporcionais.

### 1.5.3 Modelos de possibilidades proporcionais

Define-se possibilidade (*odds*) de sobrevivência para além do instante  $t$ , como sendo a razão

$$\frac{S(t)}{1 - S(t)}.$$

Nesta classe de modelos, as funções de sobrevivência relacionam-se da seguinte forma

$$\frac{S(t; \mathbf{z})}{1 - S(t; \mathbf{z})} = \exp(\mathbf{z}\boldsymbol{\beta}) \frac{S_0(t)}{1 - S_0(t)},$$

onde  $\mathbf{z}\boldsymbol{\beta} = \beta_1 z_1 + \dots + \beta_k z_k$  é uma combinação linear das  $k$  covariáveis  $z_1, \dots, z_k$ , e  $S_0(t)$  é a função de sobrevivência subjacente, correspondente a um indivíduo para o qual  $\mathbf{z} = \mathbf{0}$ . Vemos assim que as covariáveis actuam de forma multiplicativa na possibilidade de um indivíduo sobreviver para além do instante  $t$ .

Uma propriedade importante do modelo de possibilidades proporcionais diz respeito à razão  $h(t; \mathbf{z})/h_0(t)$ . Como esta razão se pode escrever na forma

$$\frac{h(t; \mathbf{z})}{h_0(t)} = [1 + (\exp(\mathbf{z}\boldsymbol{\beta}) - 1) S_0(t)]^{-1},$$

verificamos que a razão das funções hazard converge para  $\exp(-\mathbf{z}\boldsymbol{\beta})$  quando  $t \rightarrow 0$  e para a unidade quando  $t \rightarrow \infty$ . Isto significa que as funções hazard correspondentes a indivíduos com diferentes valores das covariáveis convergem ao fim de um certo tempo. Esta é uma situação que ocorre quando o efeito de uma covariável na sobrevivência dos indivíduos decresce com o tempo.

O modelo de regressão baseado na distribuição log-logística, além de pertencer a esta família de modelos, tem a particularidade única de pertencer simultaneamente à família dos modelos de tempo de vida acelerado.

## 1.6 Modelos de mistura

Uma outra forma de modelar o tempo de vida em populações heterogéneas é através de modelos de mistura. Neste caso, há uma característica que vai permitir a identificação dos vários subgrupos da população, característica essa que pode ser directamente observável ou não.

Teremos uma mistura finita de distribuições quando a função densidade  $f(t)$  de uma variável aleatória  $T$  puder ser escrita na forma

$$f(t) = \sum_{j=1}^g p_j f_j(t) \quad (0 \leq p_j \leq 1, \sum_{j=1}^g p_j = 1). \quad (1.6)$$

As funções  $f_j(t)$  são as densidades componentes da mistura e as quantidades  $p_j$  são designadas por proporções ou pesos de mistura. O número de componentes  $g$  pode ser um valor fixo, isto é, conhecido à partida, ou mais um parâmetro a ser estimado a partir da amostra. A função de sobrevivência correspondente a (1.6) é dada por

$$S(t) = \sum_{j=1}^g p_j S_j(t).$$

Uma situação em que este tipo de modelo é directamente aplicável surge quando a população em estudo é formada por  $g$  grupos,  $G_1, \dots, G_g$ , nas proporções  $p_1, \dots, p_g$ . Se a densidade de  $T$  no grupo  $G_j$  é dada por  $f_j(t)$  para  $j = 1, \dots, g$ , então a densidade de  $T$  na população tem a forma (1.6). Neste caso, as  $g$  componentes da mistura podem ser identificadas fisicamente com os  $g$  grupos da população,  $G_1, \dots, G_g$ .

Em muitas aplicações as densidades componentes da mistura  $f_j(t)$  pertencem a alguma família paramétrica. Assim, a sua representação passa a ser  $f_j(t; \boldsymbol{\theta}_j)$ , onde  $\boldsymbol{\theta}_j$  designa o vector dos parâmetros desconhecidos da  $j$ -ésima distribuição. Deste modo, o modelo (1.6) representa-se por

$$f(t; \boldsymbol{\Psi}) = \sum_{j=1}^g p_j f_j(t; \boldsymbol{\theta}_j), \quad (1.7)$$

onde  $\boldsymbol{\Psi}$  é o vector que contém todos os parâmetros desconhecidos do modelo de mistura e que pode ser definido como  $\boldsymbol{\Psi} = (p_1, \dots, p_{g-1}, \boldsymbol{\xi}^T)^T$ , sendo  $\boldsymbol{\xi}$  o vector que contém os parâmetros  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g$ .

Seja  $T$  uma v.a. que representa o tempo de vida de um indivíduo numa dada população, na qual se admite existirem indivíduos imunes e indivíduos susceptíveis. Seja  $S_d(t)$  a função de sobrevivência do tempo de vida dos indivíduos susceptíveis e  $p$  a proporção de indivíduos imunes na população. A função de sobrevivência de  $T$  (populacional) é

$$S(t) = p + (1 - p)S_d(t). \quad (1.8)$$

Este caso particular dos modelos de mistura é designado por modelo de *mistura não padrão*, ou seja, refere-se ao caso  $g = 2$  em que uma das componentes é degenerada. No contexto da análise de sobrevivência, o modelo (1.8) é designado abreviadamente por *modelo de cura*, embora mais rigorosamente seja um *modelo de cura de mistura*, pois existe outro tipo de modelos de cura, como veremos mais adiante. O estudo dos modelos de cura de mistura constituem o objectivo principal deste trabalho.

## 1.7 A função de verosimilhança

Seja  $T$  uma v.a. para a qual admitimos um modelo paramétrico, ou seja, consideramos que a distribuição do tempo de vida de  $T$  é conhecida a menos de um vector de parâmetros, digamos  $\theta$ . Como tal, a construção da função de verosimilhança tem por objectivo a realização de inferência estatística sobre o vector de parâmetros.

Ao escrever a função de verosimilhança para observações sujeitas a censura, admite-se habitualmente que os tempos de vida e os tempos de censura são independentes e que a censura é não informativa. Além disso, é necessário ter em conta qual o tipo de censura a que os indivíduos estão sujeitos (censura à direita, à esquerda, intervalar) pois para cada caso a função de verosimilhança escrever-se-à de forma diferente. Vamos apresentar apenas a situação de censura à direita por ser aquela que vamos considerar ao longo deste trabalho.

Começamos por notar que uma observação correspondente a um tempo de vida exacto, fornece informação sobre a probabilidade de ocorrer o acontecimento de interesse nesse instante, a qual é aproximadamente igual à função densidade da variável tempo de vida, nesse instante. Para uma observação censurada à direita, tudo o que sabemos é que o verdadeiro tempo de vida é superior ao tempo de vida observado, donde a informação é transmitida através da função de sobrevivência nesse instante.

Formalmente, esta situação pode ser caracterizada através de quatro variáveis. A variável correspondente ao verdadeiro tempo de vida,  $X$ , a correspondente ao tempo de censura,  $C$ , a correspondente ao tempo de vida observado,  $T$ , e a variável indicatriz,  $\delta$ , que toma o valor um se  $X$  é observado e o valor zero se  $C$  é observado. Dito de outra forma,  $\delta = 1$  se  $X \leq C$  e  $\delta = 0$  se  $X > C$  e  $T = \min(X, C)$ . Sejam ainda  $g(c)$  e  $\overline{G}(c)$  a f.d.p. e a f.s., respectivamente, de  $C$ , e  $f(x)$  e  $S(x)$  a f.d.p. e a f.s., respectivamente, de  $X$ .

Consideremos então uma amostra aleatória  $(t_1, \delta_1), \dots, (t_n, \delta_n)$ . A função densidade do par  $(t, \delta)$  pode ser obtida através da função densidade conjunta de  $X$  e  $C$ , sendo dada por

$$P(T = t, \delta = 0) = P(C = t, X > C)$$

e por

$$P(T = t, \delta = 1) = P(X = t, X \leq C).$$

Todavia, se as variáveis  $X$  e  $C$  são independentes, então

$$P(T = t, \delta = 0) = S(t)g(t)$$

e

$$P(T = t, \delta = 1) = f(t)\overline{G}(t).$$

Assim, a função de verosimilhança é

$$L = \prod_{i=1}^n [f(t_i)\overline{G}(t_i)]^{\delta_i} [g(t_i)S(t_i)]^{1-\delta_i},$$

ou ainda,

$$L = \left[ \prod_{i=1}^n g(t_i)^{1-\delta_i} \overline{G}(t_i)^{\delta_i} \right] \left[ \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \right]$$

Se a distribuição do tempo de censura não depende do vector de parâmetros de interesse, ou seja, se a censura é não informativa, a função de verosimilhança tem uma expressão mais simples:

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

ou, alternativamente,

$$L = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i).$$

Em geral, os métodos de inferência estatística usados na Análise de Sobrevivência são baseados na teoria assintótica da máxima verosimilhança, apresentando resultados válidos sob condições de regularidade bastante gerais nos processos de morte e de censura. Deste modo, o estimador de máxima verosimilhança  $\hat{\boldsymbol{\theta}}$  tem distribuição assintótica normal multivariada com valor médio  $\boldsymbol{\theta}$  e matriz de covariância  $I(\boldsymbol{\theta})^{-1}$ , onde  $I(\boldsymbol{\theta})$  representa a matriz de informação de Fisher.



# Capítulo 2

## Modelos com fragilidade

### 2.1 Introdução

A heterogeneidade existente entre os indivíduos de uma população, é considerada através da introdução de covariáveis observáveis no modelo. No entanto, em qualquer estudo da natureza humana (e não só), está sempre presente uma certa heterogeneidade, não observável. Este facto constituiu uma forte motivação para o aparecimento dos modelos com fragilidade.

Vaupel *et al.* (1979) introduziram o termo fragilidade para designar uma variável não observada que descreve factores de risco, desconhecidos ou não mensuráveis, não incluídos no modelo. De acordo com Hougaard (1995), a ideia dos modelos com fragilidade para dados de tempos de vida consiste em considerar que a variabilidade observada nos tempos de vida é proveniente de duas fontes distintas. Uma, que é simples aleatoriedade, é descrita pela função hazard. A outra, é descrita por uma variável aleatória, designada por fragilidade, e que pode representar ou uma variável individual, ou uma variável comum a um certo grupo de indivíduos. No primeiro caso, temos a situação univariada, onde a fragilidade descreve a heterogeneidade existente entre os indivíduos, ou seja, a influência dos factores de risco não observáveis num modelo de sobrevivência. No caso da variável ser comum a um grupo de indivíduos (como, por exemplo, gémeos), temos a situação multivariada, e a fragilidade reflecte a dependência entre os indivíduos do grupo. O modelo multivariado também pode ser utilizado quando existem várias observações para um mesmo indivíduo.

Neste capítulo, vamos referir o modelo multiplicativo com fragilidade, o

modelo de tempo de vida acelerado com fragilidade e o modelo aditivo com fragilidade, visto que serão utilizados posteriormente no âmbito dos modelos de cura.

## 2.2 Modelo multiplicativo com fragilidade

Seja  $T$  uma variável aleatória não negativa que representa o tempo de vida dos indivíduos. O modelo multiplicativo proposto por Vaupel *et al.* (1979), escrito à custa da função hazard no instante  $t$  para um indivíduo com fragilidade  $W = w$ , é da forma

$$\mu(t|w) = w\lambda(t), \quad (2.1)$$

onde  $W$  é uma variável aleatória não negativa (não observada) e  $\lambda(t)$  é uma função do tempo comum a todos os indivíduos e independente de  $W$ . Note-se que  $\lambda(t)$  é a função hazard correspondente aos indivíduos com fragilidade  $w = 1$ . A função de sobrevivência condicional de  $T$  dado  $W = w$  é

$$S(t|w) = \exp[-w\Lambda(t)] = [S_0(t)]^w, \quad (2.2)$$

onde  $\Lambda(t) = \int_0^t \lambda(u)du$  é a função hazard cumulativa para um indivíduo com fragilidade  $w = 1$ , que designaremos por função hazard cumulativa subjacente. Assim,  $S_0(t) = \exp[-\Lambda(t)]$  é a correspondente função de sobrevivência subjacente, associada a esse mesmo indivíduo padrão. Embora (2.2) seja equivalente na forma ao modelo de hazards proporcionais proposto por Cox (1972), os métodos convencionais da Análise de Sobrevivência deixam de ser válidos se o valor  $w$  não for observado, conforme refere Oakes (1989).

De acordo com Hougaard (1984), a função de sobrevivência e a função hazard populacionais são, respectivamente,

$$S(t) = \int_0^\infty S(t|w)dF(w) = \int_0^\infty \exp[-w\Lambda(t)] dF(w) = L[\Lambda(t)], \quad (2.3)$$

$$h(t) = -\frac{d}{dt} [\log S(t)] = -\frac{L'[\Lambda(t)]}{L[\Lambda(t)]}\lambda(t), \quad (2.4)$$

onde  $L(s)$  é a transformada de Laplace da distribuição da fragilidade  $W$ ,  $F(w)$  é a correspondente função de distribuição e  $L'(s) = dL(s)/ds$ .

Dado que podemos estimar  $S(t)$  a partir das observações  $t_1, t_2, \dots, t_n$ , é relevante perguntar se também poderemos obter estimativas únicas de  $F(w)$  e de  $S(t|w)$  usando (2.3). A resposta a esta questão, no caso geral, é negativa. Na verdade, se não dispomos de informação prévia sobre a distribuição de  $W$  e sendo  $S(t)$  a função de sobrevivência observada, por (2.3), a verdadeira função  $\Lambda(t)$  pode ser qualquer função da forma  $\Lambda(t) = L^{-1}[S(t)]$ , onde  $L(\cdot)$  é qualquer função completamente monótona, com  $L(0) = 1$ .

A definição de fragilidade dada por Vaupel *et al.* (1979) pressupõe que cada indivíduo nasce com uma determinada fragilidade, a qual se mantém constante ao longo da sua vida. Adoptando esta definição e admitindo que  $W$  é uma variável aleatória contínua que segue uma determinada distribuição com f.d.p.  $f(t)$ , Hougaard (1984) obteve a distribuição de fragilidade entre os sobreviventes no instante  $t$ , ou seja, a distribuição condicional de  $W$  dado  $T \geq t$

$$g(w|T \geq t) = \frac{\int_t^\infty f(u|w)f(w)du}{P(T \geq t)} = \frac{\exp[-w\Lambda(t)]f(w)}{L[\Lambda(t)]}, \quad (2.5)$$

onde  $f(t|w)$  representa a função densidade de  $T$  condicional à fragilidade  $W$ . De facto, para  $t = 0$ ,  $g(w|T \geq 0) = f(w)$ , ou seja, a distribuição da fragilidade à nascença coincide com a distribuição da fragilidade.

A distribuição da fragilidade entre os indivíduos que morrem no instante  $t$ , que pode ser expressa pela função densidade de  $W$  dado  $T = t$ , foi obtida pelo mesmo autor e tem a expressão

$$g(w|t) = \frac{w\lambda(t)\exp[-w\Lambda(t)]f(w)}{-L'[\Lambda(t)]\lambda(t)} = \frac{w\exp[-w\Lambda(t)]f(w)}{-L'[\Lambda(t)]}, \quad (2.6)$$

a qual, por confronto com (2.5), evidencia a diferença existente entre a distribuição da fragilidade entre os mortos e entre os sobreviventes num determinado instante *t*isto é, temos

$$g(w|t) = g(w|T \geq t) \frac{\mu(t|w)}{h(t)}.$$

A fragilidade média entre os sobreviventes no instante  $t$ , ou seja, o valor médio de  $W$  dado  $T \geq t$ , é dado por

$$E(W|T \geq t) = \frac{\int_0^\infty w \exp[-w\Lambda(t)]f(w)dw}{L[\Lambda(t)]} = -\frac{L'[\Lambda(t)]}{L[\Lambda(t)]}, \quad (2.7)$$

donde (2.4) pode ser escrito na forma

$$h(t) = E(W|T \geq t)\lambda(t). \quad (2.8)$$

Assim, a função hazard populacional  $h(t)$  não é mais do que o valor esperado, entre os sobreviventes no instante  $t$ , da verdadeira função hazard  $\mu(t|w)$ .

Várias distribuições têm sido propostas para a fragilidade, sendo a facilidade de tratamento matemático, em particular a facilidade na obtenção da transformada de Laplace da distribuição, um factor importante de decisão. Por este motivo, a distribuição gama é a mais usada, seguindo-se a distribuição Gaussiana inversa, as distribuições estáveis positivas e a distribuição de Poisson composta, entre as mais usuais.

Numa perspectiva de inferência, para uma amostra de dimensão  $n$ , a função de verosimilhança conjunta para  $T$  e  $W$  é

$$L = \prod_{i=1}^n f(t_i, w_i)^{\delta_i} S(t_i, w_i)^{1-\delta_i},$$

onde  $t_1, \dots, t_n$  representam os tempos de vida,  $\delta_1, \dots, \delta_n$  são tais que  $\delta_i = 1$  se  $t_i$  corresponde a um tempo de vida observado e  $\delta_i = 0$  se  $t_i$  corresponde a um tempo de vida censurado,  $i = 1, \dots, n$ . Para efeitos de estimação dos parâmetros da distribuição da fragilidade e da distribuição do tempo de vida através do algoritmo EM, utiliza-se a função de verosimilhança completa (assim designada pelo facto de encararmos os valores da fragilidade como observados), deduzida a partir da função de verosimilhança anterior, e que se escreve na forma

$$L_C = \prod_{i=1}^n [\mu(t_i|w_i)S(t_i|w_i)f(w_i)]^{\delta_i} [S(t_i|w_i)f(w_i)]^{1-\delta_i},$$

ou ainda,

$$L_C = \prod_{i=1}^n f(w_i) \prod_{i=1}^n \mu(t_i|w_i)^{\delta_i} S(t_i|w_i). \quad (2.9)$$

Para obtermos um modelo de fragilidade com covariáveis observadas, podemos generalizar o modelo de Cox (1972), bastando para tal substituir  $\lambda(t)$  por  $\lambda_0(t) \exp(\mathbf{z}\boldsymbol{\beta})$  em (2.1), onde  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$  designa o vector de coeficientes de regressão e  $\mathbf{z} = (z_1, z_2, \dots, z_k)$  o vector de covariáveis observadas associado ao indivíduo.

### 2.2.1 Um caso particular

Com o objectivo de ilustrar a aplicação do algoritmo EM na estimação dos parâmetros do modelo, vamos abordar um caso particular do modelo multiplicativo com fragilidade, obtido ao admitirmos a distribuição exponencial de parâmetro  $\lambda$  para o tempo de vida e a distribuição Gaussiana inversa de parâmetros  $\mu$  e  $\theta$  para a fragilidade.

#### Distribuição Gaussiana inversa como distribuição da fragilidade

De acordo com Hougaard (1984), consideremos a família de distribuições cujos elementos têm f.d.p. dada por

$$f(w) = \frac{w^\delta \exp(-\theta w)m(w)}{\phi(\delta, \theta)}, \quad w > 0, \quad (2.10)$$

onde  $\theta$  toma valores em  $(a, \infty)$  e o suporte é independente de quaisquer parâmetros desconhecidos. A distribuição correspondente a (6.29) é representada por  $D(\delta, \theta)$  e pertence à família exponencial, com suporte em  $[0, \infty)$ . Note-se que, pelo facto de  $f(w)$  ser uma f.d.p., temos

$$\phi(\delta, \theta) = \int_0^\infty w^\delta \exp(-\theta w)m(w)dw,$$

donde, para  $\delta$  fixo,

$$\phi(\delta + 1, \theta) = -\frac{d\phi(\delta, \theta)}{d\theta}.$$

Novamente por Hougaard (1984), a transformada de Laplace da distribuição  $D(\delta, \theta)$  é

$$L(s) = \frac{1}{\phi(\delta, \theta)} \int_0^\infty w^\delta \exp[-(\theta + s)w] m(w) dw = \frac{\phi(\delta, \theta + s)}{\phi(\delta, \theta)} \quad (2.11)$$

e a correspondente derivada é

$$L'(s) = -\frac{1}{\phi(\delta, \theta)} \int_0^\infty w^{\delta+1} \exp[-(\theta + s)w] m(w) dw = -\frac{\phi(\delta + 1, \theta + s)}{\phi(\delta, \theta)}. \quad (2.12)$$

Como a distribuição Gaussiana inversa é um elemento da família de distribuições acima considerada, as expressões (6.30) e (2.12) tornarão mais simples o cálculo posterior da sua transformada de Laplace e respectiva derivada.

Seja  $W$  uma v.a. com distribuição Gaussiana inversa de parâmetros  $\mu$  e  $\theta$ , isto é,  $W \sim GI(\mu, \theta)$ . Então, a correspondente f.d.p. é dada por

$$f(w) = (\mu/\pi)^{1/2} \exp[(4\mu\theta)^{1/2}] w^{-3/2} \exp(-\theta w - \mu/w), \quad w > 0,$$

em que  $\theta \geq 0$  e  $\mu > 0$ . Se considerarmos  $m(w) = (\mu/\pi)^{1/2} \exp(-\mu/w) w^{-1}$ ,  $\delta = -1/2$  e

$$\phi(-1/2, \theta) = \exp[-(4\mu\theta)^{1/2}],$$

constatamos que se trata da distribuição  $D(-1/2, \theta)$  e verificamos facilmente a estrutura de família exponencial desta distribuição.

A distribuição Gaussiana inversa é um caso particular da distribuição Gaussiana inversa generalizada, designada por  $GIG(\lambda, \mu, \theta)$ ,  $\lambda \in R$ , estudada por Jørgensen (1982). A sua f.d.p. é dada por

$$\frac{(\theta/\mu)^{\lambda/2}}{2K_\lambda(2(\theta\mu)^{1/2})} w^{\lambda-1} \exp(-\theta w - \mu/w), \quad w > 0,$$

onde  $K_\lambda$  designa a função de Bessel modificada de terceira espécie com índice  $\lambda$ , que é representada por

$$K_\lambda(u) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp[-u(y + y^{-1})/2] dy, \quad u > 0.$$

De facto,  $W \sim GI(\mu, \theta)$  quando  $W \sim GIG(-1/2, \mu, \theta)$ . Se  $\lambda = 1/2$ , obtém-se outro caso particular importante da distribuição Gaussiana inversa generalizada, a distribuição Gaussiana inversa recíproca.

Seja  $Y_1 \sim GIG(-1/2, \mu, \theta)$ . Então, de acordo com Jørgensen (1982),  $Y_1^{-1} \sim GIG(1/2, \theta, \mu)$ , logo a v.a. com distribuição Gaussiana inversa recíproca  $Y_2 \sim GIG(1/2, \mu, \theta)$  designa-se por recíproca complementar de  $Y_1$ .

O valor médio da distribuição Gaussiana inversa é  $(\mu/\theta)^{1/2}$  para  $\theta > 0$  e não existe para  $\theta = 0$ ; a variância é  $(1/2)\mu^{1/2}\theta^{-3/2}$  para  $\theta > 0$ .

Por razões de identificabilidade do modelo, admitimos que  $E(W) = 1$ , logo consideramos que  $\mu = \theta$ , para  $\theta > 0$ , ou seja, a distribuição Gaussiana inversa com apenas um parâmetro,  $\theta$ . Então  $\text{var}(W) = (2\theta)^{-1}$  e o coeficiente de variação é  $(2\theta)^{-1/2}$ .

Note-se que usando o resultado  $E(W)=1$  prova-se que  $\frac{\phi(\delta+1, \theta)}{\phi(\delta, \theta)} = 1$ , donde

$$\phi(-1/2, \theta) = \phi(1/2, \theta).$$

Por (6.30),

$$S(t) = L[\Lambda(t)] = \frac{\phi(-1/2, \theta + \Lambda(t))}{\phi(-1/2, \theta)}$$

e, por (6.30) e (2.12),

$$h(t) = -\frac{L'[\Lambda(t)]}{L[\Lambda(t)]} \lambda(t) = \frac{\phi(1/2, \theta + \Lambda(t))}{\phi(-1/2, \theta + \Lambda(t))} \lambda(t).$$

Então, ainda de acordo com Hougaard (1984), dado que estamos a considerar  $\mu = \theta$ , a função de sobrevivência e a função hazard populacionais são dadas, respectivamente, por

$$S(t) = \exp[-2[\theta(\theta + \Lambda(t))]^{1/2} + 2\theta],$$

$$h(t) = [\theta/[\theta + \Lambda(t)]]^{1/2} \lambda(t).$$

Por (2.6) e (2.12), a distribuição da fragilidade entre os mortos, no instante  $t$ , é dada por

$$g(w|t) = \frac{w^{1/2} \exp[-(\theta + \Lambda(t))w] m(w)}{\phi(1/2, \theta + \Lambda(t))},$$

ou seja, substituindo  $m(w)$  e  $\phi(1/2, \theta + \Lambda(t))$  pelas suas expressões

$$g(w|t) = \frac{w^{-1/2} [(\theta + \Lambda(t))/\pi]^{1/2} \exp[-(\theta + \Lambda(t))w - \theta/w]}{\exp[-2[\theta(\theta + \Lambda(t))]^{1/2}]}, \quad (2.13)$$

que é a f.d.p. da distribuição Gaussiana inversa recíproca  $GIG(1/2, \theta, \theta + \Lambda(t))$ .

A distribuição da fragilidade entre os sobreviventes, no instante  $t$ , é  $D(-1/2, \theta + \Lambda(t))$ , isto é,

$$g(w|T \geq t) = (\theta/\pi)^{1/2} \exp[2[\theta(\theta + \Lambda(t))]^{1/2}] w^{-3/2} \exp[-(\theta + \Lambda(t))w - \theta/w], \quad (2.14)$$

que também é Gaussiana inversa, mas com parâmetros  $\theta$  e  $\theta + \Lambda(t)$ . Então,  $W|_{T \geq t} \sim GIG(-1/2, \theta, \theta + \Lambda(t))$ . Logo  $W|_{T=t}$  designa-se por recíproca complementar de  $W|_{T \geq t}$ .

A fragilidade média entre os sobreviventes, no instante  $t$ , é

$$E(W|T \geq t) = \left[ \frac{\theta}{\theta + \Lambda(t)} \right]^{1/2}. \quad (2.15)$$

### Estimação dos parâmetros com recurso ao algoritmo EM

Cox e Oakes (1984, p. 169) referem um estudo do tempo de vida de pacientes com cancro do canal biliar, tratados com uma combinação de medicamentos e terapia por radiações. Os autores ajustam a distribuição gama a estes dados e usam o algoritmo EM para estimar os parâmetros. O nosso objectivo é

outro: ajustar a distribuição exponencial ao tempo de vida e considerar o modelo multiplicativo com fragilidade em que a fragilidade tem distribuição Gaussiana inversa. O ponto comum é a utilização do algoritmo EM para estimar os parâmetros.

Admitamos que a distribuição do tempo de vida é exponencial de parâmetro  $\lambda$ . Então,

$$S(t|w) = \exp(-w\lambda t) \quad \text{e} \quad \mu(t|w) = w\lambda(t) = w\lambda$$

logo, por (2.9), a função de verosimilhança completa é

$$L_C = \prod_{i=1}^n (\theta/\pi)^{1/2} \exp(2\theta) w_i^{-3/2} \exp(-\theta w_i - \theta/w_i) \prod_{i=1}^n (w_i \lambda)^{\delta_i} \exp(-w_i \lambda t_i).$$

O respectivo logaritmo escreve-se na forma

$$\begin{aligned} \log L_C &= \sum_{i=1}^n \left[ \frac{1}{2}(\log \theta - \log \pi) + 2\theta + (\delta_i - \frac{3}{2}) \log w_i - \theta w_i - \frac{\theta}{w_i} \right] + \\ &\quad \sum_{i=1}^n [\delta_i \log \lambda - w_i \lambda t_i] \\ &= \log L_{C1} + \log L_{C2}. \end{aligned}$$

O objectivo agora é maximizar  $\log L_C$  para encontrar os estimadores de  $\theta$  e de  $\lambda$ . No entanto, nenhum dos  $w_i$ 's,  $(\log w_i)$ 's ou  $(1/w_i)$ 's é observado, daí que tenham de ser substituídos pelo seu valor esperado, condicional aos dados, isto é, condicional a se tratar de um tempo de vida observado ou censurado. Todavia, uma vez que cada valor esperado é encarado como uma constante no processo de derivação, essa substituição só deve ser efectuada no fim deste processo ou, o que é equivalente, na expressão dos estimadores. Assim, temos

$$\frac{\partial \log L_{C1}}{\partial \theta} = \sum_{i=1}^n \left( \frac{1}{2\theta} + 2 - w_i - \frac{1}{w_i} \right)$$

e

$$\frac{\partial \log L_{C2}}{\partial \lambda} = \sum_{i=1}^n \left( \frac{\delta_i}{\lambda} - w_i t_i \right).$$

Igualando as derivadas parciais a zero, obtemos a expressão dos estimadores

$$\hat{\theta} = \left[ 2 \left( \frac{1}{n} \sum_{i=1}^n w_i + \frac{1}{n} \sum_{i=1}^n \frac{1}{w_i} - 2 \right) \right]^{-1} \quad (2.16)$$

e

$$\hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n w_i t_i}. \quad (2.17)$$

Para proceder à substituição dos  $w_i$ 's e  $(1/w_i)$ 's pelo respectivos valores esperados condicionais, precisamos de obter  $E(W|T = t)$ ,  $E(\frac{1}{W}|T \geq t)$  e  $E(\frac{1}{W}|T = t)$ , além de  $E(W|T \geq t)$  já calculado em (2.15).

Por Jørgensen (1982), dada uma variável aleatória  $X$  tal que  $X \sim GIG(\lambda, \mu, \theta)$ , então  $X^{-1} \sim GIG(-\lambda, \theta, \mu)$ . No nosso caso,  $W|_{T \geq t} \sim GIG(-1/2, \theta, \theta + \Lambda(t))$  logo  $W^{-1}|_{T \geq t} \sim GIG(1/2, \theta + \Lambda(t), \theta)$  e  $W|_{T=t} \sim GIG(1/2, \theta, \theta + \Lambda(t))$  donde  $W^{-1}|_{T=t} \sim GIG(-1/2, \theta + \Lambda(t), \theta)$ . De acordo com Jørgensen (1982), temos

$$E(W|T = t) = \left[ \frac{\theta}{\theta + \Lambda(t)} \right]^{1/2} \left[ 1 + \frac{1}{[\theta(\theta + \Lambda(t))]^{1/2}} \right] = \frac{1 + [\theta(\theta + \Lambda(t))]^{1/2}}{\theta + \Lambda(t)},$$

$$E\left(\frac{1}{W}|T \geq t\right) = \left[ 1 + \frac{1}{[\theta(\theta + \Lambda(t))]^{1/2}} \right] \left[ \frac{\theta + \Lambda(t)}{\theta} \right]^{1/2} = \frac{1 + [\theta(\theta + \Lambda(t))]^{1/2}}{\theta},$$

$$E\left(\frac{1}{W}|T = t\right) = \left[ \frac{\theta + \Lambda(t)}{\theta} \right]^{1/2}.$$

Sem perda de generalidade suponhamos que os tempos de vida estão indexados de modo a que os primeiros  $m$  correspondem a tempos censurados e os restantes  $n-m$  a tempos observados. Ao substituímos os  $w_i$ 's e os  $(1/w_i)$ 's pelos respectivos valores esperados, condicionais aos dados, em (2.16) e em (2.17), obtemos

$$\hat{\theta} =$$

$$\left[ \frac{2}{n} \left[ \frac{m}{\theta} + \sum_{i=1}^m \frac{\theta^{1/2} + \theta^{-1/2} \lambda t_i}{(\theta + \lambda t_i)^{1/2}} + \sum_{i=m+1}^n \left( \frac{1 + (\theta(\theta + \lambda t_i))^{1/2}}{\theta + \lambda t_i} + \left( \frac{\theta + \lambda t_i}{\theta} \right)^{1/2} \right) \right] - 4 \right]^{-1}$$

e

$$\hat{\lambda} = (n - m) \left[ \sum_{i=1}^m \frac{t_i}{(1 + (1/\theta)\lambda t_i)^{1/2}} + \sum_{i=m+1}^n \frac{1 + (\theta(\theta + \lambda t_i))^{1/2}}{\theta + \lambda t_i} t_i \right]^{-1}.$$

A partir de agora estamos em condições de poder implementar o algoritmo EM. Para determinar a estimativa inicial do parâmetro da distribuição do tempo de vida, considerámos a estimativa para o modelo sem fragilidade, ou seja,  $\hat{\lambda} = \sum_{i=1}^n \delta_i / \sum_{i=1}^n t_i$ , do qual resultou  $\lambda^{(0)} = 0.003388621$ . Em relação à estimativa inicial do parâmetro da distribuição da fragilidade, é prática corrente considerar  $\theta^{(0)} = 1$  e também foi isso que fizemos. No entanto, optámos por considerar ainda outros valores para  $\theta$  de modo a verificar se isso conduzia a estimativas substancialmente diferentes. Tal não aconteceu, como se pode verificar através da tabela 2.1. Os dados e o algoritmo utilizado usando a linguagem R (Ihaka e Gentleman, 1996) encontram-se no anexo B.

## 2.3 Modelo de tempo de vida acelerado com fragilidade

O modelo de tempo de vida acelerado com fragilidade pode ser encarado como um modelo em que a fragilidade actua como um factor de aceleração (ou desaceleração) no tempo. Este modelo foi proposto por Anderson e Louis

Valor inicial de $\theta$	$\hat{\lambda}$	$\hat{\theta}$
0.5	0.00225985	0.16793943
1	0.002259851	0.167939506
2	0.002259851	0.167939536
10	0.002259850	0.167939421

**Tabela 2.1:** Estimativas dos parâmetros.

(1995), na versão univariada e na versão bivariada. Tanto quanto é do nosso conhecimento, não teve até à data qualquer desenvolvimento posterior.

A função hazard no instante  $t$  para um indivíduo com fragilidade  $W = w$ , é dada por

$$\mu(t|w) = w\lambda(wt), \quad (2.18)$$

onde  $W$  é uma variável aleatória não negativa (não observada) e  $\lambda(t)$  é uma função do tempo comum a todos os indivíduos e independente de  $W$ . Note-se que, à semelhança do que acontece com o modelo multiplicativo,  $\lambda(t)$  é a função hazard correspondente aos indivíduos com fragilidade  $w = 1$ . A função de sobrevivência condicional de  $T$  dado  $W = w$  é

$$S(t|w) = \exp[-\Lambda(wt)] = S_0(wt), \quad (2.19)$$

onde  $\Lambda(t) = \int_0^t \lambda(u)du$  é a função hazard cumulativa para um indivíduo com fragilidade  $w = 1$ , que designaremos por função hazard cumulativa subjacente. Então,  $S_0(t) = \exp[-\Lambda(t)]$  é a correspondente função de sobrevivência subjacente, associada a esse mesmo indivíduo padrão. Quanto à função densidade condicional de  $T$  dado  $W = w$  é dada por

$$f(t|w) = w\lambda(wt) \exp[-\Lambda(wt)].$$

Seja  $F(w)$  a função de distribuição da fragilidade  $W$ . A função de sobrevivência e a função hazard populacionais são, respectivamente,

$$S(t) = \int_0^\infty S(t|w) dF(w) = \int_0^\infty \exp[-\Lambda(wt)] dF(w), \quad (2.20)$$

$$h(t) = \frac{\int_0^\infty w\lambda(wt) \exp[-\Lambda(wt)] dF(w)}{\int_0^\infty \exp[-\Lambda(wt)] dF(w)}. \quad (2.21)$$

Analogamente ao que foi apresentado no modelo multiplicativo, vamos agora obter a distribuição da fragilidade entre os sobreviventes no instante  $t$  e a distribuição da fragilidade entre os mortos nesse mesmo instante. Admitamos que  $W$  é uma variável aleatória contínua, com função densidade  $f(w)$ . Assim, a função densidade de  $W|_{T \geq t}$  é

$$g(w|T \geq t) = \frac{\int_t^\infty f(u|w)f(w)du}{P(T \geq t)} = \frac{\exp[-\Lambda(wt)] f(w)}{S(t)},$$

ou ainda,

$$g(w|T \geq t) = \frac{\exp[-\Lambda(wt)] f(w)}{\int_0^\infty \exp[-\Lambda(wt)] f(w)dw}. \quad (2.22)$$

Note-se que, para  $t = 0$ ,  $g(w|T \geq 0) = f(w)$ , ou seja, a distribuição da fragilidade à nascença coincide com a distribuição da fragilidade.

A distribuição da fragilidade entre os indivíduos que morrem no instante  $t$  pode ser expressa pela função densidade de  $W$  dado  $T = t$ , a qual é dada por

$$g(w|t) = \frac{w\lambda(wt) \exp[-\Lambda(wt)] f(w)}{h(t)S(t)} = \frac{w\lambda(wt) \exp[-\Lambda(wt)] f(w)}{\int_0^\infty w\lambda(wt) \exp[-\Lambda(wt)] f(w)dw}. \quad (2.23)$$

À semelhança do que se verifica no modelo multiplicativo com fragilidade, também neste caso é válida a relação

$$g(w|t) = g(w|T \geq t) \frac{\mu(t|w)}{h(t)}.$$

A fragilidade média entre os sobreviventes no instante  $t$ , ou seja, o valor médio de  $W$  dado  $T \geq t$ , é dado por

$$E(W|T \geq t) = \frac{\int_0^\infty w \exp[-\Lambda(wt)] f(w) dw}{\int_0^\infty \exp[-\Lambda(wt)] f(w) dw}, \quad (2.24)$$

Então, tendo em conta (2.21), conclui-se que  $h(t) = E(W\lambda(Wt)|T \geq t)$ , o que está de acordo com o que Anderson e Louis (1995) observaram. Assim, também neste caso continua a ser válida a interpretação de que a função hazard populacional seja o valor esperado, entre os sobreviventes no instante  $t$ , da verdadeira função hazard  $\mu(t|w)$ .

Um pormenor evidente que diferencia este modelo do modelo multiplicativo é o facto de, em geral, não ser possível escrever a função de sobrevivência populacional à custa da transformada de Laplace da distribuição da fragilidade. Na verdade, em muitos casos o cálculo da função de sobrevivência só é possível com recurso à integração numérica. Talvez seja este um dos principais motivos por este modelo ter tido tão pouco desenvolvimento.

No entanto, quando a distribuição para o tempo de vida é a distribuição exponencial, esse problema não ocorre. Basta observar que

$$\Lambda(wt) = \lambda(wt) = w\Lambda(t),$$

donde

$$S(t) = \int_0^\infty \exp[-\Lambda(wt)] f(w) dw = \int_0^\infty \exp[-w\Lambda(t)] f(w) dw = L_W[\Lambda(t)].$$

Mais geralmente, quando temos a distribuição de Weibull,

$$\Lambda(wt) = \lambda(wt)^\gamma = w^\gamma \Lambda(t).$$

Fazendo a mudança de variável  $x = w^\gamma$  e notando que a função densidade de  $X$  é  $g(x) = f(x^{\frac{1}{\gamma}}) \frac{1}{\gamma} x^{\frac{1}{\gamma}-1}$ , temos

$$S(t) = \int_0^\infty \exp[-\Lambda(wt)] f(w) dw = \int_0^\infty \exp[-x\Lambda(t)] g(x) dx = L_X[\Lambda(t)],$$

onde  $L_X[\Lambda(t)]$  é a transformada de Laplace da v.a.  $X = W^\gamma$ .

## 2.4 Modelo aditivo com fragilidade

Este modelo, proposto por Rocha (1995), deve o seu nome ao facto de a fragilidade actuar de modo aditivo na função hazard individual. Assim, seja  $T$  uma v.a. não negativa e absolutamente contínua, que representa o tempo de vida dos indivíduos. A função hazard, no instante  $t$ , para um indivíduo com fragilidade  $W = w$  é dada por

$$\mu(t|w) = w + \lambda(t), \quad (2.25)$$

onde  $W$  é uma v.a. não negativa e  $\lambda(t)$  é independente de  $W$  e comum a todos os indivíduos. À semelhança do que acontece no modelo multiplicativo,  $\lambda(t)$  também se designa por função hazard subjacente, mas corresponde aos indivíduos com fragilidade zero (e não 1).

A função de sobrevivência condicional de  $T$  dado  $W = w$  é

$$S(t|w) = \exp(-wt)S_0(t), \quad (2.26)$$

onde  $S_0(t)$  é a função de sobrevivência subjacente.

As funções anteriores, (2.25) e (2.26), correspondem a funções individuais, logo não observáveis. Vejamos então quais as correspondentes funções populacionais, ou seja, observáveis. A função de sobrevivência é

$$S(t) = L(t)S_0(t) \quad (2.27)$$

e a função hazard é

$$h(t) = \lambda(t) - L'(t)/L(t), \quad (2.28)$$

onde  $L(t)$  representa a transformada de Laplace da distribuição de  $W$ .

Uma vez que o valor médio de  $W$  dado  $T \geq t$  é dado por

$$E(W|T \geq t) = -\frac{L'(t)}{L(t)},$$

a função hazard populacional também pode ser escrita na forma

$$h(t) = \lambda(t) + E(W|T \geq t).$$

Uma característica muito interessante deste modelo é a possibilidade de ser caracterizado como modelo de riscos competitivos e, dada a relação entre estes últimos e os modelos de cura, o modelo aditivo também pode ser interpretado à luz dos modelos de cura. Retomaremos este tema mais adiante.

### 2.4.1 A distribuição $\chi_0'^2(\gamma)$

Quando a heterogeneidade não observada é devida à existência de indivíduos com susceptibilidade nula a determinado acontecimento, a distribuição  $\chi_0'^2(\gamma)$ , designada por Qui-quadrado não central com zero graus de liberdade e com parâmetro de não centralidade  $\gamma$ , pode ser usada como distribuição para a fragilidade, quer se trate do modelo multiplicativo ou do modelo aditivo, como foi proposto por Rocha (1995). Deste modo, a variável  $W$  terá massa de probabilidade não nula em zero (caracterizando os indivíduos não susceptíveis) e distribuição contínua na semi-recta positiva. Assim sendo, o uso desta distribuição nos modelos com fragilidade é uma possibilidade de inserir este tipo de modelos nos modelos com indivíduos imunes, tema principal do nosso trabalho. Vejamos algumas das suas características e como pode ser aplicada no modelo aditivo.

Seja  $W$  a variável fragilidade com distribuição  $\chi_0'^2(\gamma)$ , ou seja, com distribuição dada por

$$G(w) = \sum_{j=0}^{\infty} \left[ \frac{(\gamma/2)^j}{j!} e^{-\gamma/2} \right] F_{2j}(w),$$

onde  $F_k(x)$  representa a função de distribuição de uma variável aleatória com distribuição Qui-quadrado central com  $k$  graus de liberdade. Uma representação alternativa é

$$G(w) = \sum_{j=0}^{\infty} \left[ \frac{(\gamma/2)^j}{j!} e^{-\gamma/2} \right] F_2^{j*}(w),$$

onde  $F_2^{j*}$  representa a  $j$ -ésima convolução de  $F_2$ , logo é a distribuição da soma aleatória

$$S = \begin{cases} X_1 + \dots + X_N & \text{se } N > 0 \\ 0 & \text{se } N = 0 \end{cases}$$

em que  $N \sim Po(\gamma/2)$  e  $X_1, X_2, \dots$  são variáveis aleatórias independentes com distribuição  $\chi_2^2$ .

Por outro lado,

$$G(\infty) = \int_0^\infty dG(w) = P(W = 0) + \int_0^\infty f_\gamma(w)dw,$$

onde

$$f_\gamma(w) = \frac{1}{2}(\gamma/w)^{1/2} \exp[-(\gamma + w)/2] I_1[(\gamma w)^{1/2}]$$

e

$$I_1(w) = \sum_{k=0}^{\infty} \frac{(w/2)^{2k+1}}{k!(k+1)!}$$

é a função de Bessel modificada de primeira espécie e índice 1. Como  $P(W = 0) = \exp(-\gamma/2) > 0$ ,  $W$  é uma variável aleatória mista que não possui densidade no sentido usual.

A transformada de Laplace da distribuição  $\chi_0'^2(\gamma)$  pode ser expressa por

$$L(s) = \exp \left[ -\frac{\gamma}{2} [1 - (1 + 2s)^{-1}] \right].$$

Esta representação tem interesse, entre outros motivos, pelo facto de  $1 - (1 + 2s)^{-1}$  poder ser interpretada como uma função de distribuição, logo  $(1 + 2s)^{-1}$  como uma função de sobrevivência, como veremos adiante.

Por (2.27), a função de sobrevivência populacional é

$$S(t) = \exp \left[ -\frac{\gamma}{2} [1 - (1 + 2t)^{-1}] \right] S_0(t) \quad (2.29)$$

e por (2.28) a função hazard escreve-se na forma

$$h(t) = \lambda(t) + \frac{\gamma}{(1 + 2t)^2}.$$



# Capítulo 3

## Análise de Sobrevivência com indivíduos imunes: estado da arte

### 3.1 Introdução

Tradicionalmente, a Análise de Sobrevivência ocupa-se do tempo até à ocorrência de um determinado acontecimento. Mas há muito que esta área da estatística deixou de se restringir ao estudo do tempo de vida dos indivíduos para entrar num campo mais vasto em que se pretende saber se alguns dos indivíduos em estudo podem ser considerados curados ou imunes a determinado acontecimento. A existência destes indivíduos numa população, os quais teoricamente têm um tempo de vida infinito, faz com que a função de sobrevivência populacional seja imprópria. Os dois tipos de modelos desenvolvidos até à data, que permitem incorporar uma proporção de indivíduos imunes, são os seguintes:

1. modelos de mistura (mistura finita de distribuições) em que à subpopulação dos indivíduos imunes corresponde uma distribuição do tempo de vida degenerada e à subpopulação dos indivíduos não imunes (ou susceptíveis) corresponde uma distribuição própria (paramétrica ou não);
2. modelos de não-mistura, baseados numa única distribuição a qual é modificada de forma a que a função de sobrevivência seja imprópria

(ou, o que é equivalente, de forma a que a função hazard cumulativa seja limitada).

Os modelos que incorporam uma proporção de indivíduos imunes começaram por ser estudados na sua forma mais simples, sem a inclusão de covariáveis. O grande impulso surge com Boag (1949), que já utiliza o método da máxima verosimilhança para estimar os parâmetros do modelo de mistura.

Mas é a partir de 1990 que os modelos de cura têm tido um grande desenvolvimento, como é fácil inferir pela proliferação de artigos sobre o assunto. O tempo de vida dos indivíduos susceptíveis, na abordagem paramétrica, tem sido modelado por várias distribuições, sendo as mais frequentes, a distribuição log-normal (Boag (1949), Gamel *et al.* (1990), Peng *et al.* (2001)), a distribuição exponencial (Farewell (1977), Goldman (1984), Maller e Zhou (1992), Cantor e Shuster (1992), Sposto *et al.* (1992)), a distribuição de Weibull (Farewell (1982), Peng *et al.* (2001)), a distribuição de Gompertz (Gordon (1990)) e a distribuição gama (Peng *et al.* (2001)).

Tendo por base modelos de regressão de hazards proporcionais ou modelos de tempo de vida acelerado, foram propostos recentemente modelos semi-paramétricos (generalização do modelo de regressão de Cox (Kuk e Chen (1992), Sy e Taylor (2000), Peng e Dear (2000)), bem como modelos paramétricos: exponencial (Larson e Dinse (1985), De Angelis *et al.* (1999)), Weibull (De Angelis *et al.* (1999)), gama generalizada (Yamaguchi (1992)) e F generalizada (Peng *et al.* (1998)).

A área da saúde é a área de eleição de aplicação desta teoria, em especial, a oncologia. Nos cancros de evolução rápida, o mais importante é determinar a taxa de cura, ou seja, a proporção de indivíduos curados, pois o objectivo em termos médicos é o aumento desta taxa. Nos cancros de evolução lenta, acaba por ser mais importante conhecer o tempo médio (ou mediano) de vida. Neste último caso, deixa de ser tão relevante o indivíduo atingir a cura pois, em termos teóricos, acaba por viver o mesmo tempo do que se estivesse curado ou mesmo do que se não tivesse tido a doença.

Neste capítulo, procurámos fazer uma compilação do que tem sido feito nesta área, embora não de forma exaustiva, de modo a transmitir o mérito e a importância deste novo ramo da Análise de Sobrevivência. O critério de escolha foi essencialmente a acessibilidade dos artigos.

## 3.2 Descrição histórica: até 1990

A primeira referência encontrada data de 1926, onde Greenwood tenta definir o conceito de curado para os doentes oncológicos como sendo aqueles que, tendo em conta a sua esperança de vida, podiam ser incluídos na população em geral, na respectiva categoria quanto ao sexo, idade e ocupação. Assim, já é admitido que alguns dos doentes com cancro (com várias localizações) possam vir a estar curados, sendo essa classificação baseada essencialmente na esperança de vida após o tratamento. Note-se que não se trata aqui de cura propriamente dita, pois raramente os doentes conseguem ter a mesma esperança de vida que os indivíduos ditos saudáveis. Na verdade, a cura é encarada mais em termos de vantagem, no sentido em que os doentes conseguem ter uma esperança de vida maior se forem sujeitos ao tratamento. Aliás, a dificuldade em definir de modo rigoroso o termo cura é referida neste artigo, em especial para doenças crónicas ou de desenvolvimento lento. Um outro aspecto focado por Greenwood (1926) é a *duração natural do cancro*, que o próprio define como sendo o tempo decorrido desde o instante em que o doente se recorda de ter tido os primeiros sintomas até à sua morte pelo cancro. Como o próprio autor reconhece, esta é uma definição muito subjectiva e pouco precisa, não só porque o instante inicial é determinado pelo próprio doente como por haver uma grande variedade e intensidade de sintomas nos vários tipos de cancro.

Vários anos mais tarde, facto a que não é concerteza alheia a ocorrência da 2ª Guerra Mundial, surge Boag (1949) com um artigo inovador pois, para além de considerar a possibilidade de alguns indivíduos sujeitos a tratamento para o cancro poderem vir a estar curados, apresenta um novo modo de estimar a proporção de indivíduos curados. De facto, considera um modelo de mistura com distribuição log-normal de parâmetros  $\mu$  e  $\sigma$  para o tempo de vida dos indivíduos susceptíveis, e estima a proporção de indivíduos curados,  $c$ , conjuntamente com os dois parâmetros da distribuição, pelo método da máxima verosimilhança. Até então as estimativas dos parâmetros da distribuição baseavam-se no método actuarial, apesar de neste caso o cálculo do desvio padrão também poder ser algo complicado para a época. De realçar a genialidade deste artigo, tendo em conta a ausência de auxiliares de cálculo eficazes, além de que o método da máxima verosimilhança era dominado por poucos como é admitido pelo próprio Greenwood nos comentários que faz ao artigo. Além disso, Boag (1949) também define com precisão o tempo de

vida (tem início no momento em que é administrado o tratamento e não no momento em que o doente se lembra de ter tido os primeiros sintomas como definiu Greenwood (1926), e *terminus* no instante de morte) e o tempo de *follow-up* (definido apenas para os doentes que permanecem vivos até ao fim do estudo).

Por norma, à data do artigo, a estimativa da esperança de vida aos 5 anos é o valor de referência para avaliar a eficácia de um tratamento. Boag (1949) afirma que a eficácia de um tratamento na completa erradicação da doença é mais bem avaliada através da taxa de cura do que através da esperança de vida aos 5 anos e refere algumas razões para isso. Este valor da esperança de vida é muitas vezes interpretado como uma eventual estimativa da taxa de cura mas não tem em conta que, para alguns tipos de cancro, é usual ocorrer recaída mesmo após 5 anos. Além disso, dada a idade avançada da maioria dos doentes com cancro (novamente à data do artigo), ainda que estejam curados a sua esperança de vida é pequena, donde podem morrer de outra causa que não o cancro em estudo. Outra objecção para o uso desta estatística tem a ver com o tempo que é necessário esperar (5 anos) até se poder obter algumas conclusões. O exemplo que este autor estuda, apresenta uma estimativa da esperança de vida aos 5 anos de 43.4%, enquanto que a estimativa da taxa de cura,  $c$ , é de apenas 25.7%, o que provavelmente reflecte a existência de recaídas após 5 anos.

Em seguida, encontramos Berkson e Gage (1952) que, com base nas taxas de sobrevivência de doentes com cancro calculadas pelo método actuarial, comparam a mortalidade de dois grupos diferentes em relação ao tratamento, ao tipo de cancro, ou a outra característica. A comparação é feita tendo em conta dois parâmetros: o parâmetro  $c$ , que representa a proporção de indivíduos curados na população, isto é, com taxa de morte igual à da população dita saudável, e o parâmetro  $\beta$ , que designa a taxa instantânea de morte devida ao cancro da restante população, isto é, dos indivíduos não curados. O modelo que propõem é definido por

$$l_T = cl_0 + (1 - c)l_0e^{-\beta},$$

onde  $l_T$  representa a probabilidade de sobreviver até ao instante  $t$  para a população em geral (curados e não curados) e  $l_0$  representa a probabilidade de sobrevivência até ao instante  $t$  para os indivíduos curados, obtida a partir das tabelas de mortalidade padrão. Os autores apresentam dois métodos

de estimação dos parâmetros: um método gráfico e o método dos mínimos quadrados.

Estes autores corroboram a opinião de Boag (1949) no sentido em que também afirmam que o parâmetro  $c$  traduz melhor a eficácia de um tratamento do que a taxa de sobrevivência aos 5 anos.

Novamente se passa algum tempo até que o assunto seja retomado com Haybittle (1965), que também estuda o tempo de vida de doentes com cancro. Para este autor, o tempo de vida tem início no momento do primeiro tratamento. O modelo proposto para representar a função de sobrevivência também inclui dois parâmetros,  $c$  e  $\beta$  (com o mesmo significado que em Berkson e Gage, 1952), e é da forma

$$\frac{P_T}{\tilde{P}_T} = ce^{(-\ln c)e^{-\beta T}},$$

com  $P_T$  a representar a proporção de doentes que sobrevivem até ao instante  $T$  e  $\tilde{P}_T$  a proporção de indivíduos ditos normais que sobrevivem até ao instante  $T$ . O método de estimação utilizado já é o método da máxima verosimilhança. O modelo resulta de observar que a probabilidade condicional de morrer de cancro num determinado intervalo, calculada pelo método actuarial, tende a decrescer exponencialmente com o tempo. Assim, o modelo foi designado por *modelo actuarial extrapolado*, para distingui-lo do modelo log-normal de Boag (1949) e do modelo exponencial de Berkson e Gage (1952). O conceito de curado continua a ser próximo do definido por Greenwood (1926), pois um indivíduo curado é aquele que tem uma esperança de vida dita normal quando comparada com os indivíduos da mesma idade e do mesmo sexo. Uma observação relevante neste artigo e que se mantém perfeitamente actual, é o cuidado a ter no uso da estimativa da proporção de curados para inferências futuras, dada a constante evolução da medicina.

Alguns anos mais tarde, Farewell (1977) propõe um modelo de mistura, no qual a probabilidade de ocorrência da doença (muitas vezes designada por incidência) é estimada através de um modelo de regressão logística, e o tempo até à ocorrência da doença é modelado pela distribuição exponencial.

Note-se que esta situação é diferente das anteriores no sentido em que agora temos indivíduos saudáveis e se estuda o tempo decorrido até adoecerem, enquanto que anteriormente os indivíduos já estavam doentes e se

estudava o tempo até à morte pela doença. Embora muitas vezes se use indiscriminadamente os termos imune e curado, a verdade é que neste caso faz mais sentido designar os indivíduos por imunes e anteriormente por curados.

Assim, Farewell (1977) começa por considerar uma variável binária associada ao  $i$ -ésimo indivíduo, que toma o valor um se a doença ocorre (o indivíduo é susceptível) e zero caso contrário (o indivíduo é imune ou não susceptível). A variável  $Y_i$  é caracterizada por

$$P[Y_i = 1 | \mathbf{x}_i] = b_i(\boldsymbol{\beta}, \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}' \mathbf{x}_i)},$$

onde  $\mathbf{x}_i$  é o vector de covariáveis associado ao  $i$ -ésimo indivíduo e  $\boldsymbol{\beta}$  o vector de coeficientes de regressão. Designando por  $T$  o tempo até à ocorrência da doença, a função densidade de  $T$  para os indivíduos susceptíveis é dada por

$$f(t|Y = 1) = \lambda e^{-\lambda t}, \quad t > 0.$$

A estimação dos parâmetros também é realizada através do método da máxima verosimilhança. Se para o  $i$ -ésimo indivíduo a doença for observada no instante  $t_i$ , então a sua contribuição para a função de verosimilhança é

$$b_i(\boldsymbol{\beta}, \mathbf{x}_i) f(t_i | Y_i = 1).$$

Se o  $i$ -ésimo indivíduo for seguido até ao instante  $t_i$  sem que a doença se tenha manifestado, então a sua contribuição para a função de verosimilhança é

$$1 - b_i(\boldsymbol{\beta}, \mathbf{x}_i) + b_i(\boldsymbol{\beta}, \mathbf{x}_i) \int_{t_i}^{\infty} f(t|Y = 1) dt,$$

pois nesta situação é preciso ter em conta que ou a doença não ocorrerá, ou ocorrerá após o instante  $t_i$ .

Ainda Farewell (1982), na sequência do estudo efectuado por Pierce *et al.* (1979), analisa dados provenientes de um estudo de toxicologia, onde uma população de peixes é dividida em três grupos e cada um é exposto a

um certo nível de concentração de uma substância tóxica (zinco). Como se observou que, para determinadas concentrações, uma proporção substancial de peixes sobrevive até ao final do estudo, surge a motivação para utilizar um modelo de mistura. O modelo é muito semelhante ao já utilizado por Farewell (1977), diferindo essencialmente na distribuição utilizada para o tempo de vida dos indivíduos susceptíveis; é usada a distribuição de Weibull em vez da distribuição exponencial:

$$f(t|Y = 1, \mathbf{x}) = \delta\lambda(\lambda t)^{\delta-1} \exp [-(\lambda t)^\delta], \quad t > 0.$$

Neste caso,  $\mathbf{x} = (x_0 = 1, x_1, \dots, x_k)$  é o vector de covariáveis,  $\lambda = \exp(-\boldsymbol{\gamma}'\mathbf{x})$  e  $\boldsymbol{\gamma}$  é o vector de parâmetros de regressão.

Haybittle (1983), consciente da controvérsia existente em torno da possibilidade de cura dos doentes oncológicos, apresenta três interpretações para o conceito de cura.

1. Para um indivíduo cujo cancro foi completamente erradicado, é de esperar que o seu risco de morte devido a esse cancro não seja maior que o risco de morte da restante população do mesmo sexo e com a mesma idade. Temos assim a *cura clínica*.
2. Quanto à *cura pessoal*, um doente considerar-se-à curado se não voltar a apresentar sintomas do cancro até o fim da sua vida, quer esta seja longa ou curta. Claro que isto não corresponde obrigatoriamente à completa erradicação da doença, mas para o doente já é considerada uma cura satisfatória. Além disso, o doente irá morrer de uma causa que não o cancro, não apresentando quaisquer manifestações da doença na morte.
3. Relativamente à *cura estatística*, pode ser interpretada da seguinte forma: se um grupo de doentes passa a exibir uma taxa de morte anual devida a qualquer causa semelhante à restante população do mesmo sexo e com a mesma idade, então o grupo pode ser considerado curado. Note-se que neste último caso, os doentes podem apresentar uma maior taxa de morte devido ao cancro, mas que será de certa forma compensada pelo facto de então morrerem menos devido a outras causas.

Pouco depois surge Goldman (1984) que admite que, após os pacientes entrarem no estudo, para uma proporção desconhecida  $(1 - \pi)$  dessa população, é observado o acontecimento de interesse com uma taxa de falha  $\lambda$ . Assim, o modelo que esta autora considera, escrito em termos da função de sobrevivência, é

$$S(t) = \pi + (1 - \pi)e^{-\lambda t}, \quad t \geq 0,$$

ou seja, é um modelo de mistura baseado na distribuição exponencial, sem covariáveis.

O logaritmo da função de verosimilhança para uma amostra de  $n$  indivíduos é dada por

$$\ln L(\lambda, \pi) = d[\ln(1 - \pi) + \ln \lambda] - \lambda t_+ + \sum_{i=1}^n (1 - \delta_i) \ln [(1 - \pi)e^{-\lambda c_i} + \pi],$$

onde  $c_i$  é o tempo de censura de um indivíduo para o qual não foi observado o acontecimento de interesse,  $\delta_i$  é a variável indicatriz que toma o valor 1 se para o  $i$ -ésimo indivíduo o acontecimento de interesse ocorreu no instante  $t_i \leq c_i$  e o valor zero se para o  $i$ -ésimo indivíduo o acontecimento de interesse não ocorreu até ao instante  $c_i$ ,  $d = \sum_{i=1}^n \delta_i$  é o número de indivíduos para os

quais o acontecimento de interesse foi observado e  $t_+ = \sum_{i=1}^n \delta_i t_i$  é o tempo total em observação desses indivíduos. Goldman (1984) obtém estimativas de máxima verosimilhança para a taxa de cura e para  $\lambda$  de forma análoga a Farewell (1977). Através do método de Monte Carlo, efectua um estudo sobre várias características dos estimadores de máxima verosimilhança de  $\pi$  e de  $\lambda$ . O estimador para a taxa de cura é centrado e o estimador para o parâmetro da função hazard apresenta algum enviesamento quando a taxa de cura é elevada mas, no contexto do modelo, isso não é grave uma vez que esse parâmetro é um parâmetro perturbador de pouco interesse.

Larson e Dinse (1985) efectuaram um estudo com dados relativos ao tempo de vida após o transplante de coração. Aplicaram um modelo de mistura paramétrico, no âmbito dos riscos competitivos, pois admitiram que os indivíduos susceptíveis podiam estar sujeitos a vários tipos de falha.

Os parâmetros de mistura correspondem às probabilidades marginais de ocorrência dos vários tipos de falha (que se supõe terem distribuição multinomial) e são estimados através de modelos de regressão logística. Assim, denotando por  $\mathbf{z}$  o vector coluna das  $K$  covariáveis, a probabilidade de morte devido à causa  $j$  ( $j = 1, \dots, J$ ) dado o vector de covariáveis  $\mathbf{z}$  é dada pelo modelo de regressão logística

$$P_j(\mathbf{z}) = \Pr(D = j|\mathbf{z}) = \frac{\exp(\mu_j + \boldsymbol{\pi}_j \mathbf{z})}{\sum_{l=1}^J \exp(\mu_l + \boldsymbol{\pi}_l \mathbf{z})},$$

com  $\mu_j$  constante e  $\boldsymbol{\pi}_j$  vector dos coeficientes de regressão ( $j = 1, \dots, J$ ). Por questões de unicidade,  $\mu_J = 0$  e  $\boldsymbol{\pi}_J = 0$ . Quanto à função de sobrevivência condicional à falha devido à causa  $j$ , é dada por

$$S_j(t|\mathbf{z}) = \Pr(T > t|\mathbf{z}, D = j) = \exp \left[ - \int_0^t h_j(x) \exp(\boldsymbol{\beta}_j \mathbf{z}) dx \right],$$

onde  $h_j(x)$  é a função hazard subjacente ( $\mathbf{z} = 0$ ) para a falha do tipo  $j$  e  $\boldsymbol{\beta}_j$  é o vector dos coeficientes de regressão. Por simplicidade, foi adoptado o modelo que se segue para a função hazard

$$h_j(x) = \exp(\alpha_{jm}) \text{ se } x \in I_m,$$

onde  $I_1, I_2, \dots, I_m$  são  $M$  intervalos previamente especificados, que constituem uma partição da semi-recta real não negativa, ou seja, foi adoptada a distribuição exponencial com o parâmetro a variar em cada um dos intervalos referidos. Sob este modelo,  $h_j(x)$  é constante em cada intervalo e  $\alpha_m$  é um parâmetro que representa o logaritmo da função hazard subjacente correspondente aos indivíduos sujeitos à falha do tipo  $j$  no intervalo  $I_m$  ( $j = 1, \dots, J; m = 1, \dots, M$ ). Quanto ao método de estimação dos parâmetros de regressão, é usado o método da máxima verosimilhança (com recurso ao algoritmo EM), tendo em conta que a contribuição para a verosimilhança de uma observação não censurada do tipo  $j$  no instante  $t$  é  $h_j(t) \exp(\boldsymbol{\beta}_j \mathbf{z}) P_j(\mathbf{z}) S_j(t|\mathbf{z})$ , e de uma observação censurada no mesmo instante é  $\sum_{l=1}^J P_l(\mathbf{z}) S_l(t|\mathbf{z})$ .

Embora o modelo descrito não leve em consideração a existência de indivíduos imunes, a metodologia aplicada é muito semelhante à aplicada nos modelos de mistura com indivíduos imunes e, como os próprios autores referem, o modelo pode ser visto como uma generalização do proposto por Farewell (1982). Assim, além de este artigo ter dado um impulso determinante na teoria dos riscos competitivos, a estrutura do modelo considerado e o método de estimação dos parâmetros sem dúvida que constituem uma boa referência para os modelos com indivíduos curados.

Em Farewell (1986), o autor chama a atenção para os perigos que podem advir do uso indiscriminado dos modelos de mistura, salientando que só os aconselha quando houver forte evidência científica da existência de duas subpopulações: susceptíveis e curados (ou imunes). Do ponto de vista estatístico, esta não é uma decisão fácil, até porque se trata de um problema de inferência não padrão pois a hipótese nula encontra-se na fronteira do espaço de parâmetros. Além disso, em termos clínicos, é necessário ter em conta que só faz sentido falar em taxas de cura se os dados forem referentes a um *follow-up* suficientemente longo.

Numa perspectiva de comparar tratamentos através dum possível aumento do valor da mediana do tempo de vida dos indivíduos, Halpern e Brown (1987) compararam a potência do teste *log-rank* e do teste de Wilcoxon generalizado para o modelo de cura de Goldman (1984)

$$1 - F_i(t) = \pi_i + (1 - \pi_i)e^{-\frac{t \ln 2}{\lambda_i}}, \quad i = 0, 1,$$

onde  $F_i$  representa a função de distribuição do tempo de vida correspondente a cada grupo de tratamento,  $\pi_i$  a taxa de cura em cada grupo e  $\lambda_i$  é a mediana da distribuição exponencial para cada grupo. A comparação foi efectuada através de um estudo de simulação, considerando  $\lambda_0 = 1$  e fazendo variar  $\lambda_1, \pi_0$  e  $\pi_1$ , assim como o tempo de *follow-up*. Concluíram que o teste *log-rank* é o mais potente, excepto se as taxas de cura forem elevadas e aproximadamente iguais nos dois grupos.

Se num estudo de análise de sobrevivência houver indícios de existência de uma probabilidade positiva de cura, havendo dois grupos de tratamento, o maior interesse recai na diferença entre as taxas de cura dos dois grupos e não nas diferenças iniciais entre as distribuições dos tempos de vida. A

partir deste pressuposto, Gray e Tsiatis (1989) desenvolveram um teste para detectar diferenças entre as funções de sobrevivência, com realce para as diferenças tardias, mas em que a hipótese nula continua a ser a igualdade das funções de sobrevivência e não apenas a igualdade das taxas de cura. Assim, o teste desenvolvido (pertencente à classe dos testes de ordem lineares) difere do teste *log-rank* pelo facto de atribuir um peso maior às diferenças tardias entre as funções de sobrevivência. Além disso, a hipótese alternativa é a proporcionalidade das funções de distribuição do tempo de vida de cada grupo de tratamento. No entanto, mesmo nas situações em que esta última hipótese não seja válida, se o objectivo for detectar diferenças entre as taxas de cura e não diferenças entre as funções de sobrevivência, o teste desenvolvido pelos autores continua a ser preferível ao teste *log-rank* devido a atribuir maior peso às diferenças tardias. Se a taxa de cura for inferior a 50% é aconselhado o uso do teste proposto; se a taxa de cura for pelo menos 50% então não haverá grande diferença entre o teste proposto e o teste *log-rank*, pelo que pode ser usado qualquer um deles.

### 3.3 Descrição histórica: a partir de 1990

No entanto, é a partir de 1990 que este tema começa a merecer maior atenção por parte dos investigadores, com o aparecimento de vários artigos, talvez um pouco motivados não só pelos progressos clínicos a nível da cura do cancro, como também pelos fundamentos teóricos entretanto estudados e pelo recente aparecimento de novas ferramentas de trabalho, de que se destaca o algoritmo EM.

Gordon (1990) tenta estimar a proporção de curados usando o conceito de *cura pessoal* definido por Haybittle (1983). Para tal, emprega a distribuição de Gompertz num modelo de mistura, pois é uma distribuição que se ajusta bem aos tempos de vida da população em geral, ou seja, aos obtidos a partir dos censos. Assim, num grupo de doentes oncológicos sujeitos a tratamento, admite que uma parte é formada pelos indivíduos que morrem duma causa que não o cancro e cuja função de sobrevivência é  $S_0$ , e a outra é formada pelos indivíduos que morrem devido ao cancro e cuja função de sobrevivência é  $S_d$ . Os tempos de vida de ambas as subpopulações são modelados pela distribuição de Gompertz, condicional à idade do doente ou no momento da cirurgia, ou no momento do tratamento inicial (no exemplo apresentado,

coincidem). Designando por  $t$  o tempo decorrido desde a cirurgia e por  $a$  a idade do doente no momento da cirurgia, Gordon (1990) considerou as funções de sobrevivência nos instantes  $t+a$  e  $a$  dadas, respectivamente, por

$$S_0(t+a) = \exp \left[ -\frac{\exp(\lambda_0 + \gamma_0(t+a)) - \exp(\lambda_0)}{\gamma_0} \right],$$

$$S_0(a) = \exp \left[ -\frac{\exp(\lambda_0 + \gamma_0 a) - \exp(\lambda_0)}{\gamma_0} \right],$$

onde  $\lambda_0$  e  $\gamma_0$  são os parâmetros da distribuição de Gompertz. Então, as funções de sobrevivência e de densidade condicionais à idade do doente no momento da cirurgia são, respectivamente,

$$S_0(t|a) = \exp \left[ -\frac{\exp(\lambda_0 + \gamma_0(t+a)) - \exp(\lambda_0 + \gamma_0 a)}{\gamma_0} \right],$$

$$f_0(t|a) = \exp[\lambda_0 + \gamma_0(t+a)] S_0(t|a).$$

As funções de sobrevivência e de densidade, condicionais à idade do doente no momento da cirurgia, da subpopulação que morre do cancro são, respectivamente:

$$S_d(t|a) = \exp \left[ -\frac{\exp(\lambda_d + \gamma_d(t+a)) - \exp(\lambda_d + \gamma_d a)}{\gamma_d} \right],$$

$$f_d(t|a) = \exp[(\lambda_d + \gamma_d(t+a))] S_d(t|a).$$

O modelo de mistura é então caracterizado por

$$S(t|a) = cS_0(t|a) + (1-c)S_d(t|a),$$

onde  $S(t|a)$  representa a função de sobrevivência para um qualquer doente com idade  $a$  no momento da cirurgia e  $c$  ( $0 \leq c \leq 1$ ) representa a proporção

de curados, ou seja, dos indivíduos que morrem de outras causas. Gordon (1990) supõe que a proporção de indivíduos curados não depende da idade no momento da cirurgia e que, em cada instante  $t$ , a proporção  $c(t)$  de indivíduos curados é estimada por

$$c(t) = \frac{cS_0(t|a^*)}{S(t|a^*)},$$

onde  $a^*$  é a mediana ou a média de idades no momento da cirurgia. As estimativas de  $\lambda_0$  e de  $\gamma_0$  são obtidas através das tabelas de mortalidade da população em geral, com a mesma idade e o mesmo sexo. Quanto às estimativas de máxima verosimilhança de  $c$ ,  $\lambda_d$  e  $\gamma_d$ , são obtidas por um processo iterativo, mais concretamente, pelo método de Newton-Raphson.

Por outro lado, Gamel *et al.* (1990) caracterizam a gravidade de um cancro através de dois parâmetros,  $\pi_c$  (proporção de curados) e  $\mu_t$  (valor médio do logaritmo do tempo de vida dos indivíduos não curados), os quais, no exemplo apresentado, estão inversamente correlacionados com o tamanho do tumor. A distribuição utilizada para o tempo de vida dos indivíduos não curados após o tratamento é a distribuição log-normal pois, para muitos tipos de cancro, parece ser uma distribuição apropriada. Assim, se  $X$  representar o tempo decorrido desde o tratamento até à morte pelo cancro e  $C$  designar o acontecimento “o doente estar curado do cancro em estudo“, a função de distribuição do tempo de vida é

$$F(t) = P [X \leq t|\overline{C}] (1 - \pi_c),$$

onde  $P [X \leq t|\overline{C}]$  é uma função de distribuição correspondente aos indivíduos que não estão curados. Supondo que  $P [X \leq t|\overline{C}]$  tem distribuição log-normal e tendo em conta que  $S(t) = 1 - F(t)$ , a função de sobrevivência dos doentes sujeitos a tratamento (curados e não curados) é dada por

$$S(t) = 1 - (1 - \pi_c) \int_0^t \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[ -\frac{\left(\frac{x-\mu_t}{\sigma_t}\right)^2}{2} \right] dx,$$

onde  $t$  é o logaritmo do tempo de vida e  $\sigma_t$  é o desvio padrão do logaritmo do tempo de vida dos indivíduos não curados. Trata-se novamente de um modelo de mistura.

Para estimar os parâmetros  $\pi_c$ ,  $\mu_t$  e  $\sigma_t$ , Gamel *et al.* (1990) começam por estratificar a amostra de acordo com o tamanho do tumor e depois usam o método da máxima verosimilhança, tal como descrito em Boag (1949), para cada estrato. Em seguida, representam graficamente as estimativas dos parâmetros *versus* o tamanho do tumor, com o objectivo de encontrar funções contínuas que descrevam adequadamente a relação entre  $\pi_c$  e o tamanho do tumor e entre  $\mu_t$  e o tamanho do tumor. Note-se que este processo pode ser adaptado para uma qualquer variável de prognóstico, pelo que este modelo é uma generalização do utilizado por Boag (1949) pois permite a inclusão de covariáveis. Gamel *et al.* (1990) salientam as limitações da distribuição log-normal no exemplo que estudaram (melanoma da úvea), possivelmente devido ao número de potenciais factores de prognóstico para este tipo de tumor que não foram considerados, nomeadamente o grau de diferenciação histológica.

Em 1992 (um dos anos em que encontramos mais artigos), Kuk e Chen estudaram um modelo de mistura que combinava a regressão logística (para a probabilidade de ocorrência de um acontecimento) com o modelo de regressão de hazards proporcionais (para o tempo até à ocorrência do acontecimento). Este modelo pode ser considerado uma generalização semi-paramétrica do modelo proposto por Farewell (1982), pois não é especificada uma distribuição para o tempo de vida dos indivíduos susceptíveis. As estimativas dos parâmetros de regressão foram obtidas maximizando uma aproximação da função de verosimilhança marginal obtida por métodos de Monte Carlo. Foi escolhida esta verosimilhança e não a verosimilhança parcial pois apresentava a vantagem de permitir eliminar o parâmetro perturbador, ou seja, a função de sobrevivência subjacente dos indivíduos susceptíveis. Esta última função foi estimada após terem sido obtidas as estimativas dos parâmetros de regressão. Para tal, foi usada a função de verosimilhança completa e a sua maximização fez-se com recurso ao algoritmo EM. Este modelo semi-paramétrico apresentou-se vantajoso em relação aos modelos paramétricos, em especial, quando eram violadas algumas das condições dos modelos paramétricos.

No mesmo ano, Maller e Zhou (1992) propuseram um estimador não

paramétrico para a proporção de indivíduos susceptíveis na população. Com base numa amostra com observações censuradas, a estimativa é dada pelo valor da estimativa de Kaplan-Meier da função de distribuição empírica calculada na maior observação. Os autores provaram que este era um bom estimador (consistente, assintoticamente normal) desde que fossem verificadas algumas condições, nomeadamente que a proporção de indivíduos imunes não fosse muito próxima de zero e que o *follow-up* fosse considerado suficiente. Para verificar a suficiência do *follow-up*, desenvolveram um teste baseado na magnitude da diferença entre a maior observação e a maior observação não censurada. Em geral, se a maior observação for não censurada, não faz sentido tentar ajustar um modelo de cura pois a probabilidade de existirem indivíduos curados será nula. No entanto, salientaram que este estimador não era necessariamente preferível ao estimador de máxima verosimilhança obtido a partir de um modelo paramétrico, mas que podia ser um bom ponto de partida para técnicas mais sofisticadas.

Também em 1992, Cantor e Shuster fizeram um confronto entre a abordagem paramétrica e a não-paramétrica na estimação da taxa de cura. Começaram por advertir dos perigos de considerar a proporção de curados ( $\pi$ ) igual ao valor da estimativa de Kaplan-Meier na maior observação, pois podia originar resultados contraditórios ao senso comum. Este facto deve-se, em grande parte, a uma característica muitas vezes ignorada e algo anómala do estimador de Kaplan-Meier da função de sobrevivência: quando substituímos uma observação  $t_i$  não censurada por outra  $t'_i > t_i$  também não censurada, onde  $t'_i$  é escolhida por forma a que exista pelo menos uma observação censurada no intervalo  $(t_i, t'_i)$ , a sobrevivência global do grupo melhora e, no entanto, para  $t > t'_i$ ,  $\hat{S}(t)$  diminui, conforme referido em Oakes (1993). Por outro lado, como nos instantes iniciais temos um número elevado de indivíduos em risco comparativamente ao número de mortes em cada instante, os decréscimos da curva de sobrevivência estimada são pequenos, ao contrário do que acontece com as falhas tardias, onde os decréscimos são grandes.

Em seguida, os autores fizeram uma abordagem paramétrica ao problema com o estudo de dois modelos. O primeiro, que é um modelo de mistura, usa a distribuição exponencial para os tempos de vida dos indivíduos cujo acontecimento de interesse foi observado, à semelhança de Goldman (1984):

$$S_1(t) = \pi + (1 - \pi) \exp(-\lambda t), \quad 0 < \pi < 1, \quad \lambda > 0.$$

O segundo, que não é um modelo de mistura, é o modelo de Gompertz modificado,

$$S_2(t) = \exp [\beta^{-1}\alpha(1 - \exp(\beta t))], \quad \alpha > 0, \beta < 0.$$

A razão da modificação apresentada ( $\beta < 0$  em vez de  $\beta > 0$ ) é para que  $\lim_{t \rightarrow \infty} S_2(t) = S_2(\infty) = \exp(\alpha/\beta) > 0$ , ou seja, para que a função de distribuição seja imprópria, como é característico nos modelos para populações com indivíduos curados. Deste modo, a taxa de cura é estimada por  $\exp(\hat{\alpha}/\hat{\beta})$ . Recorde-se que a distribuição de Gompertz já tinha sido usada por Gordon (1990), mas não com a modificação referida e num modelo de mistura.

Os autores verificaram que, para dados que indicavam uma taxa de cura não nula, qualquer um dos dois modelos estudados fornecia curvas de sobrevivência muito similares e taxas de cura que diferiam por menos de 3%. No entanto, preferiram o modelo de Gompertz modificado, não só por ser mais geral, como também por apresentar melhor desempenho em relação à convergência do algoritmo empregue na maximização da função de verosimilhança. De facto, o processo de convergência do método de Newton-Raphson na maximização da função de verosimilhança baseada na distribuição de Gompertz é bastante rápido, ao contrário do que acontece com  $S_1(t)$ , onde frequentemente ou é necessário um grande número de iterações ou o próprio método de Newton-Raphson falha.

Para concluir, consideraram dados referentes a leucemia e, por substituição de uma observação igual a dois anos por outra igual a doze anos, tornou-se evidente a anomalia referida anteriormente, pois o valor da estimativa de Kaplan-Meier na maior observação passa de 45.5% para 39.2%. Já com o modelo de Gompertz modificado esse decréscimo é de apenas 0.6%, pelo que concluem ser preferível usar esta distribuição para estimar a taxa de cura em vez da estimativa de Kaplan-Meier.

Spoto *et al.* (1992) publicaram um artigo onde, através de um estudo de simulação, comparam um teste não paramétrico simples, baseado na estimativa de Kaplan-Meier da função de sobrevivência, e três testes paramétricos (teste da razão de verosimilhanças, teste de Wald e teste *score*) para a diferença das proporções de indivíduos curados em dois grupos. O objectivo foi determinar qual o melhor teste com base na respectiva potência. Além disso,

também lhes interessou saber se algum dos testes anteriores tinha algumas boas propriedades para pequenos valores de  $t$ , ou seja, quando ainda não tinham ocorrido muitas falhas. O principal modelo que consideraram foi o modelo de mistura usual

$$S_i(t) = \pi_i + (1 - \pi_i)H_i(t),$$

onde  $\pi_i$  é a proporção de indivíduos curados no grupo  $i$  e  $H_i(t)$  é a função de sobrevivência dos indivíduos susceptíveis (não curados).  $\pi_i$  foi parametrizado por

$$\pi_i = \frac{\exp(\alpha + \beta i)}{1 + \exp(\alpha + \beta i)}, \quad (3.1)$$

com  $\alpha$  e  $\beta \in R$  (modelo de regressão logística). Para  $H_i(t)$  foi usada a distribuição exponencial de parâmetro  $\lambda_i$ , embora algumas das amostras fossem geradas a partir da distribuição de Weibull. Esta situação é um caso particular do modelo usado por Farewell (1982).

A hipótese nula de interesse é  $H_0 : \pi_0 = \pi_1$ , ou seja,  $H_0 : \beta = 0$ . A função de verosimilhança é dada por

$$L = \prod_{i=0}^1 \prod_{j=1}^{n_i} [(1 - \pi_i)\lambda_i \exp(-\lambda_i t_{ij})]^{\delta_{ij}} [\pi_i + (1 - \pi_i) \exp(-\lambda_i t_{ij})]^{1 - \delta_{ij}}, \quad (3.2)$$

onde  $n_i$  é o número de indivíduos no grupo  $i$ ,  $t_{ij}$  ( $j = 1, \dots, n_i, i = 0, 1$ ) é o tempo até à morte ou até à censura e  $\delta_{ij}$  é o indicador de morte ( $\delta_{ij} = 1$ ) ou de censura ( $\delta_{ij} = 0$ ).

Os três testes paramétricos basearam-se na verosimilhança (3.2). As correspondentes estatísticas de teste, sob  $H_0$ , têm distribuição assintótica Qui-quadrado não central com um grau de liberdade e parâmetro de não centralidade  $\delta^2 = n\beta^2/\sigma_{\beta\beta}$ , onde  $\sigma_{\beta\beta}$  é o elemento apropriado da inversa da matriz de informação esperada.

A estatística de teste do teste não paramétrico é

$$\chi_{PL}^2 = \frac{[\hat{S}_1(u) - \hat{S}_0(v)]^2}{\hat{\sigma}_1^2 + \hat{\sigma}_0^2}, \quad (3.3)$$

onde  $u$  e  $v$  representam o instante imediatamente após o maior instante de morte em cada grupo,  $\hat{S}(t) = \prod_{k:t_k < t} \left(1 - \frac{d_k}{r_k}\right)$  é o estimador de Kaplan-Meier da função de sobrevivência em cada grupo e  $\hat{\sigma}^2$  é a estimativa usual da variância de  $\hat{S}(t)$  dada pela fórmula de Greenwood:

$$\hat{\sigma}^2 = \hat{S}(t)^2 \sum_{k:t_k < t} \frac{d_k}{r_k(r_k - d_k)}.$$

Sob  $H_0$ , a estatística de teste  $\chi_{PL}^2$  tem distribuição Qui-quadrado com um grau de liberdade, condicional a  $u$  e a  $v$ .

Para contemplar a possibilidade de haver um tempo de cura fixo, os autores propuseram ainda o modelo

$$S_i^*(t) = \begin{cases} \exp(-\lambda_i^* t), & t < T^* \\ \exp(-\lambda_i^* T^*) = \pi_i, & t > T^* \end{cases}$$

onde  $T^*$  é o tempo de cura e  $\lambda_i^*$  é a taxa de falha que actua até que seja atingido  $T^*$ .

Spisto *et al.* (1992) concluíram que os três testes paramétricos tinham propriedades semelhantes mesmo para pequenas amostras, desde que a maioria das possíveis falhas tivessem sido observadas, ou seja, desde que o *follow-up* fosse suficientemente longo. No entanto, a preferência recaiu sobre o teste de Wald por ser aquele que apresentou maior potência em situações mais frequentes na prática (por exemplo, quando as funções de sobrevivência são paralelas nos dois grupos) e porque o teste *score* não estava definido em muitas situações, devido à matriz das covariâncias não ser definida positiva, mas com a ressalva de isso ser verdadeiro apenas para o caso em estudo, isto é, usando a parametrização (3.1). Quanto ao teste não paramétrico, teve um desempenho comparável aos testes paramétricos, em especial quando as taxas de falha iniciais nos dois grupos eram as mesmas. Por isso, é referido como uma alternativa razoável, em especial quando as amostras são pequenas e algumas falhas ainda não foram observadas.

Laska e Meisner (1992) abordaram o problema clássico de testar a igualdade de duas funções de sobrevivência, ou seja,  $H_1(t) \equiv H_2(t)$ , na situação em que é definido um tempo de cura. Assim, a variável resposta pode ser de

três tipos: morte pela doença, censura ou cura. Consideraram o modelo de cura

$$H_i(t) = r_i + (1 - r_i)S_i(t), \quad i = 1, 2,$$

donde a igualdade das funções de sobrevivência é verdadeira se e só se  $r_1 = r_2$  e  $S_1(t) \equiv S_2(t)$ , ou seja, se a proporção de curados for igual nos dois grupos e se o mesmo acontecer relativamente às funções de sobrevivência dos susceptíveis. Note-se que se soubermos que  $S_1$  e  $S_2$  são iguais então um teste para  $H_1(t) \equiv H_2(t)$  é um teste para  $r_1 = r_2$ .

Estes autores testaram a igualdade das taxas de cura (para dois e mais grupos), quer as funções de sobrevivência dos susceptíveis fossem iguais ou não. Este último caso foi possível devido ao teste não ser sensível às diferenças nos instantes de morte dos susceptíveis. Se as funções de sobrevivência dos susceptíveis não forem iguais nos dois grupos, os testes de ordem (*rank tests*) podem ser enganadores, pois não testam a igualdade das taxas de cura mas sim a igualdade das funções de sobrevivência. O teste não paramétrico proposto por Sposto *et al.* (1992), cuja estatística de teste é dada por (3.3), é outro candidato a testar a igualdade das taxas de cura quando as funções de sobrevivência dos susceptíveis não são iguais. No entanto, tem como grande limitação o facto de não poder ser generalizado para mais do que dois grupos, ao contrário do que acontece com o teste proposto por Laska e Meisner (1992). Relativamente à estimação das taxas de cura, os autores notaram que a existência de censura pesada ou de *follow-up* insuficiente podia levar a conclusões erradas, nomeadamente, a produzir estimativas das taxas de cura demasiado optimistas.

Ainda em 1992, Yamaguchi apresenta uma abordagem diferente das anteriores, pois baseia-se num modelo de tempo de vida acelerado com covariáveis em vez de num modelo com funções hazard proporcionais. Para estimar a proporção  $p$  de indivíduos curados é usada, como habitualmente, a regressão logística (com vector de covariáveis  $\mathbf{x}$  e vector de parâmetros  $\boldsymbol{\alpha}$ )

$$p(\boldsymbol{\alpha}) = \frac{\exp(\mathbf{x}'\boldsymbol{\alpha})}{1 + \exp(\mathbf{x}'\boldsymbol{\alpha})},$$

mas para modelar o tempo de vida dos indivíduos susceptíveis é utilizada uma extensão da família de distribuições Gama generalizadas. Assim, designa por  $T$  e por  $Y = \log(T)$  as variáveis aleatórias que representam o tempo de vida e o logaritmo do tempo de vida, respectivamente. Yamaguchi considera o modelo de mistura usual

$$S(y) = (1 - p)S_d(y) + p. \quad (3.4)$$

Quanto ao modelo de tempo de vida acelerado, pode ser escrito na forma

$$Y = \mathbf{x}'\boldsymbol{\beta} + \sigma z,$$

onde  $\mathbf{x}$  é o vector de covariáveis,  $\boldsymbol{\beta}$  o respectivo vector de parâmetros de regressão,  $\sigma$  o parâmetro de escala e  $z$  tem distribuição log-gama padrão com parâmetro de forma  $k$ , tal que,

$$f(z; k) = \begin{cases} \frac{k^{k-1/2}}{\Gamma(k)} \exp \left[ \sqrt{kz} - k \exp(z/\sqrt{k}) \right] & \text{se } 0 \leq k < \infty \\ \frac{1}{(2\pi)^{1/2}} \exp(-z^2/2) & \text{se } k = \infty \end{cases} \quad (3.5)$$

Para obter a extensão da família Gama generalizada basta substituir o parâmetro  $k$  pelo parâmetro  $\lambda = k^{-1/2}$  e permitir que  $\lambda$  tanto possa assumir valores positivos como negativos. A função densidade passa a ser

$$f(z; \lambda) = \begin{cases} \frac{|\lambda|}{\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}-2} \exp[\lambda^{-2}(\lambda z - \exp(\lambda z))] & \text{se } \lambda \neq 0 \\ \frac{1}{(2\pi)^{1/2}} \exp(-z^2/2) & \text{se } \lambda = 0 \end{cases}$$

A contribuição da  $i$ -ésima observação de  $Y$ , para o logaritmo da função de verosimilhança no caso do modelo log-gama generalizado modificado devido à inclusão da proporção de curados é

$$\log L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, k) = \delta_i [\log(1 - p(\boldsymbol{\alpha})) + \log f(z_i; k) - \log \sigma] + (1 - \delta_i) \log \left[ (1 - p(\boldsymbol{\alpha})) Q \left( k, k \exp(z_i/\sqrt{k}) \right) + p(\boldsymbol{\alpha}) \right],$$

$0 < k < \infty$ , onde  $z_i = [\log(t_i) - \mathbf{x}'_i \boldsymbol{\beta}] / \sigma$ ,  $\delta_i = 1$  se  $t_i$  corresponde a um tempo de vida observado,  $\delta_i = 0$  se  $t_i$  corresponde a um tempo de vida censurado,  $f(z; k)$  é dado pela expressão (3.5) e  $Q(k, a)$  é a função gama incompleta

$$Q(k, a) = \int_a^{\infty} \frac{x^{k-1}}{\Gamma(k)} \exp(-x) dx.$$

O modelo é aplicado na área da sociologia, mais precisamente ao “emprego permanente” no Japão, um dos poucos exemplos fora da área da medicina. Generalizar o modelo de tempo de vida acelerado através da introdução de uma proporção de curados tem vantagens e desvantagens quando comparado com o correspondente modelo de hazards proporcionais. A principal vantagem reside no facto de permitir separar o efeito das covariáveis no tempo até à ocorrência do acontecimento e na probabilidade de ocorrência do acontecimento. Quanto às desvantagens, existem algumas, tal como o modelo não admitir covariáveis dependentes do tempo.

Maller e Zhou (1994) voltaram a abordar a questão do teste para a suficiência do *follow-up*, desta vez de um modo mais formal do que em Maller e Zhou (1992). Motivados pelas características especiais do exemplo que apresentam (na área da criminologia), incluíram ainda um teste para a possível existência de outliers, mais concretamente, para a possibilidade de a maior observação não censurada ser um outlier. Este teste é baseado na comparação entre as duas maiores observações não censuradas. Estudar a presença de outliers é sobremaneira importante pois dela pode depender a existência ou não de indivíduos curados numa população.

Novamente Maller e Zhou (1995) desenvolveram um teste para detectar a presença de indivíduos imunes, baseado na razão de verosimilhanças. Note-se que, sendo  $p$  a proporção de indivíduos susceptíveis e devido à hipótese a testar,  $H_0 : p = 1$ , se encontrar na fronteira do espaço de parâmetros  $(0, 1]$ , a distribuição assintótica da estatística de teste é não padrão, ou seja, não é uma distribuição Qui-quadrado mas sim uma mistura de uma variável com distribuição Qui-quadrado com um grau de liberdade e uma distribuição degenerada em zero, em que as proporções de mistura são ambas iguais a  $1/2$ .

Com base no exemplo que estudaram, os autores propuseram um teste de ajustamento quando a distribuição dos indivíduos susceptíveis é a distribuição exponencial. Além disso, como existiam dois grupos em estudo, consideraram um teste não paramétrico para a diferença entre a proporção de indivíduos imunes nos dois grupos, que podemos interpretar como uma outra versão de (3.3), pois a estatística de teste é

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

e tem distribuição assintótica Gaussiana padrão.

Paralelamente, Ghitany *et al.* (1995) analisaram dois tipos de estimador para a proporção de indivíduos curados numa população, paramétrico (obtido através do método da máxima verosimilhança) e não paramétrico ( $\hat{p}_n = \hat{F}_n(T_n)$ , sendo  $p$  a proporção de indivíduos susceptíveis e  $T_n$  tal que  $\hat{F}_n(T_n)$  é máximo), com base em estudos de simulação.

Assim, examinaram as vantagens e desvantagens do estimador não paramétrico quando comparado com o estimador paramétrico. Utilizaram a distribuição exponencial para o tempo de vida (com valor médio  $\lambda_0^{-1} = 1$ ) e para o tempo de censura (com valor médio  $\mu^{-1} = 1, 2, 3, 5$ ). Para a proporção de curados, geraram uma amostra proveniente de uma distribuição de Bernoulli de parâmetro  $p_0$ ; a variável associada toma o valor zero se o indivíduo pertence ao grupo dos imunes. Adicionalmente, Ghitany *et al.* fizeram variar a dimensão das amostras ( $n = 50, 100, 200$ ) assim como a proporção de susceptíveis ( $p_0 = 0.1, 0.5, 0.9, 1$ ) com vista a explorar o efeito destes factores nas propriedades assintóticas dos estimadores. A simulação foi efectuada com 1000 réplicas de cada caso para posterior cálculo das médias e dos desvios padrão de  $\hat{p}$  (a estimativa obtido pelo método da máxima verosimilhança) e de  $\bar{p}$  (a estimativa obtido pelo método não paramétrico).

Das simulações efectuadas, os autores concluíram que, quando a proporção de susceptíveis era muito pequena ( $0.1 \leq p_0 \leq 0.3$ ) e  $\lambda_0 = \mu$ , o estimador paramétrico era um mau estimador de  $p_0$  mesmo para grandes amostras ( $n = 200$ ), enquanto que o estimador não paramétrico era muito mais rigoroso e preciso. Além disso, as simulações mostraram que a precisão de ambos os estimadores aumenta à medida que a dimensão da amostra aumenta e também à medida que o valor médio do tempo de censura aumenta em relação ao valor médio do tempo de vida. O estimador não paramétrico de  $p$ , além de ser muito mais simples de calcular, pode ainda ser tomado para valor inicial do processo iterativo para obtenção do estimador paramétrico de  $p$  ou de  $\lambda$ . Podem surgir alguns problemas de convergência que obriguem a um grande número de iterações, mas isto acontece em situações pouco frequentes na prática, como é o caso de amostras relativamente pequenas com uma grande proporção de curados.

Habitualmente, num modelo de cura, pretende-se estimar a taxa de cura, a função de sobrevivência dos indivíduos susceptíveis e o efeito das covariáveis, se tiverem sido incluídas no modelo. Assim, interessa saber se o acontecimento ocorreu (que se designa por incidência) e quando ocorreu, dado que ocorreu (que se designa por latência). Para cada covariável existem dois parâmetros: um que descreve como a covariável afecta a incidência a longo prazo e outro que descreve como afecta a latência. Uma covariável que seja importante para a incidência pode não ser importante para a latência e vice-versa. A importância relativa da incidência e da latência difere consoante as aplicações. No entanto, em geral, a incidência é a parte que costuma ter maior interesse científico, caso em que será apropriado utilizar um pequeno número de covariáveis (ou mesmo nenhuma) na parte da latência, para evitar demasiadas covariáveis no modelo.

Taylor (1995) estudou um modelo semi-paramétrico de mistura caracterizado pela regressão logística para a parte da incidência e por uma abordagem não paramétrica para a parte da latência. Difere do modelo apresentado por Kuk e Chen (1992) por permitir que as covariáveis afectem apenas a parte da incidência, por admitir ser possível saber que, para alguns indivíduos, o acontecimento de interesse nunca ocorre, e pelo método de estimação usado: os coeficientes de regressão e a função hazard foram estimados em conjunto utilizando o algoritmo EM para maximizar a verosimilhança completa. O estimador encontrado apresenta algumas semelhanças com o estimador de Kaplan-Meier, daí que tenha sido designado por estimador logístico/Kaplan-Meier. Todavia, por forma a minimizar a eventualidade da ocorrência de coeficientes de regressão infinitos (em especial com amostras pequenas), foi necessário fazer a restrição de que a função de sobrevivência dos susceptíveis seja zero a partir da última observação não censurada. Saliente-se que esta restrição faz sentido quando estamos a trabalhar com modelos de mistura, uma vez que admitimos haver evidência da existência de dois grupos (susceptíveis e não susceptíveis) e também que o *follow-up* é suficientemente longo de forma a que a maioria dos acontecimentos já tenha ocorrido.

Em 1996, Maller e Zhou publicaram um livro que, além de conter o material dos seus artigos (que não se esgotam nos que aqui mencionamos), naturalmente é mais minucioso e mais vasto. Constitui uma boa referência para um primeiro contacto com esta área da Análise de Sobrevivência e, tanto quanto sabemos, é o único livro existente até à data exclusivamente sobre este assunto. No entanto, a sua abordagem baseia-se na função de distribuição e

não na função de sobrevivência, ao contrário do que é habitual nesta área.

Como temos vindo a constatar até agora, os modelos de cura pressupõem a existência de uma subpopulação de sobreviventes a longo prazo (habitualmente com tempos de vida censurados), designados por indivíduos curados ou imunes. Tsodikov *et al.* (1998) referiram que uma outra forma de abordar estes modelos é considerar que a existência de um factor de prognóstico não observado determinaria quais os indivíduos imunes. Surge então uma primeira referência à inclusão de uma variável fragilidade no contexto dos modelos de cura.

A doença de Hodgkin é um tipo especial de linfoma (cancro), identificada pelo aparecimento de uma célula típica chamada célula de Reed-Sterberg no exame microscópico. O seu tratamento, por norma agressivo, parece poder induzir o aparecimento de outros tipos de cancro, entre os quais a leucemia mielóide aguda. No entanto, não deixa de ser pertinente tentar saber se, de facto, a leucemia é induzida pelo tratamento ou se os doentes com doença de Hodgkin já têm uma maior propensão para o desenvolvimento de cancros.

Seja  $T$  o tempo de vida, que neste caso corresponde ao tempo desde o primeiro tratamento da doença de Hodgkin até ao aparecimento da leucemia. Como uma função de sobrevivência  $\bar{G}$ , ( $\bar{G} = 1 - G$ ) se pode escrever à custa da correspondente função hazard cumulativa,  $H$ , e como a função de sobrevivência de uma população em que existem indivíduos curados e indivíduos susceptíveis é imprópria, resulta que a função hazard cumulativa é limitada. Então, se  $H(t) \leq \theta$ , temos  $H(t) = \theta F(t)$ , onde  $F(t)$  é a função de distribuição de uma variável aleatória não negativa que representa o tempo de progressão do tumor, e a função de sobrevivência escreve-se na forma

$$\bar{G}(t) = \exp[-\theta F(t)]. \quad (3.6)$$

O modelo anterior pode ser generalizado de forma a incluir covariáveis, que tanto podem ser incluídas na função  $F$  como no parâmetro  $\theta$ . No entanto, se  $F$  não incluir covariáveis, o modelo resultante é um caso especial de modelo de hazards proporcionais.

Esta formulação permite que a estimação da taxa de cura seja feita sem ter que estimar o parâmetro perturbador ( $F$ ) em conjunto, uma vez que é dada por  $\exp(-\hat{\theta})$ .

Infelizmente as conclusões do estudo ficaram um pouco aquém do esperado, mais pela complexidade da situação do que pelo modelo adoptado.

Sobre o mesmo assunto e seguindo o mesmo argumento em relação à função hazard cumulativa, Tsodikov (1998) considerou um modelo de hazards proporcionais modificado para permitir a existência de sobreviventes a longo prazo, ou seja, novamente (3.6) mas com  $\theta = \theta(\mathbf{z})$ . De facto, como a função hazard cumulativa correspondente a  $\overline{G}(t)$  é dada por  $-\log \overline{G}(t) = \theta(\mathbf{z})F(t)$ , verifica-se que dados dois vectores de covariáveis  $\mathbf{z}_1$  e  $\mathbf{z}_2$ , a razão  $\theta(\mathbf{z}_1)F(t)/\theta(\mathbf{z}_2)F(t)$  não depende de  $t$ .

Quanto ao método de estimação dos parâmetros, encarou várias possibilidades, nomeadamente a verosimilhança parcial (a qual produz estimadores consistentes), a verosimilhança marginal (que permite a estimação de  $\theta(\mathbf{z})$  sendo a função  $F$  encarada como um parâmetro perturbador) e a verosimilhança *profile*, que é a verosimilhança completa maximizada com respeito a  $F$ , sendo esta a forma mais eficiente de estimar  $\theta$  se  $F$  é completamente desconhecida. Em relação ao último caso, o habitual é recorrer ao método de Newton-Raphson, apesar de poder apresentar alguns problemas de convergência. Por isso, foi sugerida uma outra abordagem, assente na resolução de equações algébricas aninhadas.

Num outro estudo sobre linfomas, mas desta vez linfoma não-Hodgkin, Peng *et al.* (1998) propuseram um modelo que usa a distribuição  $F$  generalizada, que designaram por modelo de mistura  $F$  generalizado. O modelo referido é uma extensão de modelos já anteriormente abordados, nomeadamente por Boag (1949), Gamel (1990), Farewell (1982), Ghitany *et al.* (1995) e muitos outros, uma vez que a distribuição  $F$  generalizada, dependendo do número de parâmetros e dos valores que tomam, apresenta como casos particulares a distribuição log-normal, Weibull, exponencial (entre outras). Visto por outro prisma, o modelo também pode ser considerado uma generalização de um modelo de tempo de vida acelerado, pois é constituído por duas partes: um modelo de tempo de vida acelerado (para os tempos de vida dos indivíduos susceptíveis) e um modelo de regressão logística (para a taxa de cura). Esta abordagem é semelhante à de Yamaguchi (1992), já referida anteriormente, residindo a diferença essencialmente na distribuição utilizada.

As vantagens encontradas devem-se em grande parte à flexibilidade da distribuição  $F$  generalizada, permitindo descobrir a estrutura dos dados sem impôr restrições à partida. Neste sentido, quase pode ser interpretado como um modelo semi-paramétrico. A principal desvantagem são as dificuldades computacionais na maximização da função de verosimilhança, em grande parte devido à função beta. Apesar de ser necessário algum cuidado ao

analisar os resultados deste modelo, essa precaução deve-se essencialmente à natureza mais complexa do problema em si e não ao demérito do modelo de mistura  $F$  generalizado. Aliás, o facto de poder ser usado em problemas mais complexos é outra das suas vantagens.

Gieser *et al.* (1998), interessados na estimação da taxa de cura, generalizaram o modelo descrito por Cantor e Shuster (1992), através da introdução de covariáveis. Consideraram a existência de  $I$  tratamentos,  $p$  covariáveis,  $z_1, z_2, \dots, z_p$  e a função de sobrevivência correspondente a um indivíduo no  $i$ -ésimo tratamento com vector de covariáveis  $\mathbf{z} = (z_1, z_2, \dots, z_p)$ ,  $S(t; i, \mathbf{z})$ . O modelo proposto é tal que

$$S(t; i, \mathbf{z}) = [S_{0,i}(t)]^{\exp(\boldsymbol{\gamma}'\mathbf{z})} = \exp \left[ \exp(\boldsymbol{\gamma}'\mathbf{z}) \beta_i^{-1} \alpha_i^* (1 - \exp(\beta_i t)) \right], i = 1, \dots, I, \quad (3.7)$$

onde  $\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \dots, \gamma_p)$  são os coeficientes de regressão associados a  $(z_1, z_2, \dots, z_p)$ , respectivamente, e  $S_{0,i}(t)$  é a função de sobrevivência específica do  $i$ -ésimo tratamento com parâmetros  $(\alpha_i^*, \beta_i)$ ,  $\alpha_i^* > 0$ ,  $\beta_i < 0$ . Para  $\mathbf{z}$  fixo e para o tratamento  $i$ , a função hazard correspondente é

$$\lambda(t; i, \mathbf{z}) = \exp(\alpha_i + \beta_i t + \boldsymbol{\gamma}'\mathbf{z}),$$

onde  $\alpha_i = \log \alpha_i^*$  e  $\lambda_{0,i}(t) = \exp(\alpha_i + \beta_i t)$ . Através desta representação, é imediato observar que o logaritmo da função hazard é linear nos parâmetros donde, em cada grupo de tratamento, o modelo (3.7) é um modelo de hazards proporcionais. Determinando o limite em (3.7) quando  $t \rightarrow \infty$ , os autores obtiveram a expressão para a taxa de cura:

$$\pi(i, \mathbf{z}) = S(\infty; i, \mathbf{z}) = \exp \left[ \beta_i^{-1} \exp(\boldsymbol{\gamma}'\mathbf{z} + \alpha_i) \right]. \quad (3.8)$$

A função de verosimilhança é

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^I \prod_{j=1}^{n_i} S(t_{ij}; i, \mathbf{z}_{ij}) [\lambda(t_{ij}; i, \mathbf{z}_{ij})]^{\delta_{ij}}. \quad (3.9)$$

Gieser *et al.* (1998) mostraram que o logaritmo da função de verosimilhança é uma função convexa e, conseqüentemente, é garantida a existência, unicidade e normalidade assintótica do estimador de máxima verosimilhança.

A estimação dos parâmetros não apresentou problemas, tendo sido usado o método de Newton-Raphson na maximização de (3.9).

Os coeficientes de regressão  $\gamma_i$  medem o efeito das covariáveis  $z_i$  na taxa de cura. Para testar  $H_0 : \gamma_i = 0$ , os autores usaram a estatística de teste  $Z = \hat{\gamma}_i / \text{s.e.}(\hat{\gamma}_i)$  a qual, sob  $H_0$ , tem uma distribuição assintótica  $N(0, 1)$ . Por (3.8), uma medida da diferença entre as taxas de cura correspondentes aos tratamentos  $i$  e  $j$  é expressa por

$$\theta_{ij} = \frac{\log \pi(i, \mathbf{z})}{\log \pi(j, \mathbf{z})} = \frac{\beta_j \exp \alpha_i}{\beta_i \exp \alpha_j}.$$

Assim, para testar a igualdade da taxa de cura nos grupos de tratamento  $i$  e  $j$ ,  $H_0 : \log \theta_{ij} = 0$ , Gieser *et al.* (1998) utilizaram a estatística de teste

$$Z = \frac{\hat{\alpha}_i + \log(-\hat{\beta}_j) - \log(-\hat{\beta}_i) - \hat{\alpha}_j}{\text{s.e.} \left[ \hat{\alpha}_i + \log(-\hat{\beta}_j) - \log(-\hat{\beta}_i) - \hat{\alpha}_j \right]}$$

que, sob  $H_0$ , também tem uma distribuição assintótica  $N(0, 1)$ .

Quanto ao exemplo que abordaram, além de ajustarem aos dados o modelo proposto, para efeitos de comparação também ajustaram o modelo de regressão de Cox, bem como o utilizado por Farewell (1982). O modelo (3.7), baseado na distribuição de Gompertz, apresentou estimativas dos coeficientes de regressão com valores semelhantes às estimativas obtidas com o modelo de regressão de Cox, mas tem a vantagem de fornecer adicionalmente estimativas das taxas de cura. As estimativas da taxa de cura obtidas através do modelo (3.7) ou através do modelo de Farewell (1982), apesar de terem valores próximos dos obtidos através do valor máximo da estimativa de Kaplan-Meier da função de sobrevivência, apresentaram desvios padrão substancialmente mais pequenos.

O modelo (3.7) pode ser generalizado um pouco mais com a introdução de covariáveis na função hazard subjacente, ou seja, em  $\lambda_{0,i}(t)$ . Na prática, esta generalização é útil quando o modelo de hazards proporcionais (3.7) não se ajusta aos dados dentro de cada grupo de tratamento mas há um bom ajustamento para certas combinações apropriadas das covariáveis e dos tratamentos.

Uma outra alternativa proposta para a inclusão das covariáveis é o modelo

$$S(t; \mathbf{z}) = [\pi + (1 - \pi) \exp(\lambda t)]^{\exp(\gamma' \mathbf{z})}, \quad \lambda < 0,$$

ou uma sua generalização em que a função hazard subjacente é específica para cada grupo de tratamento, apesar de poder apresentar algumas dificuldades computacionais.

Também em 1998, no âmbito da Análise Bayesiana, Chao desenvolveu um modelo de cura de riscos competitivos, onde não é assumida a independência das causas de morte. O exemplo que estudou refere-se a indivíduos com leucemia (crónica ou aguda), mielodisplasia ou linfoma sujeitos a transplante da medula óssea, em que a situação após tratamento (transplante de medula óssea) pode ser: curado pelo tratamento, recaída da doença ou morte sem ocorrência de recaída (mas relacionada com o tratamento). Neste caso, só os dois últimos acontecimentos são observáveis, dando origem a observações não censuradas, visto que o tempo em estudo é o tempo decorrido até à falha do transplante. Note-se que embora a morte possa ocorrer após a recaída, nesta situação o tempo de vida é o tempo até à recaída e não o tempo até à morte. Uma observação censurada corresponde a um indivíduo vivo e livre de recaída no fim do seu *follow-up* e é apenas nesta situação que a cura pode ocorrer. Assim, sem considerar que a cura é observável, o modelo de mistura desenvolvido providencia não só uma estimativa da probabilidade de cura, condicional ao tempo de censura, como também uma estimativa da probabilidade de recaída, condicional a não estar curado.

Considere-se a variável  $C$  que toma o valor zero (quando o indivíduo está curado) com probabilidade  $p$  e o valor um com probabilidade  $q = 1 - p$ ,  $X$  a variável que representa o tempo que decorre desde o tratamento até à morte e  $x$  o respectivo tempo de vida observado (censurado ou não). A probabilidade de sobreviver para além do instante  $x$ , condicional a estar curado, é representada por  $S(x)$ , enquanto que  $G(x; \boldsymbol{\theta})$  designa a função de sobrevivência dos indivíduos não curados. O modelo de mistura inicialmente considerado foi

$$P(X > x | \boldsymbol{\eta}) = pS(x) + qS(x)V(x; \boldsymbol{\theta}),$$

onde  $\boldsymbol{\eta} = (p, \boldsymbol{\theta})$  é o vector de parâmetros do modelo e  $V(x; \boldsymbol{\theta}) = G(x; \boldsymbol{\theta})/S(x)$  é uma função de sobrevivência relativa tal que  $0 \leq V(x; \boldsymbol{\theta}) \leq 1$ . Em termos da função hazard, o modelo pode ser escrito na forma

$$h(x; \boldsymbol{\eta}) = \mu(x) + \frac{qV(x; \boldsymbol{\theta})}{p + qV(x; \boldsymbol{\theta})} \frac{v(x; \boldsymbol{\theta})}{V(x; \boldsymbol{\theta})},$$

onde  $\mu(x)$  designa a função hazard para os indivíduos curados e  $v(x; \boldsymbol{\theta}) = -(d/dx)V(x; \boldsymbol{\theta})$ , ou mais simplesmente,

$$h(x; \boldsymbol{\eta}) = \mu(x) + \frac{qv(x; \boldsymbol{\theta})}{p + qV(x; \boldsymbol{\theta})}.$$

A probabilidade *a posteriori* de um indivíduo não estar curado, dado que o seu tempo de vida é maior do que  $x$ , é dada por

$$P(\bar{C}|X > x, \boldsymbol{\eta}) = \frac{qV(x; \boldsymbol{\theta})}{p + qV(x; \boldsymbol{\theta})}.$$

Note-se a analogia com a maioria dos modelos referidos até agora, onde para os indivíduos curados se considerou  $S(x) = 1, \forall x < \infty$ , do que resultaria neste caso  $V(x; \boldsymbol{\theta}) = G(x; \boldsymbol{\theta})$ .

No entanto, atendendo a que o exemplo considera dois tipos de falha de transplante (recaída ou morte sem recaída), houve necessidade de construir um modelo que tivesse em conta qual a causa de falha. Então, Chao (1998) considerou a variável  $R$ , que tem o valor 1 para os indivíduos que sofreram recaída, e zero para os que morreram sem terem tido recaída e  $\gamma = P(R|\bar{C})$  a probabilidade de um indivíduo ter recaída dado que não está curado. O índice  $r$  foi associado aos indivíduos não curados que sofreram recaída e o índice  $d$  aos não curados que morreram sem recaída. Note-se que o tempo de vida observado,  $x$ , não é mais do que o mínimo entre o tempo até à recaída, o tempo até à morte sem ocorrência de recaída e o tempo de censura. A função de sobrevivência é então dada por

$$P(X > x|\boldsymbol{\eta}) = p + q[\gamma V_r(x; \boldsymbol{\theta}_r) + (1 - \gamma)V_d(x; \boldsymbol{\theta}_d)],$$

com  $V_r(x; \boldsymbol{\theta}_r)$  a função de sobrevivência relativa para os indivíduos que tiveram recaída e  $V_d(x; \boldsymbol{\theta}_d)$  a função de sobrevivência relativa para os indivíduos que morreram sem terem tido recaída. Quando  $V_r(x; \boldsymbol{\theta}_r)$  e  $V_d(x; \boldsymbol{\theta}_d)$  são funções correspondentes à distribuição exponencial, a probabilidade *a posteriori* de cura, dado que o tempo de vida é superior ao tempo de censura  $x$ , é

$$P(C|x; p, \gamma, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d) = \frac{p}{p + q [\gamma \exp(-\boldsymbol{\theta}_r x) + (1 - \gamma) \exp(-\boldsymbol{\theta}_d x)]},$$

enquanto que a probabilidade *a posteriori* de recaída, dado que o indivíduo é não curado e dado o tempo de censura  $x$  é

$$P(R|x, \bar{C}; p, \gamma, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d) = \frac{\gamma \exp(-\boldsymbol{\theta}_r x)}{\gamma \exp(-\boldsymbol{\theta}_r x) + (1 - \gamma) \exp(-\boldsymbol{\theta}_d x)}.$$

As estimativas do modelo foram determinadas através do método de amostragem Gibbs. Quando os parâmetros são substituídos pelas suas estimativas, estas duas últimas probabilidades podem ser utilizadas para atribuir o estatuto de cura ou de recaída, respectivamente, aos indivíduos com observações censuradas.

Já em 1999, De Angelis *et al.* debruçaram-se sobre modelos de mistura logístico/exponencial e logístico/Weibull com inclusão de covariáveis, com o objectivo de estudar a sobrevivência de indivíduos com cancro do cólon, com base no Registo Nacional de Cancro da Finlândia. Assim, consideraram que os indivíduos em estudo podiam ser divididos em dois grupos distintos: (i) uma proporção  $P$  de indivíduos curados para a doença em causa, que têm a mesma função hazard ( $h_0$ ), função densidade ( $f_0$ ) e função de sobrevivência ( $S_0$ ) do que a população em geral que seja comparável; e (ii) uma proporção  $(1 - P)$  de indivíduos susceptíveis que vão morrer da doença em causa, com função hazard ( $h_1$ ), função de sobrevivência ( $S_1$ ) e função densidade ( $f_1$ ) dadas, respectivamente, por

$$h_1(t, \boldsymbol{x}) = h_0(t, \boldsymbol{x}) + h_D(t, \boldsymbol{x}),$$

$$S_1(t, \boldsymbol{x}) = \exp \left[ - \int_0^t h_1(u, \boldsymbol{x}) du \right] = S_0(t, \boldsymbol{x}) S_D(t, \boldsymbol{x}),$$

$$f_1(t, \boldsymbol{x}) = h_1(t, \boldsymbol{x}) S_1(t, \boldsymbol{x}),$$

onde  $t$  representa o tempo de vida de um indivíduo,  $\mathbf{x}$  o respectivo vector de covariáveis,  $h_D$  o acréscimo do risco de morte devido à doença e  $S_D$  a correspondente função de sobrevivência relativa. A proporção de indivíduos curados é modelada por

$$P(\mathbf{x}) = \frac{1}{1 + \pi_0 \exp(\boldsymbol{\pi} \mathbf{x})},$$

onde  $\pi_0$  e  $\boldsymbol{\pi}$  são os parâmetros de regressão ( $\boldsymbol{\pi} = \pi_1, \dots, \pi_{k1}$ ) e a função de sobrevivência relativa para os dois grupos em conjunto é dada por

$$S(t, \mathbf{x}) = P(\mathbf{x}) + (1 - P(\mathbf{x})) S_D(t, \mathbf{x}).$$

Para a função de sobrevivência relativa  $S_D$ , foram considerados o modelo exponencial,

$$S_D(t, \mathbf{x}) = \exp[-\lambda(\mathbf{x})t]$$

e o modelo Weibull,

$$S_D(t, \mathbf{x}) = \exp\left[-(\lambda(\mathbf{x})t)^\beta\right].$$

Para o parâmetro de escala  $\lambda(\mathbf{x})$  que coincide com a função  $h_D$  no modelo exponencial, é adoptado um modelo de hazards proporcionais

$$\lambda(\mathbf{x}) = \gamma_0 \exp(\boldsymbol{\gamma} \mathbf{x}),$$

onde  $\gamma_0$  e  $\boldsymbol{\gamma}$  são os parâmetros de regressão ( $\boldsymbol{\gamma} = \gamma_1, \dots, \gamma_{k1}$ ).

Relativamente ao efeito das covariáveis, os dois tipos de modelos apresentaram resultados semelhantes; quanto à estimação da proporção de curados, o modelo de regressão logística/exponencial revelou uma estimativa demasiado optimista, pelo que o modelo de regressão logística/Weibull foi o mais adequado. Os autores salientaram as vantagens dos modelos de mistura para este tipo de estudo, dado permitirem a avaliação simultânea dos factores associados à probabilidade de cura e dos factores que afectam o tempo de vida

dos casos fatais. Afirmam mesmo que a consideração de que as populações são homogêneas pode camuflar características importantes para a sobrevivência. Porém, a validade dos resultados deste tipo de modelos depende fortemente da escolha adequada das funções de distribuição envolvidas.

No mesmo ano, também no âmbito da Análise Bayesiana, Chen *et al.* (1999) propuseram um modelo um pouco diferente dos anteriores, sobretudo na sua formulação, mais adequada a uma interpretação clínica directa, embora pareça ser algo semelhante ao modelo já referido por Tsodikov (1998). A função de sobrevivência da população é dada por

$$S_p(t) = \exp(-\theta F(t)), \quad (3.10)$$

onde  $\theta$  está relacionado com o número inicial de células carcinogénicas e  $F(t)$  com a sua taxa de progressão, visto que  $F(t)$  representa a função de distribuição do tempo necessário para a  $i$ -ésima célula originar uma massa tumoral. Note-se que  $S_p(\infty) = \exp(-\theta) > 0$ , donde  $S_p(t)$  é uma função de sobrevivência imprópria. As correspondentes funções densidade e hazard são, respectivamente,

$$f_p(t) = \theta f(t) \exp(-\theta F(t)) \quad \text{e} \quad h_p(t) = \theta f(t),$$

com  $f(t) = (d/dt)F(t)$ . Em relação aos indivíduos susceptíveis, Chen *et al.* (1999) deduziram as seguintes funções:

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

$$\text{e} \quad h^*(t) = \left( \frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)} \right) \theta f(t).$$

A função hazard pode também ser escrita na forma

$$h^*(t) = \left( \frac{1}{P(T < \infty | T > t)} \right) h_p(t)$$

para evidenciar que  $h^*(t) > h_p(t)$ .

Uma das grandes vantagens desta abordagem é o facto de o modelo de hazards proporcionais ser válido para toda a população se  $h_p(t; \mathbf{z}) = \theta(\mathbf{z})f(t)$ , mas então tem como desvantagem já não ser um modelo de hazards proporcionais para os indivíduos susceptíveis. Do ponto de vista bayesiano, conduz a distribuições *a priori* e *a posteriori* com características desejáveis, para além de ser um modelo computacionalmente atractivo. São propostas algumas possíveis extensões (em estudo), nomeadamente, uma ligeira modificação no modelo de forma a permitir uma maior heterogeneidade entre os indivíduos, e uma versão multivariada de (3.10).

Sy e Taylor (2000), voltaram a abordar o modelo descrito por Kuk e Chen (1992), apresentando um método de estimação diferente para os parâmetros de regressão: usaram o algoritmo EM em vez do método de Monte Carlo. Este trabalho vem na sequência do descrito em Taylor (1995), que considerou que as covariáveis influenciavam apenas a incidência. Agora os autores consideram também a sua influência na função de sobrevivência dos susceptíveis.

O exemplo que estudaram refere-se ao carcinoma das células escamosas das amígdalas. Algumas covariáveis presentes no modelo revelaram-se importantes para a latência e para a incidência, outras apenas para a incidência (dose e duração do tratamento) ou vice-versa (caso da idade) e outras nem para a latência nem para a incidência.

Uma observação relevante neste artigo é o facto de o modelo de cura de hazards proporcionais, ou seja, o modelo de cura de mistura em que as funções hazard dos indivíduos susceptíveis são proporcionais, constituir um caso particular do modelo multiplicativo com fragilidade (do qual fazemos referência no capítulo 6). No entanto, foi feita a ressalva de que neste caso a variável fragilidade não era inteiramente não observável (quando era observado o acontecimento de interesse), logo não era uma fragilidade no sentido literal.

Paralelamente, Peng e Dear (2000) também usaram o algoritmo EM na estimação dos parâmetros de regressão do modelo de Kuk e Chen (1992). Mais precisamente, o seu método de estimação não paramétrico é uma combinação da verosimilhança marginal usada para o modelo de regressão de Cox e do algoritmo EM. Na ausência de covariáveis para os indivíduos susceptíveis, o modelo e o modo de estimação reduz-se ao já abordado por Taylor (1995). Por outro lado, quando não existe uma proporção de curados, o modelo e o método de estimação resume-se ao modelo de regressão de Cox, logo pode

ser considerado uma generalização ao caso de existência de uma proporção de curados.

Bernardo e Ibrahim (2000) tornaram a abordar a questão de testar a igualdade das funções de sobrevivência de dois grupos de tratamento, dando sequência a um trabalho já desenvolvido pelo segundo autor. Fizeram uma abordagem bayesiana, mas os seus resultados são uma generalização dos obtidos por Gray e Tsiatis (1989), por também se basearem no teste *log-rank* e por admitirem que quer a taxa de cura quer a função de sobrevivência dos indivíduos susceptíveis pode diferir nos dois grupos.

Como habitualmente, consideraram a função de sobrevivência descrita por

$$Q(t) = \theta + (1 - \theta)H(t),$$

onde  $T$  é a variável aleatória que representa o tempo de vida,  $\theta$  é a proporção de indivíduos curados e  $H(t)$  é a função de sobrevivência dos indivíduos susceptíveis. Testaram a hipótese  $H_0 : Q_1(t) = Q_2(t)$  versus  $H_1 : Q_1(t) \neq Q_2(t)$ , onde  $Q_j(t)$  é a função de sobrevivência dos indivíduos submetidos ao tratamento  $j$ , com  $j = 1, 2$ .

O trabalho desenvolvido, na nossa opinião, tem uma importância prática bastante acentuada, uma vez que permite que os ensaios clínicos tenham uma duração mais curta. Para além das vantagens económicas que daí advêm, importa realçar as vantagens para os doentes pois, ou recebem o melhor tratamento mais cedo, ou não estão tanto tempo sujeitos ao tratamento experimental, se este não se revelar melhor que o tratamento padrão ou mesmo se apresentar efeitos secundários mais nefastos.

Em Tamura *et al.* (2000) encontrámos um exemplo relacionado com o tempo até à resposta a um tratamento com anti-depressivos, sendo dos poucos que não se referem a doentes com cancro. Talvez por isso, o modelo estudado apresentou alguns pormenores um pouco diferentes. Os autores usaram um modelo de cura, que tem em linha de conta a existência de indivíduos que não respondem ao tratamento e, para os que respondem, modelaram o tempo que decorre até à resposta. Como consideraram dois grupos de tratamento, para o  $i$ -ésimo ( $i = 1, 2$ ) grupo,  $H_i(t)$  representa a probabilidade de que a resposta ocorra após o instante  $t$ , e  $p_i$  a proporção de indivíduos que respondem ao tratamento. Além disso,  $T_{ij}$  representa o tempo de resposta do  $j$ -ésimo indivíduo no  $i$ -ésimo grupo de tratamento e toma valores em  $(0, u)$  para os indivíduos que respondem, uma vez que é assumido que a resposta

só ocorre nesse intervalo (todos os indivíduos que tenham tempo superior a  $u$  são considerados não respondentes, ou seja, não susceptíveis ao tratamento). Então,

$$H_i(t) = p_i S_i(t) + (1 - p_i), \quad 0 \leq t \leq u,$$

onde  $S_i(t) = P(T_{ij} > t | T_{ij} \leq u)$  é a probabilidade de a resposta ocorrer após o instante  $t$ , dado que ocorre. Claro que  $u$  varia de exemplo para exemplo, mas em estudos de depressão é prática corrente considerar  $u$  igual a 6 semanas. Em qualquer dos casos,  $u$  tem que ser suficientemente grande de forma a que a identificação dos indivíduos que respondem seja inequívoca. A resposta ao tratamento não deve ser confundida com a ausência de sintomas da doença, situação que pode ocorrer para além do instante  $u$ . Note-se que neste caso, e ao contrário do que costuma acontecer com os estudos do foro oncológico, é muito fácil identificar quais os indivíduos susceptíveis (são os que respondem ao tratamento no prazo de 6 semanas) e quais os não susceptíveis (são os que não respondem nesse prazo).

Dado que muitas vezes os pacientes podem estar em risco de cometer suicídio, o “melhor” anti-depressivo será aquele que não só produza efeito na maioria dos pacientes, como sobretudo que o faça no mais curto espaço de tempo possível. Neste último factor residiu a motivação para preferir comparar a eficiência de dois anti-depressivos com base nas funções de sobrevivência dos indivíduos e não nas correspondentes proporções de respondentes, como já tinham feito Gray e Tsiatis (1989) e Laska e Meisner (1992). Assim, a hipótese testada foi  $H_0 : S_1(t) = S_2(t)$ , com recurso ao teste de Cramer-von Mises e também à metodologia *bootstrap*. Como o  $p$ -value obtido foi 0.30, Tamura *et al.* (2000) concluíram que não existia evidência de que um anti-depressivo fosse melhor do que o outro.

Voltando ao problema de testar a presença de indivíduos curados, Peng *et al.* (2001) realizaram um estudo de simulação, no qual pretendiam analisar a distribuição da estatística de teste do teste de razão de verosimilhanças,  $d_n$ , quando a distribuição do tempo de vida dos indivíduos susceptíveis é a distribuição de Weibull ou a distribuição log-normal. Este estudo é uma generalização do já desenvolvido por Maller e Zhou (1995), que utilizaram a distribuição exponencial. Além disso, também pretendiam saber qual a dimensão da amostra necessária para que  $d_n$  tivesse aproximadamente a mesma

distribuição, quer fosse utilizada a distribuição exponencial, Weibull ou log-normal para o tempo de vida dos indivíduos susceptíveis.

Nalgumas situações, a presença de indivíduos curados pode ser inferida através da experiência clínica da doença ou através de evidência biológica. No entanto, quando tal não é possível, interessa desenvolver um teste para a sua presença, baseado no aparecimento de um número relativamente grande de observações censuradas após a maior observação não censurada. Seja  $\pi$  a proporção de indivíduos susceptíveis. A hipótese de não existência de indivíduos curados é dada por  $H_0 : \pi = 1$  (hipótese na fronteira do espaço de parâmetros:  $(0, 1]$ ), que pode ser testada através do teste de razão de verossimilhanças, baseado na estatística

$$d_n = -2 [\ell_0(\theta) - \ell_1(\theta, \pi)],$$

onde  $\ell_1(\theta, \pi)$  é o logaritmo da função de verossimilhança maximizada sob o modelo de mistura habitual e  $\ell_0(\theta)$  é o logaritmo da função de verossimilhança maximizada sob a hipótese  $H_0$ . Como Maller e Zhou (1995) observaram para a distribuição exponencial, a distribuição de  $d_n$  é uma mistura de uma distribuição Qui-quadrado com um grau de liberdade e uma distribuição degenerada em zero, em que as proporções de mistura são ambas iguais a  $1/2$ , ou seja,

$$P(d_n < t) \rightarrow \frac{1}{2} + \frac{1}{2}P(\chi_1^2 < t). \quad (3.11)$$

Quando, em vez da distribuição exponencial é utilizada a distribuição log-normal ou Weibull, a distribuição (3.11) é apenas uma aproximação da distribuição assintótica de  $d_n$ . Se a amostra for pequena, essa aproximação é tanto pior quanto menor for a censura (10% ou menos) e quanto maior for o valor da função hazard para valores elevados de  $t$ . Qualquer uma destas situações origina uma aproximação considerada inaceitável, ainda que a dimensão da amostra seja moderada. De facto, nestes casos, a verdadeira distribuição atribui um peso maior à distribuição degenerada em zero. Os autores remetem para um artigo posterior a questão da potência do teste e da determinação da dimensão da amostra.

Yau e Ng (2001) generalizaram o uso de modelos de mistura à análise multivariada do tempo de vida, com uma abordagem através dos modelos de

mistura lineares generalizados. Este tipo de modelos surge para dar resposta a situações em que ou a população se encontra agrupada em *clusters* ou para cada indivíduo podem ser contabilizadas várias falhas do mesmo tipo (por exemplo, várias recaídas de uma mesma doença). Em qualquer dos casos, existe algum tipo de correlação entre alguns dos tempos de vida. No exemplo que os autores consideraram, os objectivos foram identificar os factores de risco, avaliar o seu efeito no tempo de vida e comparar os dois tratamentos com base na sobrevivência dos pacientes. O tempo de vida foi definido como o tempo desde o diagnóstico até à morte pela doença; para os indivíduos em que a morte foi observada mas sem que estivesse relacionada com a doença em causa, o tempo de vida foi considerado censurado.

Para um indivíduo com vector de covariáveis  $\mathbf{x}_i$ ,  $i = 1, \dots, N$ , o habitual é que a proporção  $\pi$  seja especificada pela função logística

$$\pi(\mathbf{x}_i) = P(Y = 1|\mathbf{x}_i) = \frac{1}{1 + \exp(\boldsymbol{\xi}_i)}, \quad (3.12)$$

onde  $Y$  é a variável indicatriz de susceptibilidade,  $\boldsymbol{\xi}_i = \mathbf{w}_i' \boldsymbol{\gamma}$ ,  $\mathbf{w}_i = [1 \ \mathbf{x}_i']$ , e  $\boldsymbol{\gamma}$  é o vector de parâmetros de regressão. Relativamente à função de sobrevivência, é usual adoptar o modelo

$$S(t; \mathbf{x}) = 1 - \pi(\mathbf{x}) + \pi(\mathbf{x})S_2(t; \mathbf{x}), \quad (3.13)$$

onde  $S_2(t; \mathbf{x})$  designa a função de sobrevivência dos indivíduos susceptíveis, ou seja, daqueles que não respondem favoravelmente ao tratamento. Se para os indivíduos susceptíveis for considerado o modelo de regressão de Cox, isto é, se

$$h_2(t; \mathbf{x}) = h_{20}(t) \exp(\boldsymbol{\eta}), \quad (3.14)$$

tal que  $\boldsymbol{\eta} = \mathbf{x}' \boldsymbol{\beta}$ ,  $\boldsymbol{\beta}$  é o vector de parâmetros de regressão e  $h_{20}(t)$  é a função hazard subjacente dos indivíduos susceptíveis, a função hazard populacional escreve-se na forma

$$h(t; \mathbf{x}) = \pi(\mathbf{x})h_{20}(t) \exp(\boldsymbol{\eta}) \frac{S_{20}(t)^{\exp(\boldsymbol{\eta})}}{S(t; \mathbf{x})}.$$

Os pacientes provieram de seis instituições distintas. Este facto fez com que a eventual variabilidade (não observada) existente entre as instituições fosse modelada por um efeito aleatório, ou seja, através de uma fragilidade.

Na análise multivariada, a expressão (3.12) escreve-se na forma

$$\pi(\mathbf{x}_{ij}) = \frac{1}{1 + \exp(\boldsymbol{\xi}_{ij})}, \quad \boldsymbol{\xi}_{ij} = \mathbf{w}'_{ij}\boldsymbol{\gamma} + V_i \quad i = 1, \dots, M; \quad j = 1, \dots, n_i,$$

onde  $V_i$  representa o efeito aleatório não observado da  $i$ -ésima instituição na proporção de indivíduos curados,  $M$  designa o número de instituições,  $n_i$  indica o número de observações na  $i$ -ésima instituição e  $N = \sum_{i=1}^M n_i$ . Yau e Ng (2001) admitiram que  $V \sim N(0, \theta_1)$ . Note-se que, de acordo com a definição do efeito aleatório, valores elevados de  $V$  correspondem a uma maior probabilidade de cura.

Quanto à expressão (3.14), passa a escrever-se na forma

$$h_2(t_{ij}; \mathbf{x}_{ij}) = h_{20}(t_{ij}) \exp(\boldsymbol{\eta}_{ij}), \quad \boldsymbol{\eta}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + U_i,$$

com  $U_i$  a representar o efeito aleatório não observado da  $i$ -ésima instituição no risco de formação de tumor para os indivíduos susceptíveis. Os autores admitem que  $U \sim N(0, \theta_2)$ . Um valor elevado de  $U_i$  indica que os pacientes da  $i$ -ésima instituição têm um maior risco de terem um tumor, se pertencem ao grupo dos indivíduos susceptíveis.

Yau e Ng (2001) salientam que, à semelhança do que acontece quando o modelo (3.13) é adoptado, também no caso da análise multivariada podem surgir problemas de não identificabilidade do modelo e estimativas dos parâmetros de regressão infinitas, em especial quando não é clara a existência de indivíduos curados.

Broët *et al.* (2001) desenvolveram testes para detectar diferenças entre as funções de sobrevivência de dois grupos, admitindo em qualquer um deles a presença de indivíduos imunes. Diferem dos testes propostos até à data, entre outras coisas, por partirem de um modelo de cura de não-mistura e não do modelo de mistura habitual.

Seja  $n_i$  o número de indivíduos em cada grupo ( $i = 0, 1$ ,  $n = n_0 + n_1$ ). Para o  $j$ -ésimo paciente no  $i$ -ésimo grupo, sejam  $T_{ij}$  e  $C_{ij}$  as variáveis aleatórias que representam o tempo de vida e o tempo de censura, respectivamente,  $X_{ij} = \min(T_{ij}, C_{ij})$  o tempo de *follow-up* observado com função

de sobrevivência  $H_i(t)$ ,  $\delta_{ij} = 1_{(X_{ij}=T_{ij})}$  a variável de morte,  $Y_{ij} = 1_{(t \leq X_{ij})}$  a variável indicatriz de o indivíduo estar em risco no instante  $t$  e  $Z_j$  a variável indicatriz do grupo 1. A função hazard de  $T_{ij}$  é denotada por  $\lambda_i(t)$  e a correspondente função hazard cumulativa por  $\Lambda_i(t)$ .

Para o  $i$ -ésimo grupo, a função de sobrevivência foi definida por

$$S_i(t) = \exp[-\theta \exp(\beta_1 i)[1 - \exp(-K(t) \exp(\beta_2 i))]], \quad (3.15)$$

onde  $K(t)$  representa uma função contínua positiva crescente (de zero a infinito) não especificada e  $\theta$  é um parâmetro positivo. Este modelo é uma generalização do modelo descrito em Cantor e Shuster (1992), em que  $K(t) = (\alpha/\theta)t$ , com  $\alpha$  parâmetro positivo. Por (3.15), verifica-se que  $\lim_{t \rightarrow \infty} \Lambda_i(t) = \theta \exp(\beta_1 i)$ , logo a hipótese de  $\beta_1 = 0$  significa ausência de efeito a longo prazo. A outra hipótese de interesse,  $\beta_2 = 0$ , é a hipótese de proporcionalidade das funções hazard.

A nossa revisão bibliográfica, primeira etapa crucial neste trabalho, terminou neste ano (2001), embora a actualização fosse uma preocupação constante. Assim, ao longo do trabalho podem ser encontradas algumas referências posteriores, mas já inseridas no contexto próprio, nomeadamente Choi e Zhou (2002), Maller e Zhou (2002), Phillips *et al.* (2002), Sposto (2002), Ng e McLachlan (2003), Steele (2003), Tsodikov (2003), Tsodikov *et al.* (2003), Shao e Zhou (2004) e Yu *et al.* (2004).

A continuação da proliferação de artigos sobre modelos de cura é um forte indicador do interesse que esta área da estatística desperta, bem como da necessidade do seu desenvolvimento.



# Capítulo 4

## Modelos de Cura

### 4.1 Introdução

Tipicamente, a Análise de Sobrevivência consiste na análise de dados referentes a medições do tempo decorrido desde um instante inicial até à ocorrência de um determinado acontecimento, numa população. No entanto, nem sempre é possível observar o acontecimento de interesse, situação em que surgem as chamadas observações censuradas. Geralmente, tais observações devem-se à existência de certas restrições no processo de recolha dos dados.

Contudo, se a maioria das observações censuradas corresponder às maiores observações, podemos suspeitar da existência de indivíduos que nunca vão experimentar o acontecimento de interesse, ainda que o tempo em observação pudesse ser prolongado indefinidamente. Tais indivíduos são considerados indivíduos imunes ao acontecimento de interesse e, nesta perspectiva, a função de sobrevivência que lhes irá corresponder tomará sempre o valor um em qualquer instante (função de sobrevivência degenerada). Assim, numa população poderemos ter uma mistura de indivíduos imunes (ou não susceptíveis, ou curados) e de indivíduos não imunes (ou susceptíveis, ou doentes); a estes últimos irá corresponder uma função de sobrevivência não degenerada.

A observação do acontecimento de interesse permite que os indivíduos sejam classificados como susceptíveis. Quando esse acontecimento não é observado, em geral não é possível identificar os indivíduos imunes, apesar de a sua presença ser admitida. Apenas no caso em que é definido algum critério de classificação para os indivíduos curados é possível identificar inequivoca-

mente esses indivíduos. Laska e Meisner (1992), Taylor (1995), Tamura *et al.* (2000) e Betensky e Schoenfeld (2001), fornecem dos poucos exemplos desta situação. Em Laska e Meisner (1992), em Taylor (1995) e em Tamura *et al.* (2000) é definido um valor *a priori* e os indivíduos que tenham um tempo de vida superior a esse valor são considerados curados. Em Betensky e Schoenfeld (2001), a ocorrência de um certo acontecimento (alta hospitalar) distinto do acontecimento de interesse (morte pela doença) vai possibilitar a identificação dos indivíduos curados.

Sem dúvida que a situação mais usual é aquela em que os indivíduos curados podem ser encontrados entre aqueles a que correspondem os maiores tempos de vida censurados. Neste caso, entre os indivíduos com tempos de vida censurados, também poderão estar alguns indivíduos susceptíveis.

Neste capítulo, vamos começar por definir formalmente os modelos de cura de mistura e os modelos de cura de não-mistura, bem como estabelecer a relação entre ambos. No âmbito dos modelos de cura de não-mistura, propomos uma nova distribuição imprópria, a qual pode ser considerada uma generalização da distribuição de Gompertz modificada. A questão da estimação dos parâmetros nos modelos de cura é algo que não se encontra com facilidade na literatura, pelo menos, não com grande detalhe. Por isso, entendemos dar o nosso contributo neste aspecto, ilustrando a metodologia com um exemplo de aplicação a dados reais da distribuição log-logística para o tempo de vida dos indivíduos susceptíveis. Em seguida, propomos também um novo modelo de cura de mistura paramétrico, uma vez que se baseia numa distribuição ainda não usada neste contexto.

## 4.2 Modelos de cura

Os modelos de cura são modelos desenvolvidos para contemplar situações em que é admissível a existência de alguns indivíduos curados na população. Embora todos estes modelos tenham como característica comum o facto de a função de sobrevivência populacional ser uma função imprópria, são divididos em dois grupos: os modelos de cura de mistura e os modelos de cura de não-mistura. No entanto, é usual designar abreviadamente qualquer um destes modelos por modelo de cura, sempre que seja claro no contexto qual o tipo que está a ser considerado.

Apesar de a esmagadora maioria das aplicações destes modelos ser na área da saúde, algumas distribuições impróprias já se têm revelado mais

adequadas do que distribuições próprias para representar dados relativos a casamentos, divórcios e mobilidade profissional, conforme refere Yamaguchi (1992).

### 4.2.1 Modelo de cura de mistura

Seja  $T$  uma v.a. que representa o tempo de vida de um indivíduo numa dada população, na qual se admite existirem indivíduos imunes e indivíduos susceptíveis. Seja  $S_d(t)$  a função de sobrevivência do tempo de vida dos indivíduos susceptíveis e  $p$  a proporção de indivíduos imunes na população. A função de sobrevivência de  $T$  (populacional) é

$$S(t) = p + (1 - p)S_d(t). \quad (4.1)$$

Este caso particular dos modelos de mistura é designado por modelo de *mistura não padrão*, ou seja, refere-se ao caso de uma mistura de duas distribuições, uma das quais degenerada. Note-se que  $S(t)$  é uma f.s. imprópria pois  $S(\infty) = p$ , enquanto que  $S_d(t)$  é uma f.s. própria.

Um outro modo de formular o modelo de cura, de acordo com Farewell (1977, 1982), Ng e McLachlan (1998), entre muitos outros, consiste em definir uma v.a. binária  $Y$ , onde  $Y = 1$  se para um determinado indivíduo o acontecimento ocorrerá, isto é, se o indivíduo for susceptível, e  $Y = 0$  se o acontecimento nunca ocorrer. A função de sobrevivência condicional a  $Y = 0$  é sempre 1, para qualquer valor finito de  $t$ . Admite-se que um indivíduo tem probabilidade  $(1 - p)$  de ser susceptível e probabilidade  $p$  de ser imune. Assim, a função de sobrevivência de  $T$  é

$$S(t) = p + (1 - p)S(t|Y = 1). \quad (4.2)$$

Tendo em conta a igualdade  $h(t) = -\frac{d \log S(t)}{dt}$ , podemos definir o modelo de cura em termos da função hazard do seguinte modo:

$$h(t) = \frac{(1 - p)f(t|Y = 1)}{S(t)}. \quad (4.3)$$

Importa realçar que a função hazard definida por (4.3) tem integral finito em  $(0, \infty)$ , logo não se trata de uma função hazard no sentido usual visto corresponder a uma distribuição imprópria, conforme referimos no capítulo 1.

Note-se que (4.2) é equivalente a

$$S(t) = pS(t|Y = 0) + (1 - p)S(t|Y = 1),$$

ou seja, a função de sobrevivência populacional é escrita à custa das funções de sobrevivência das respectivas subpopulações, sendo  $S(t|Y = 0) = 1, \forall t$ . Também a correspondente função hazard pode ser escrita como uma mistura das funções hazard correspondentes. De facto, temos

$$h(t) = w_c(t)h(t|Y = 0) + w_d(t)h(t|Y = 1),$$

onde

$$w_d(t) = 1 - w_c(t) = \frac{(1 - p)S(t|Y = 1)}{pS(t|Y = 0) + (1 - p)S(t|Y = 1)} = \frac{(1 - p)S(t|Y = 1)}{S(t)}.$$

Repare-se que  $h(t|Y = 0)$  é sempre zero, qualquer que seja o valor de  $t$ , uma vez que é a derivada de uma constante, o que aliás está de acordo com o de facto representar o risco de morte dos indivíduos curados. Assim sendo,

$$h(t) = \frac{(1 - p)S(t|Y = 1)h(t|Y = 1)}{S(t)}. \quad (4.4)$$

Neste tipo de modelos de cura interessa estimar não só os parâmetros da função de distribuição dos indivíduos susceptíveis, como também a proporção de indivíduos curados. Maller e Zhou (1992) propuseram uma estimativa não paramétrica desta proporção, que é obtida através do valor da estimativa de Kaplan-Meier da função de sobrevivência na maior observação.

Para ter em conta a heterogeneidade existente entre os indivíduos relativamente a eventuais factores de risco, podem ser incorporadas covariáveis no modelo (4.1) e assim obter um modelo de regressão. Habitualmente

pretende-se estimar a taxa de cura, a função de sobrevivência dos indivíduos susceptíveis e o efeito das covariáveis na taxa de cura, ou na função de sobrevivência ou em ambas. Assim, interessa saber se o acontecimento de interesse pode ocorrer (que se designa por incidência) e quando ocorrerá, dado que pode ocorrer (que se designa por latência).

Deste modo, cada covariável  $z_i$ ,  $i = 1, \dots, k$  estão associados dois parâmetros, digamos  $\beta_i$  e  $\gamma_i$ : um que descreve como a covariável afecta a incidência a longo prazo (ou a taxa de cura) e outro que descreve como afecta a latência (ou o tempo de vida). Uma covariável que seja importante para a incidência pode não ser importante para a latência e vice-versa. Note-se que alguns  $\beta_i$  (ou mesmo todos) podem coincidir com os respectivos  $\gamma_i$  (isto é, o efeito da covariável no tempo de vida é o mesmo do que na taxa de cura) ou alguns  $\beta_i$  ou  $\gamma_i$  podem ser nulos (ou seja, a covariável não influencia o tempo de vida ou a taxa de cura, respectivamente).

O modelo (4.1) com covariáveis escreve-se na forma

$$S(t; \mathbf{z}) = p(\mathbf{z}) + (1 - p(\mathbf{z}))S_d(t; \mathbf{z}) \quad (4.5)$$

ou, alternativamente,

$$S(t; \mathbf{z}) = p(\mathbf{z}) + (1 - p(\mathbf{z}))S(t|Y = 1; \mathbf{z}),$$

onde  $\mathbf{z}$  representa o vector de covariáveis associado a um determinado indivíduo e  $S_d(t; \mathbf{z})$  pode ter uma formulação paramétrica ou semi-paramétrica. Relativamente à taxa de cura, fazendo uso da regressão logística, a variável  $Y$  é caracterizada por

$$P[Y = 1|\mathbf{z}] = 1 - p(\mathbf{z}) = \frac{\exp(\gamma_0 + \mathbf{z}\boldsymbol{\gamma})}{1 + \exp(\gamma_0 + \mathbf{z}\boldsymbol{\gamma})},$$

onde  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)'$  e  $\gamma_0, \gamma_1, \dots, \gamma_k$  são os coeficientes de regressão.

Como referem Kuk e Chen (1992) e Sy e Taylor (2000), um modelo de mistura em que as funções hazard dos indivíduos susceptíveis sejam proporcionais (que estes autores designam por *modelo de cura de hazards proporcionais*), não é de facto um modelo de hazards proporcionais. Na verdade, sejam  $\mathbf{z}$  e  $\mathbf{z}^*$  os vectores de covariáveis de dois indivíduos susceptíveis e

suponhamos então que a razão das respectivas funções hazard não depende de  $t$ , isto é, que

$$\frac{h_d(t; \mathbf{z})}{h_d(t; \mathbf{z}^*)} = \frac{h_{d0}(t)\alpha(\mathbf{z})}{h_{d0}(t)\alpha(\mathbf{z}^*)} = \frac{\alpha(\mathbf{z})}{\alpha(\mathbf{z}^*)},$$

onde  $h_{d0}$  representa a função hazard subjacente dos indivíduos susceptíveis. Admitindo que  $p(\mathbf{z}) = p$ , a função hazard relativa ao modelo (4.5) é representada na forma

$$h(t; \mathbf{z}) = \frac{(1-p)S_d(t; \mathbf{z})h_d(t; \mathbf{z})}{S(t; \mathbf{z})} = \frac{(1-p)S_d(t; \mathbf{z})h_d(t; \mathbf{z})}{p + (1-p)S_d(t; \mathbf{z})}.$$

Então, dados  $\mathbf{z}$  e  $\mathbf{z}^*$ , os vectores de covariáveis associados aos dois indivíduos, temos

$$\frac{h(t; \mathbf{z})}{h(t; \mathbf{z}^*)} = \frac{(1-p)S_d(t; \mathbf{z})h_d(t; \mathbf{z})}{p + (1-p)S_d(t; \mathbf{z})} \left[ \frac{(1-p)S_d(t; \mathbf{z}^*)h_d(t; \mathbf{z}^*)}{p + (1-p)S_d(t; \mathbf{z}^*)} \right]^{-1},$$

o que é equivalente a

$$\frac{h(t; \mathbf{z})}{h(t; \mathbf{z}^*)} = \left[ \frac{p(S_d(t; \mathbf{z}) - S_d(t; \mathbf{z}^*))}{(p + (1-p)S_d(t; \mathbf{z}))S_d(t; \mathbf{z}^*)} + 1 \right] \frac{\alpha(\mathbf{z})}{\alpha(\mathbf{z}^*)}. \quad (4.6)$$

Claramente, (4.6) é dependente do tempo, o que está de acordo com a afirmação de Kuk e Chen (1992) e de Sy e Taylor (2000). Sy e Taylor (2000) referem ainda que, para uma covariável binária, pode mesmo acontecer que as funções hazard populacionais se cruzem. No entanto, não deixa de ser verdade que o modelo de hazards proporcionais usual é um caso particular do modelo de cura de hazards proporcionais, bastando para tal fazer  $p(\mathbf{z}) = 0, \forall \mathbf{z}$ .

### 4.2.2 Modelo de cura de não-mistura

Seja qual for o tipo de modelo de cura, a função de sobrevivência populacional que lhe está associada é uma função imprópria. Este facto, em

conjunto com a relação existente entre uma função de sobrevivência e a correspondente função hazard cumulativa, tem como consequência que a função hazard cumulativa é limitada superiormente. Assim, seja  $S(t)$  a função de sobrevivência de  $T$  e  $H(t)$  é tal que  $S(t) = \exp[-H(t)]$ . Se  $S(\infty) > 0$  então existe  $\theta < \infty$ , tal que  $H(\infty) = \theta$ . De acordo com Yakovlev *et al.* (1993), uma forma possível de caracterizar esta propriedade é considerar  $H(t) = \theta F(t)$ , onde  $F(t)$  designa a função de distribuição (própria) de uma variável aleatória não negativa. Então, o modelo de cura de não-mistura pode ser escrito na forma

$$S(t) = \exp[-\theta F(t)]. \quad (4.7)$$

A correspondente função hazard é

$$h(t) = \theta f(t),$$

onde  $f(t)$  é a função densidade correspondente a  $F(t)$ .

Neste caso, a probabilidade de cura é  $P(\text{Cura}) = e^{-\theta}$ , que corresponde a  $S(\infty)$ . Importa observar que é possível determinar a probabilidade de cura em cada instante, pois ela é dada por

$$P(\text{Cura}|T \geq t) = \exp[-\theta(1 - F(t))].$$

Notemos que, neste contexto, existem duas hipóteses para obter uma função de sobrevivência imprópria. Ou se modifica o espaço de parâmetros de alguma distribuição própria como acontece, por exemplo, com a distribuição de Gompertz modificada, ou se considera uma variável fragilidade para caracterizar a susceptibilidade de cada indivíduo ao acontecimento de interesse. Nesta situação, é a escolha da distribuição da fragilidade e não da distribuição do tempo de vida que vai dar origem a uma função de sobrevivência imprópria. Um exemplo deste caso é a distribuição Qui-quadrado não central com zero graus de liberdade (Rocha, 1995), que se representa por  $\chi_0'^2(\gamma)$ , onde  $\gamma > 0$  designa o parâmetro de não centralidade.

Vejamos então estas duas situações com maior detalhe.

- Quando se opta pela distribuição  $\chi_0'^2(\gamma)$  para a fragilidade, sob o modelo multiplicativo (2.1), a função de sobrevivência (populacional) é

$$S(t) = \exp \left[ -\frac{\gamma}{2} \left( 1 - \frac{1}{1 + 2\Lambda(t)} \right) \right],$$

em que  $\Lambda(t)$  representa a função hazard cumulativa dos indivíduos com fragilidade  $w = 1$ . Constatamos então que estamos perante um modelo da forma (4.7), onde  $\theta = \gamma/2$  e  $F(t) = 1 - [1 + 2\Lambda(t)]^{-1}$ .

Curiosamente, quando  $\Lambda(t) = \lambda t^\alpha$ , ou seja, quando consideramos uma distribuição de Weibull de parâmetros  $\lambda$  e  $\alpha$ , o resultado é a obtenção de uma distribuição log-logística de parâmetros  $2\lambda$  e  $\alpha$  para  $F(t)$ , pois

$$S(t) = \exp \left[ -\frac{\gamma}{2} \left( 1 - \frac{1}{1 + 2\lambda t^\alpha} \right) \right].$$

Claro que o caso em que  $\alpha = 1$  corresponde à escolha da distribuição exponencial para  $\Lambda(t)$ , resultando na distribuição log-logística uniparamétrica para  $F(t)$ .

- Quando se considera a distribuição de Gompertz modificada (Cantor e Shuster, 1992)

$$S(t) = \exp [\lambda^{-1}\alpha(1 - \exp(\lambda t))], \quad \alpha > 0, \lambda < 0,$$

de acordo com o modelo (4.7), temos  $\theta = -\lambda^{-1}\alpha$  e  $F(t) = 1 - \exp(\lambda t)$ , ou seja,  $F(t)$  corresponde à distribuição exponencial de parâmetro  $-\lambda$ .

Tsodikov (1998) comentou que, quando se admite a existência de indivíduos curados, optarmos por um modelo com fragilidade para explicar a heterogeneidade não observada ou considerarmos a população como homogênea embora com função de sobrevivência imprópria, é mais uma questão de filosofia do que de estatística.

Uma extensão possível do modelo (4.7), consiste em incluir covariáveis. Uma possibilidade é construir modelos de regressão da forma  $S(t; \mathbf{z}) = \exp[-\theta F(t; \mathbf{z})]$ . Podemos então considerar

1.  $\bar{F}(t; \mathbf{z}) = \bar{F}_0(t)^{\exp(\mathbf{z}\boldsymbol{\beta})}$
2.  $\bar{F}(t; \mathbf{z}) = \bar{F}_0(t \exp(\mathbf{z}\boldsymbol{\beta}))$
3.  $\frac{\bar{F}(t; \mathbf{z})}{1 - \bar{F}(t; \mathbf{z})} = \exp(-\mathbf{z}\boldsymbol{\beta}) \frac{\bar{F}_0(t)}{1 - \bar{F}_0(t)}$

onde  $\bar{F} = 1 - F$ ,  $\bar{F}_0(t)$  é a função de sobrevivência subjacente,  $\mathbf{z}$  é o vector de covariáveis e  $\boldsymbol{\beta}$  o correspondente vector de parâmetros de regressão. Obtemos

pois modelos de cura de não-mistura, cujas funções próprias seguem um modelo de regressão de hazards proporcionais, de tempo de vida acelerado ou de possibilidades proporcionais, respectivamente. A distribuição de Weibull pode ser escolhida para  $\overline{F}_0(t)$  nos casos 1 e 2 e a distribuição log-logística nos casos 2 e 3.

Note-se que no caso 2, a estrutura de modelo de tempo de vida acelerado é mantida por  $S(t; \mathbf{z})$ , ou seja,  $S(t; \mathbf{z}) = S_0(t \exp(\mathbf{z}\boldsymbol{\beta}))$ , mas o mesmo não acontece com os restantes dois casos.

Como já referimos no capítulo anterior, Tsodikov (1998) notou que ao reescrever o modelo (4.7) na forma  $S(t) = [\exp[-F(t)]]^\theta$ , se  $\theta$  dependesse de covariáveis observadas, teria uma generalização do modelo de hazards proporcionais. Ao contrário do que sucede com o modelo de cura de mistura, o modelo assim obtido é um verdadeiro modelo de hazards proporcionais e, além disso, permite contemplar a existência de indivíduos imunes. Tsodikov *et al.* (2003) designaram este modelo por *modelo de hazards proporcionais impróprio*.

O modelo de cura de não-mistura apresenta algumas vantagens relativamente ao modelo de cura de mistura, nomeadamente:

1. Permite a construção de uma vasta classe de modelos de regressão semi-paramétricos através de transformações não lineares (Tsodikov 2002, 2003), na qual se inclui como caso particular o modelo de hazards proporcionais impróprio.
2. Nalgumas situações, proporciona uma interpretação biológica mais natural. Esta característica torna-se especialmente importante quando simultaneamente se utiliza outras abordagens matemáticas ao problema em estudo como, por exemplo, na optimização da terapia oncológica.
3. Apresenta algumas vantagens técnicas no processo de maximização da função de verosimilhança e na estimação Bayesiana.

### Distribuição de Gompertz/Weibull modificada

Notando que, em (4.7),  $F(t)$  pode ser qualquer função de distribuição própria de uma variável aleatória não negativa, propomos agora generalizar um pouco mais esse modelo, considerando a distribuição de Weibull para  $F(t)$ . Obtemos assim uma nova distribuição imprópria, que é definida para  $t > 0$ , por

$$S(t) = \exp[\lambda^{-1}\alpha(1 - \exp(\lambda t^\gamma))], \quad \alpha, \gamma > 0, \lambda < 0, \quad (4.8)$$

distribuição esta que é caracterizada pelos parâmetros,  $\alpha$ ,  $\lambda$  e  $\gamma$ , em que  $\gamma$  é o parâmetro de forma. Designaremos por distribuição Gompertz/Weibull modificada a distribuição definida por (4.8). A função hazard correspondente é

$$h(t) = \alpha\gamma t^{\gamma-1} \exp(\lambda t^\gamma), \quad t > 0.$$

Como

$$h'(t) = \alpha\gamma t^{\gamma-2} \exp(\lambda t^\gamma) [\gamma - 1 + \lambda\gamma t^{\gamma-1}],$$

$h(t)$  é unimodal para  $\gamma > 1$ , sendo crescente para  $0 < t < \left(\frac{1-\gamma}{\lambda\gamma}\right)^{1/\gamma}$  e decrescente para  $t \geq \left(\frac{1-\gamma}{\lambda\gamma}\right)^{1/\gamma}$ . Para  $\gamma \leq 1$ ,  $h(t)$  é monótona decrescente. A Fig. 4.1 ilustra o que acabamos de referir. No anexo C.1 estão representados mais alguns gráficos da função hazard, de modo a ilustrar o seu comportamento em função da variação do valor dos parâmetros.

Comparando esta distribuição com a distribuição de Gompertz modificada, notamos que em ambos os casos  $S(\infty) = \exp(\alpha/\lambda)$ . No entanto, a nossa proposta tem a vantagem de permitir uma maior flexibilidade da função hazard pois, além de poder ser decrescente, também pode ser unimodal. Nesta última situação, a partir de certo valor de  $t$ , a função é praticamente constante, num valor próximo de zero. Julgamos que este tipo de função hazard é perfeitamente consentâneo com situações em que existem indivíduos curados.

Importa notar que, para  $\gamma > 1$ , a função hazard deste modelo é do mesmo tipo que a do modelo proposto por Shao e Zhou (2004). Estes autores admitem um modelo de cura de mistura, em que a função de sobrevivência dos indivíduos susceptíveis é a distribuição de Burr de tipo *XII*.

Deixamos para trabalho futuro a questão da estimação dos parâmetros deste modelo.

Note-se que se  $\lambda > 0$  em (4.8), temos também uma nova distribuição, que poderemos designar por distribuição de Gompertz/Weibull, mas que é uma distribuição própria. Como tal, não a iremos referir com mais detalhe, visto não ser adequada ao problema que estamos a tratar.

**Figura 4.1:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\lambda=-0.5$  e  $\alpha=2$ .

### 4.2.3 Relação entre os modelos

Embora os modelos de cura sejam caracterizados como modelos de cura de mistura ou modelos de cura de não-mistura, a verdade é que com a parametrização adequada, eles dão origem ao mesmo modelo. No entanto, a distinção não é meramente formal, pois a interpretação dos parâmetros é diferente consoante se trate de um ou de outro tipo de modelo. Esta é uma das justificações para esta diferenciação. Assim, consoante a natureza do estudo, pode ter interesse optar por um ou outro tipo.

Um outro motivo, diz respeito à proporcionalidade das funções hazard. Assim, se dada uma amostra, as funções hazard dos indivíduos susceptíveis forem proporcionais, a primeira opção deve ser ajustar um modelo de cura de mistura aos dados. Contudo, se dados dois quaisquer indivíduos, susceptíveis ou não, as correspondentes funções hazard forem proporcionais,

então a primeira opção já deverá ser ajustar um modelo de cura de não-mistura.

Vejamos então como relacionar os modelos.

Partindo de (4.7), uma vez que  $p = e^{-\theta}$ , obtemos (4.1), em que

$$S_d(t) = \frac{e^{-\theta F(t)} - e^{-\theta}}{1 - e^{-\theta}}.$$

Reciprocamente, temos

$$\theta = -\log p \quad \text{e} \quad F(t) = \frac{\log[p + (1 - p)S_d(t)]}{\log p}.$$

Vários autores chamaram a atenção para esta relação, entre os quais, Chen *et al.* (1999), Tsodikov (2002), Tsodikov *et al.* (2003). Tsodikov *et al.* (2003) referiu ainda que se for seguida uma abordagem não paramétrica, a forma de estimar a probabilidade de cura é a mesma, mas o mesmo já não acontece com a abordagem paramétrica. Na verdade, basta observar que se for especificada uma distribuição para os indivíduos susceptíveis, a função de sobrevivência (4.7) pode ser representada na forma (4.1) mas tanto  $p$  como  $S_d(t)$  são escritas à custa de um parâmetro comum,  $\theta$ . Por outro lado, Sposto (2002) apresenta alguns exemplos de aplicação onde são usados, ora os modelos de mistura, ora os modelos de não-mistura e refere vantagens e desvantagens na aplicação destes modelos.

### 4.3 Estimação dos parâmetros

Devido à relação existente entre o modelo de cura de mistura e o modelo de cura de não-mistura, vamos detalhar a estimação dos parâmetros apenas para o modelo de mistura. Admitamos o modelo (4.1)

$$S(t) = p + (1 - p)S_d(t)$$

e recordemos que  $S_d(t) = S(t|Y = 1)$ . Numa abordagem paramétrica admitamos, em particular, uma distribuição contínua para o tempo de vida dos indivíduos susceptíveis. Vejamos então a questão da estimação dos parâmetros, seguindo no essencial a abordagem de Steele (2003), embora a autora se referisse a tempos de vida discretos.

Consideremos uma amostra de dimensão  $n$  e designemos por  $t_1, \dots, t_n$  os tempos de vida. Sem perda de generalidade, suponhamos que os primeiros  $m$  ( $m < n$ ) tempos de vida são censurados. Sejam  $\delta_1, \dots, \delta_n$  tais que

$$\delta_i = \begin{cases} 0 & \text{se } 1 \leq i \leq m \\ 1 & \text{se } m + 1 \leq i \leq n \end{cases}$$

e  $y_1, \dots, y_n$  tais que

$$y_i = \begin{cases} 0 & \text{se o indivíduo é imune} \\ 1 & \text{se o indivíduo é susceptível} \end{cases}$$

Começemos por analisar quais as situações possíveis para o par  $(\delta_i, y_i)$ , a que casos correspondem e que reflexos têm na função de verosimilhança. Na verdade, podemos sintetizar tudo isto na tabela 4.1.

Valores possíveis para $(\delta_i, y_i)$	Descrição da situação	Contribuição para a verosimilhança
(0, 0)	Censurado, imune	$p$
(0, 1)	Censurado, susceptível	$(1 - p)S_d(t_i)$
(1, 1)	Observado, susceptível	$(1 - p)h_d(t_i)S_d(t_i)$

**Tabela 4.1:** Resumo das possibilidades para  $(\delta_i, y_i)$ .

Quando temos observações censuradas, não nos é possível, regra geral, identificar inequivocamente quais são os indivíduos imunes. O facto de não observarmos todos os  $y_i$ 's (aqueles correspondentes aos tempos de vida censurados), conduz-nos a uma situação de dados incompletos. Ora neste caso específico, o método de eleição é o algoritmo EM, visto ser um método iterativo que permite obter as estimativas de máxima verosimilhança dos parâmetros em situações em que existem observações omissas.

Uma vez que admitimos que a censura é não informativa, a função de verosimilhança para uma amostra de dimensão  $n$  é dada por

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}.$$

No caso do modelo (4.1), a verosimilhança total observada é

$$L_O = \prod_{i=1}^n [(1-p)f_d(t_i)]^{\delta_i} [p + (1-p)S_d(t_i)]^{1-\delta_i}. \quad (4.9)$$

Se todos os  $y_i$ 's fossem observados, de acordo com a tabela 4.1, teríamos a seguinte verosimilhança completa

$$L_C = \prod_{i=1}^n [[(1-p)f_d(t_i)]^{y_i}]^{\delta_i} [p^{1-y_i} [(1-p)S_d(t_i)]^{y_i}]^{1-\delta_i}. \quad (4.10)$$

Tendo em conta que  $f_d(t_i) = h_d(t_i)S_d(t_i)$  e que, dados os valores possíveis para  $(\delta_i, y_i)$ ,  $(1-\delta_i)(1-y_i) = 1-y_i$ , a verosimilhança anterior pode ser factorizada em

$$L_C = \prod_{i=1}^n (1-p)^{y_i} p^{1-y_i} \prod_{i=1}^n h_d(t_i)^{y_i \delta_i} S_d(t_i)^{y_i} = L_{C_1} L_{C_2}. \quad (4.11)$$

Notando que  $q = 1-p$ , o logaritmo da função de verosimilhança completa é

$$\log L_C = \sum_{i=1}^n [y_i \log q + (1-y_i) \log(1-q)] + \sum_{i=1}^n [\delta_i y_i \log h_d(t_i) + y_i \log S_d(t_i)] \quad (4.12)$$

A etapa E do algoritmo EM consiste em determinar o valor esperado do logaritmo da verosimilhança completa em relação à distribuição dos  $Y$ 's não observados, condicional aos valores actuais dos parâmetros e aos dados observados  $\mathcal{O}$ , onde  $\mathcal{O} = \{y_i \text{ observado}, (t_i, \delta_i), i = 1, \dots, n\}$ . No entanto, como  $\log L_C$  é linear em  $Y$ , para calcular o valor esperado de  $\log L_C$  basta substituir, em (4.12), os valores não observados de  $Y$  pelos respectivos valores esperados, denotados por  $\tau_i$ . Então, temos:

$$\tau_i = E(Y|\mathcal{O}) = P(Y_i = 1|T_i > t_i, \delta_i = 0) = \frac{(1-p)S_d(t_i)}{S(t_i)}.$$

Assim, no logaritmo da verosimilhança completa, cada  $y_i$  é substituído por  $\omega_i$ , onde  $\omega_i$  é definido na forma que se segue:

$$\omega_i = \begin{cases} 1 & \text{se } \delta_i = 1 \\ \tau_i & \text{se } \delta_i = 0 \end{cases}$$

Na etapa E, cada observação censurada  $i$  é atribuída à subpopulação  $Y = 1$  com probabilidade  $\tau_i$  e à subpopulação  $Y = 0$  com probabilidade  $1 - \tau_i$ .

Para efectuar a etapa M, precisamos de maximizar as duas componentes do valor esperado do logaritmo da função de verosimilhança

$$\begin{aligned} \log L_{E_1} &= \sum_{i=1}^n [\omega_i \log q + (1 - \omega_i) \log(1 - q)] \\ &= (n - m) \log q + m \log(1 - q) + \sum_{i=1}^m \tau_i [\log q - \log(1 - q)], \end{aligned} \tag{4.13}$$

$$\begin{aligned} \log L_{E_2} &= \sum_{i=1}^n [\delta_i \omega_i \log h_d(t_i) + \omega_i \log S_d(t_i)] \\ &= \sum_{i=1}^m \tau_i \log S_d(t_i) + \sum_{i=m+1}^n [\log h_d(t_i) + \log S_d(t_i)], \end{aligned} \tag{4.14}$$

Relativamente a  $\log L_{E_1}$  a expressão do estimador de  $q$  na  $(k + 1)$ -ésima iteração é

$$q^{(k+1)} = \frac{1}{n} \left[ (n - m) + \sum_{i=1}^m \tau_i^{(k)} \right], \tag{4.15}$$

e em relação a  $\log L_{E_2}$ , naturalmente, vai depender da distribuição escolhida para o tempo de vida dos indivíduos susceptíveis.

Em termos gerais, a estratégia a adoptar para a implementação do algoritmo EM, resume-se a:

1. determinar os valores iniciais;
2. calcular  $\omega_i$  com base nos valores actuais dos parâmetros;
3. maximizar  $\log L_{E_1}$  e  $\log L_{E_2}$  com base nos valores actuais de  $\omega_i$  e dos parâmetros;
4. repetir 2. e 3. até à convergência.

## 4.4 Modelo de cura baseado na distribuição log-logística

### 4.4.1 Motivação

Conforme refere Yamaguchi (1992), o modelo de hazards proporcionais não permite uma distinção paramétrica útil entre o efeito das covariáveis no tempo até à ocorrência do acontecimento de interesse e o efeito das covariáveis na probabilidade de ocorrência do acontecimento. Na verdade, embora quer o modelo de hazards proporcionais quer o modelo de tempo de vida acelerado possam ser generalizados de modo a permitir a inclusão de uma proporção de indivíduos imunes, essa mesma proporção só é independente dos restantes parâmetros do modelo neste último caso. De facto, considerando  $S(t, \theta)$  a função de sobrevivência do tempo de vida com  $\theta = \alpha(\mathbf{z})$ , no modelo de tempo de vida acelerado, em que  $S(t, \theta) = S_0(\theta t)$ ,  $S(\infty, \theta)$  é independente do parâmetro  $\theta$ , quer  $S_0(\infty) = 0$  ou  $S_0(\infty) > 0$  (distribuição imprópria). Já no caso do modelo de hazards proporcionais,  $S(t, \theta) = S_0(t)^\theta$ ,  $S(\infty, \theta)$  só é independente de  $\theta$  se  $S_0(\infty) = 0$ .

Nesta perspectiva, será útil considerar distribuições que levem à obtenção de um modelo de tempo de vida acelerado, tais como a distribuição de Weibull, a distribuição log-normal, a distribuição log-logística e de forma mais geral, a distribuição gama generalizada. Em situações para as quais é adequada a utilização de uma função hazard unimodal, a distribuição log-

-logística apresenta, relativamente à log-normal, a vantagem de ter uma expressão analítica simples para a função de sobrevivência e para a função hazard.

Por outro lado, uma característica única do modelo de regressão baseado na distribuição log-logística, é o facto de pertencer simultaneamente à família de modelos de tempo de vida acelerado e à família de modelos com possibilidades proporcionais. Este facto permite que os parâmetros do modelo possam ser obtidos através do modelo de tempo de vida acelerado e vice-versa. Além disso, a interpretação do resultado da análise tanto pode ser feita em termos do factor de aceleração como em termos das possibilidades de sobreviver para além de determinado instante, aquela que for mais adequada aos dados em estudo. Como este último tipo de modelos ainda foi pouco utilizado no contexto dos modelos de cura, este será mais um argumento a favor da distribuição log-logística.

#### 4.4.2 Desenvolvimento

Admitamos que o tempo de vida dos indivíduos susceptíveis segue uma distribuição log-logística. Então as correspondentes funções densidade, de sobrevivência e hazard são, respectivamente:

$$f_d(t) = \frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2}, \quad S_d(t) = \frac{1}{1 + \lambda t^\alpha}, \quad h_d(t) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha}, \quad t \geq 0, \quad \alpha, \lambda > 0. \quad (4.16)$$

A função hazard da distribuição log-logística é monótona decrescente para  $\alpha \leq 1$ ; para  $\alpha > 1$ , é crescente até atingir um máximo no instante  $(\frac{\alpha-1}{\lambda})^{\frac{1}{\alpha}}$  e depois é decrescente, tendendo para zero à medida que  $t \rightarrow \infty$ . Como o percentil  $p$  é determinado através da expressão  $t_p = \left[ \frac{p}{\lambda(1-p)} \right]^{\frac{1}{\alpha}}$ , podemos notar que no caso em que  $\alpha = 2$  a mediana coincide com o instante em que a função hazard atinge o seu máximo.

Se em (4.1) substituirmos  $S_d(t)$  pela sua expressão em (4.16), temos:

$$S(t) = \frac{1 + p \lambda t^\alpha}{1 + \lambda t^\alpha}. \quad (4.17)$$

Dada a relação entre a função de sobrevivência e a função hazard, a partir de (4.17) obtemos:

$$h(t) = \frac{(1-p)\alpha\lambda t^{\alpha-1}}{(1+p\lambda t^\alpha)(1+\lambda t^\alpha)}. \quad (4.18)$$

Podemos observar que a função hazard resultante é o produto da função hazard dos indivíduos susceptíveis por uma outra função de  $t$  que é decrescente. Assim, para o caso em que  $\alpha \leq 1$ , a função hazard populacional decresce mais rapidamente do que a função hazard dos susceptíveis, uma vez que é o resultado do produto de duas funções decrescentes.

Como estamos a admitir censura não informativa, a função de verosimilhança para uma amostra de dimensão  $n$  pode ser expressa por

$$L = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i), \quad (4.19)$$

onde  $\delta_i$  é uma variável indicatriz tal que  $\delta_i = 1$  se  $t_i$  corresponde a um tempo de vida e  $\delta_i = 0$  se  $t_i$  é uma observação censurada. Então, por (4.17) e (4.18), (4.19) tem a seguinte expressão:

$$L(t_1, \dots, t_n; \alpha, \lambda, p) = \prod_{i=1}^n \left[ \frac{(1-p)\alpha\lambda t_i^{\alpha-1}}{(1+\lambda t_i^\alpha)(1+p\lambda t_i^\alpha)} \right]^{\delta_i} \frac{1+p\lambda t_i^\alpha}{1+\lambda t_i^\alpha}. \quad (4.20)$$

A formulação (4.2) do modelo de cura fornece-nos outra via para determinar (4.20). A contribuição para a função de verosimilhança de um indivíduo para o qual foi observado o acontecimento de interesse em  $t_i$  é

$$(1-p)f(t_i|Y=1), \quad (4.21)$$

onde  $f(t_i|Y=1)$  é a função densidade de  $T$  condicional ao indivíduo falhar. Um indivíduo para o qual não foi observado esse mesmo acontecimento até  $t_i$ , contribui para a função de verosimilhança com o factor

$$p + (1-p)S(t_i|Y=1), \quad (4.22)$$

o qual representa a probabilidade de um indivíduo ou nunca falhar, ou falhar apenas depois do instante  $t_i$ . Assim, como a função de verosimilhança, para uma amostra de dimensão  $n$ , também pode ser expressa por

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \quad (4.23)$$

por (4.21) e (4.22), (4.23) escreve-se na forma

$$L = \prod_{i=1}^n [(1-p)f(t_i|Y=1)]^{\delta_i} [p+(1-p)S(t_i|Y=1)]^{1-\delta_i},$$

ou seja, usando a distribuição log-logística,

$$L = \prod_{i=1}^n \left[ \frac{(1-p)\alpha\lambda t_i^{\alpha-1}}{(1+\lambda t_i^\alpha)^2} \right]^{\delta_i} \left[ p + (1-p)\frac{1}{1+\lambda t_i^\alpha} \right]^{1-\delta_i}. \quad (4.24)$$

Escrevendo (4.24) de outro modo,

$$L = \prod_{i=1}^n \left[ \frac{(1-p)\alpha\lambda t_i^{\alpha-1}}{(1+\lambda t_i^\alpha)(1+p\lambda t_i^\alpha)} \right]^{\delta_i} \left[ \frac{1+p\lambda t_i^\alpha}{1+\lambda t_i^\alpha} \right]^{\delta_i} \left[ \frac{1+p\lambda t_i^\alpha}{1+\lambda t_i^\alpha} \right]^{1-\delta_i},$$

conclui-se facilmente que obtemos de novo (4.20).

Temos que:

$$\frac{\partial \log L}{\partial p} = \sum_{i=1}^n \frac{(1-p)\lambda t_i^\alpha - \delta_i(1+\lambda t_i^\alpha)}{(1-p)(1+p\lambda t_i^\alpha)},$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^n \frac{\delta_i - \lambda t_i^\alpha(1-p + \delta_i p \lambda t_i^\alpha)}{\lambda(1+\lambda t_i^\alpha)(1+p\lambda t_i^\alpha)},$$

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^n \left[ \frac{\delta_i}{\alpha} + \delta_i \log t_i - \frac{(\delta_i + 1)\lambda t_i^\alpha \log t_i}{1 + \lambda t_i^\alpha} + \frac{(1 - \delta_i)p\lambda t_i^\alpha \log t_i}{1 + p\lambda t_i^\alpha} \right].$$

Como não é possível obter expressões explícitas para os estimadores de máxima verosimilhança dos parâmetros, as estimativas dos parâmetros têm

que ser determinados por métodos numéricos. Uma possibilidade, é a utilização do método de N-R, por isso, apresentamos ainda as derivadas parciais de segunda ordem.

$$\frac{\partial^2 \log L}{\partial p^2} = - \sum_{i=1}^n \frac{\delta_i(1 + \lambda t_i^\alpha)}{(1-p)^2(1 + p\lambda t_i^\alpha)} - \sum_{i=1}^n \frac{\lambda^2 t_i^{2\alpha}}{(1 + p\lambda t_i^\alpha)^2} + \sum_{i=1}^n \frac{\lambda \delta_i t_i^\alpha (1 + \lambda t_i^\alpha)}{(1-p)(1 + p\lambda t_i^\alpha)^2}$$

$$\frac{\partial^2 \log L}{\partial p \partial \lambda} = \sum_{i=1}^n \frac{(1 - \delta_i \lambda t_i^\alpha) t_i^\alpha}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)} - \sum_{i=1}^n \frac{[\delta_i - \lambda t_i^\alpha (1 - p + \delta_i p \lambda t_i^\alpha)] t_i^\alpha}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)^2}$$

$$\frac{\partial^2 \log L}{\partial p \partial \alpha} = \lambda \sum_{i=1}^n \frac{(1 - \delta_i) t_i^\alpha \log t_i}{1 + p\lambda t_i^\alpha} - p \lambda^2 \sum_{i=1}^n \frac{(1 - \delta_i) t_i^{2\alpha} \log t_i}{(1 + p\lambda t_i^\alpha)^2}$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \lambda^2} &= - \frac{1}{\lambda^2} \sum_{i=1}^n \frac{\delta_i}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)} - \frac{1}{\lambda} \sum_{i=1}^n \frac{\delta_i t_i^\alpha}{(1 + \lambda t_i^\alpha)^2 (1 + p\lambda t_i^\alpha)} \\ &\quad - \frac{p}{\lambda} \sum_{i=1}^n \frac{\delta_i t_i^\alpha}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)^2} - p \sum_{i=1}^n \frac{\delta_i t_i^{2\alpha}}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)} \\ &\quad + \sum_{i=1}^n \frac{t_i^{2\alpha} (1 - p + \delta_i p \lambda t_i^\alpha)}{(1 + \lambda t_i^\alpha)^2 (1 + p\lambda t_i^\alpha)} + p \sum_{i=1}^n \frac{t_i^{2\alpha} (1 - p + \delta_i p \lambda t_i^\alpha)}{(1 + \lambda t_i^\alpha)(1 + p\lambda t_i^\alpha)^2} \end{aligned}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = - \frac{1}{\alpha^2} \sum_{i=1}^n \delta_i - \lambda \sum_{i=1}^n \frac{(\delta_i + 1) t_i^\alpha (\log t_i)^2}{1 + \lambda t_i^\alpha} + \lambda^2 \sum_{i=1}^n (\delta_i + 1) \left( \frac{t_i^\alpha \log t_i}{1 + \lambda t_i^\alpha} \right)^2$$

$$+ p \lambda \sum_{i=1}^n \frac{(1 - \delta_i) t_i^\alpha (\log t_i)^2}{1 + p\lambda t_i^\alpha} - (p\lambda)^2 \sum_{i=1}^n (1 - \delta_i) \left( \frac{t_i^\alpha \log t_i}{1 + p\lambda t_i^\alpha} \right)^2$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial \alpha} = - \sum_{i=1}^n \frac{(\delta_i + 1) t_i^\alpha \log t_i}{1 + \lambda t_i^\alpha} + \lambda \sum_{i=1}^n \frac{(\delta_i + 1) t_i^{2\alpha} \log t_i}{(1 + \lambda t_i^\alpha)^2}$$

$$+ p \sum_{i=1}^n \frac{(1 - \delta_i) t_i^\alpha \log t_i}{1 + p\lambda t_i^\alpha} - p^2 \sum_{i=1}^n \frac{(1 - \delta_i) t_i^{2\alpha} \log t_i}{(1 + p\lambda t_i^\alpha)^2}$$

Conforme referimos anteriormente, uma possibilidade interessante consiste em incorporar covariáveis no modelo de cura e assim obter um modelo de regressão. Assim, se no modelo (4.5) a distribuição para o tempo de vida dos indivíduos susceptíveis for a log-logística, a função de sobrevivência de  $T$  (populacional) é

$$S(t; \mathbf{z}) = p(\mathbf{z}) + (1 - p(\mathbf{z})) \frac{1}{1 + \lambda \exp(\mathbf{z}\boldsymbol{\beta})t^\alpha},$$

onde  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$  é o vector de coeficientes de regressão.

A situação mais simples ocorre quando  $z$  é uma variável binária e pode corresponder ao estudo de um novo tratamento ( $z = 1$ ) relativamente ao tratamento padrão ( $z = 0$ ). Neste caso, a probabilidade de cura de um indivíduo sujeito ao tratamento padrão é dada por

$$P[Y = 0|z = 0] = \frac{1}{1 + \exp(\gamma_0)},$$

enquanto que para um indivíduo sujeito ao novo tratamento é

$$P[Y = 0|z = 1] = \frac{1}{1 + \exp(\gamma_0 + \gamma_1)}.$$

A eficácia do tratamento será aferida pelo aumento da probabilidade de cura e isso corresponde a termos  $\gamma_1 < 0$ . Por outro lado, se  $\gamma_0 = 0$ ,  $P[Y = 0|z = 0] = P[Y = 1|z = 0] = \frac{1}{2}$ , ou seja, nos indivíduos sujeitos ao tratamento padrão, a probabilidade de ficar curado é igual à probabilidade de não ficar. Esta situação faz mais sentido quando, em vez de termos um grupo de indivíduos sujeitos a um tratamento padrão, esse mesmo grupo não é sujeito a qualquer tratamento, recebendo apenas um placebo. Outra possibilidade é a ausência de conhecimento prévio do efeito do tratamento na cura do indivíduo.

### 4.4.3 Aplicação do algoritmo EM

Nesta secção, vamos partir da expressão (4.14) e percorrer o caminho necessário até à implementação de um algoritmo (EM ou alguma variante)

para o cálculo das estimativas dos parâmetros  $q (= 1 - p)$ ,  $\lambda$  e  $\alpha$ . Além disso, exemplificaremos com dados referidos em Klein e Moeschberger (1998).

Recordemos que a expressão do estimador de  $q$  é referida em (4.15). No caso da distribuição log-logística, por (4.16), (4.14) escreve-se na forma

$$\begin{aligned} \log L_{E_2} &= \sum_{i=1}^m [-\tau_i \log(1 + \lambda t_i^\alpha)] + \sum_{i=m+1}^n [\log(\alpha \lambda t_i^{(\alpha-1)}) - 2 \log(1 + \lambda t_i^\alpha)] \\ &= (n - m)(\log \alpha + \log \lambda) + \sum_{i=m+1}^n [(\alpha - 1) \log t_i - 2 \log(1 + \lambda t_i^\alpha)] - \\ &\quad \sum_{i=1}^m \tau_i \log(1 + \lambda t_i^\alpha). \end{aligned} \tag{4.25}$$

Como a etapa M do algoritmo EM consiste em maximizar (4.13) e (4.25) em ordem a  $q$ ,  $\lambda$  e  $\alpha$  para  $\tau_i$  fixo (Peng e Dear, 2000), isto significa que ao derivarmos (4.25),  $\tau_i$  é encarado como uma constante. Assim, as derivadas parciais de  $\log L_{E_2}$  em ordem a  $\alpha$  e a  $\lambda$  são, respectivamente,

$$\begin{aligned} \frac{\partial \log L_{E_2}}{\partial \alpha} &= \frac{n - m}{\alpha} + \sum_{i=m+1}^n \left[ \log t_i - 2\lambda \frac{t_i^\alpha \log t_i}{1 + \lambda t_i^\alpha} \right] - \lambda \sum_{i=1}^m \tau_i \frac{t_i^\alpha \log t_i}{1 + \lambda t_i^\alpha}, \\ \frac{\partial \log L_{E_2}}{\partial \lambda} &= \frac{n - m}{\lambda} - 2 \sum_{i=m+1}^n \frac{t_i^\alpha}{1 + \lambda t_i^\alpha} - \sum_{i=1}^m \tau_i \frac{t_i^\alpha}{1 + \lambda t_i^\alpha}. \end{aligned}$$

Ao considerar  $\frac{\partial \log L_{E_2}}{\partial \alpha} = 0$  e  $\frac{\partial \log L_{E_2}}{\partial \lambda} = 0$ , verifica-se que não existem expressões explícitas para os estimadores de máxima verosimilhança de  $\alpha$  e de  $\lambda$ , logo é necessário recorrer a algum método numérico. A escolha recai no método de Newton-Raphson, o que obriga ao cálculo das segundas derivadas parciais. Assim, numa forma simplificada, temos:

$$\frac{\partial^2 \log L_{E_2}}{\partial \alpha^2} = - \left[ \frac{n - m}{\alpha^2} + 2\lambda \sum_{i=m+1}^n \frac{t_i^\alpha (\log t_i)^2}{(1 + \lambda t_i^\alpha)^2} + \lambda \sum_{i=1}^m \tau_i \frac{t_i^\alpha (\log t_i)^2}{(1 + \lambda t_i^\alpha)^2} \right],$$

$$\frac{\partial^2 \log L_{E_2}}{\partial \lambda^2} = - \left[ \frac{n-m}{\lambda^2} - 2 \sum_{i=m+1}^n \frac{t_i^{2\alpha}}{(1 + \lambda t_i^\alpha)^2} - \sum_{i=1}^m \tau_i \frac{t_i^{2\alpha}}{(1 + \lambda t_i^\alpha)^2} \right],$$

$$\frac{\partial^2 \log L_{E_2}}{\partial \lambda \partial \alpha} = \frac{\partial^2 \log L_{E_2}}{\partial \alpha \partial \lambda} = - \left[ 2 \sum_{i=m+1}^n \frac{t_i^\alpha \log t_i}{(1 + \lambda t_i^\alpha)^2} - \sum_{i=1}^m \tau_i \frac{t_i^\alpha \log t_i}{(1 + \lambda t_i^\alpha)^2} \right].$$

Vamos agora ilustrar a metodologia anterior com a aplicação a dados reais. Klein e Moeschberger (1998, p.10) referem um estudo envolvendo 101 pacientes com leucemia mielógena aguda avançada. A todos estes pacientes foi feito transplante da medula óssea, 51 com a sua própria medula (os do grupo designado por *auto*) e 50 com medula de algum parente compatível (os do grupo designado por *allo*). Como o grupo *allo* tem uma apreciável quantidade de observações censuradas correspondentes aos valores mais elevados das observações, apresenta boas perspectivas de conter indivíduos imunes, daí a razão por termos optado por trabalhar com os dados deste grupo. O nivelamento da estimativa de Kaplan-Meier da função de sobrevivência é bem evidente na Fig 4.2. Além disso, também podemos observar que após aproximadamente 20 meses, a curva estabiliza num valor próximo de 0.5.

Antes de iniciar a implementação propriamente dita, surgiram-nos algumas questões pertinentes, nomeadamente:

1. qual a variante do algoritmo EM a utilizar?
2. como obter os valores iniciais para os parâmetros?

Quanto à primeira questão, por razões pedagógicas optámos por experimentar várias variantes. Relativamente à segunda, temos dois tipos de parâmetros (o parâmetro  $q$  e os parâmetros da distribuição), logo dois tipos de questões. Para  $q$ , com base em Maller e Zhou (1992), considerámos a estimativa da proporção de susceptíveis  $q^{(0)} = 1 - \hat{S}(t_{(r)})$ , onde  $t_{(r)}$  representa a maior observação não censurada e  $\hat{S}(t)$  é a estimativa de Kaplan-Meier da função de sobrevivência. Para os parâmetros da distribuição, optámos por fazer um ajustamento apenas com os dados com valores até  $t_{(r)}$  e as estimativas obtidas constituem os valores iniciais. A ideia intuitiva subjacente

**Figura 4.2:** Estimativa de Kaplan-Meier da função de sobrevivência.

é a de que a maioria das observações censuradas com valores superiores a  $t_{(r)}$  irá corresponder a indivíduos imunes, e a maioria das observações censuradas com valores inferiores ou iguais a  $t_{(r)}$  irá corresponder a indivíduos susceptíveis.

Utilizamos a linguagem R (Ihaka e Gentleman, 1996), e os algoritmos que programamos podem ser encontrados no anexo C. Partimos dos valores iniciais  $q^{(0)} = 0.467857523$ ,  $\lambda^{(0)} = 0.132413185$  e  $\alpha^{(0)} = 1.098664946$  e aplicamos 6 processos alternativos para a estimação dos parâmetros:

- o algoritmo EM;
- o algoritmo ECM com  $R = 2$  e o método de N-R unidimensional, isto é, usamos o método de N-R para  $\alpha$  e para  $\lambda$  em separado;

4.4. Modelo de cura baseado na distribuição log-logística

- o algoritmo ECM com  $R = 2$  e o método N-R bidimensional (método de N-R para  $\alpha$  e para  $\lambda$  em conjunto);
- o algoritmo ECM com  $R = 3$ ;
- o algoritmo EM híbrido;
- o algoritmo EM gradiente.

Os resultados a que chegámos estão patentes na tabela 4.2, onde os valores entre parêntesis representam o desvio padrão das estimativas.

Neste caso, os valores finais dos parâmetros são praticamente iguais mas o mesmo já não podemos afirmar relativamente ao número de iterações. Sem dúvida que o algoritmo EM simples é aquele que requiere um maior número de iterações, confirmando a eventual necessidade de recorrer a variantes que acelerem o processo de convergência.

Algoritmos	Número de iterações	$q$	$\lambda$	$\alpha$
EM	52	0.5092644 (0.09087)	0.1780919 (0.07361)	1.1934109 (0.24877)
ECM,R=2,N-R unid.	24	0.5092644	0.1780918	1.1934110
ECM,R=2,N-R bid.	15	0.5092645	0.1780919	1.1934109
ECM,R=3	14	0.5092644	0.1780919	1.1934108
EM híbrido	29	0.5092644	0.1780919	1.1934110
EM gradiente	33	0.5092645	0.1780919	1.1934109

**Tabela 4.2:** Valores das estimativas em função do algoritmo aplicado.

Apesar da forte suspeita da existência de indivíduos curados na nossa amostra, ajustamos também um modelo homogéneo com a distribuição log-logística. Os valores que obtivemos para as estimativas foram  $\hat{\lambda} = 0.114$  e  $\hat{\alpha} = 0.631$ , o último dos quais evidenciando uma função hazard decrescente.

A Fig. 4.3 ilustra os ajustamentos efectuados. O modelo de cura parece fornecer um bom ajustamento aos dados, com a proporção de indivíduos curados a ser captada pela cauda direita da curva. De acordo com este modelo, aproximadamente 49% dos pacientes estão curados após terem feito o transplante.

**Figura 4.3:** Funções de sobrevivência ajustadas.

Para testar formalmente a presença de indivíduos imunes, consideramos a hipótese  $H_0 : q = 1$ . Denote-se por  $\Gamma$  a estatística de teste da razão de verossimilhanças. Por Maller e Zhou (1996), a distribuição assintótica de  $-2\log(\Gamma)$  é uma mistura 50-50 de uma variável aleatória com distribuição  $\chi^2$  com 1 grau de liberdade e um ponto com massa em 0. Como o valor do logaritmo da verossimilhança maximizada do modelo de cura é -193.2301 e o do modelo homogêneo é -94.8328, então  $-2\log(\Gamma) = 196.7946 > 2.71$ . Note-se que 2.71 corresponde ao percentil 95 da distribuição de  $-2\log(\Gamma)$ . Concluimos, pois, que existe uma forte evidência da presença de indivíduos curados, como já suspeitávamos.

Yu *et al.* (2004) salientam que a duração do *follow-up* tem um papel importante na obtenção de melhores estimativas dos parâmetros nos modelos de cura de mistura. De facto, estes autores enfatizam a necessidade de o *follow-up* ser maior do que o valor da mediana do tempo de vida dos in-

divíduos susceptíveis, por forma a que a maioria dos acontecimentos possa ser observada antes do fim do *follow-up*. Os dados que escolhemos estão precisamente nessa situação, já que a mediana é 4.25 meses e a maior observação é 60.625 meses.

## 4.5 Um novo modelo de cura paramétrico

O modelo de cura que propomos é um modelo de mistura, em que a distribuição do tempo de vida dos indivíduos susceptíveis é uma distribuição biparamétrica, proposta por Chen (2000). Esta distribuição foi desenvolvida no contexto usual da Análise de Sobrevivência, isto é, sem admitir a possibilidade de existência de indivíduos curados na população em estudo. Um dos méritos deste modelo advém do facto de a função hazard dos indivíduos susceptíveis ter alguma flexibilidade, no sentido em que, de acordo com o valor de um dos seus parâmetros, tanto pode ser monótona crescente como *bathtub-shaped*. Outra vantagem reside na facilidade do tratamento matemático, como veremos adiante, no âmbito da estimação dos parâmetros.

### 4.5.1 A função de distribuição dos indivíduos susceptíveis

A função de distribuição proposta por Chen (2000) é

$$F(t) = 1 - \exp[\lambda(1 - \exp(t^\beta))], \quad t > 0, \quad (4.26)$$

onde  $\lambda$  e  $\beta$  são os parâmetros da distribuição, sendo  $\lambda$  o parâmetro de escala e  $\beta$  o parâmetro de forma. Assim, as correspondentes função de sobrevivência e função hazard são, respectivamente,

$$\bar{F}(t) = \exp[\lambda(1 - \exp(t^\beta))], \quad t > 0 \quad (4.27)$$

e

$$h^*(t) = \lambda\beta t^{\beta-1} \exp(t^\beta), \quad t > 0. \quad (4.28)$$

O autor refere que  $h^*(t)$  pode ser *bathtub-shaped* quando  $\beta < 1$  e que é crescente quando  $\beta \geq 1$ , o que não é habitual nas distribuições mais usadas em análise de sobrevivência. Na verdade, como

$$h^{*'}(t) = \lambda \beta t^{\beta-2} \exp(t^\beta) ((\beta - 1) + \beta t^\beta),$$

para  $\beta < 1$ , temos que  $h^*(t)$  é decrescente para  $t \in [0, (\frac{1}{\beta} - 1)^{\frac{1}{\beta}}[$  e, para  $t \geq (\frac{1}{\beta} - 1)^{\frac{1}{\beta}}$ ,  $h^*(t)$  é crescente. Assim, o intervalo para  $t$  onde  $h^*(t)$  é decrescente será tanto maior quanto menor o valor de  $\beta$ , como se pode verificar na Fig.4.4. Deste modo, para valores de  $\beta$  próximos de zero, por exemplo  $\beta = 0.1$ , o intervalo tem uma amplitude tão grande que, do ponto de vista prático, é como se tivéssemos uma função hazard decrescente. Já para valores de  $\beta$  próximos de 1, o intervalo em que a função hazard é decrescente é tão pequeno que é quase como se a função hazard fosse sempre crescente.

**Figura 4.4:** Funções hazard da distribuição de Chen, com  $\lambda = 2$  e  $\beta$  a variar entre 0.4 e 1.2.

## 4.5.2 O novo modelo de cura

Recordemos que o modelo de cura de mistura pode ser escrito em termos da função de sobrevivência na forma

$$S(t) = 1 - q + qS(t|Y = 1),$$

onde  $q$  representa a proporção de indivíduos susceptíveis,  $S(t|Y = 1)$  a função de sobrevivência dos mesmos e  $Y$  a variável que toma o valor um se o indivíduo é susceptível e zero caso contrário.

Uma vez que a função  $S(t|Y = 1)$  será caracterizada por (4.27), o modelo é definido do modo que se segue

$$S(t) = 1 - q + q \exp[\lambda(1 - \exp(t^\beta))], \quad t > 0. \quad (4.29)$$

Se o modelo for escrito à custa da função hazard, temos

$$h(t) = \frac{q\lambda\beta t^{\beta-1} \exp(t^\beta) \exp[\lambda(1 - \exp(t^\beta))]}{1 - q + q \exp[\lambda(1 - \exp(t^\beta))]}. \quad (4.30)$$

**Figura 4.5:** Funções hazard do modelo de cura baseado na distribuição de Chen, com  $q = 0.3$ ,  $\lambda = 2$  e  $\beta$  a variar entre 0.4 e 1.4.

Importa notar que, embora o parâmetro  $\beta$  continue a ser um parâmetro de forma do modelo populacional, o comportamento da função hazard populacional é completamente diferente do da função hazard dos indivíduos susceptíveis. Assim,  $h(t)$  é decrescente para  $\beta \leq 1$  e é unimodal para  $\beta > 1$ , como podemos observar na Fig.4.5. Além disso, no caso em que  $\beta > 1$ , o valor da moda será tanto maior quanto maior o valor de  $q$ , o que faz todo o sentido uma vez que  $q$  representa a proporção de indivíduos susceptíveis. Neste caso,  $q$  é o parâmetro de escala. Relativamente ao parâmetro  $\lambda$ , de certa forma vai definir o valor de  $t$  a partir do qual a função hazard irá estabilizar num valor próximo de zero, ou seja, neste contexto, o valor de  $t$  a partir do qual poderemos esperar que os indivíduos que sobrevivam até esse instante estejam curados. Mais especificamente, quanto mais pequeno for o valor de  $\lambda$ , maior o valor de  $t$  a partir do qual a função hazard é aproximadamente zero. Em anexo, encontram-se alguns gráficos da função hazard para este modelo, elucidativos das características atrás referidas.

É interessante notar que as funções hazard deste modelo e do modelo de Gompertz/Weibull modificado são do mesmo tipo. Tal facto não é de estranhar pois, tendo em conta a relação entre os modelos de cura de mistura e os modelos de cura de não-mistura, as funções de sobrevivência têm algumas semelhanças.

Recordemos ainda que, ao considerar uma amostra de dimensão  $n$ , na presença de censura não informativa, caso todos os  $y_i$ 's fossem observados, a denominada verosimilhança completa escreve-se na forma

$$L_C = \prod_{i=1}^n q^{y_i} (1 - q)^{1-y_i} \prod_{i=1}^n h(t_i|Y = 1)^{y_i \delta_i} S(t_i|Y = 1)^{y_i},$$

onde  $\delta_i$  representa a função indicatriz que tem o valor um se  $t_i$  corresponde a um tempo de vida observado e tem o valor zero, caso contrário.

Assim, para o nosso caso, a função de verosimilhança completa é

$$L_C = \prod_{i=1}^n q^{y_i} (1 - q)^{1-y_i} \prod_{i=1}^n [\lambda \beta t_i^{\beta-1} \exp(t_i^\beta)]^{y_i \delta_i} [\exp[\lambda(1 - \exp(t_i^\beta))]]^{y_i}.$$

### 4.5.3 Estimação dos parâmetros

O procedimento que seguimos é o mesmo do que o usado na secção 4.3 e na subsecção 5.4.2. Deste modo, temos  $m$  tempos de vida censurados e  $n - m$  observados. Além disso,  $\tau_i$  continua a representar o valor esperado de  $Y$  quando  $\delta_i = 0$ , que agora se escreve na forma

$$\tau_i = \frac{q \exp[\lambda(1 - \exp(t_i^\beta))]}{1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]} \quad (4.31)$$

Assim, como  $\omega_i = 1$  se  $\delta_i = 1$  e  $\omega_i = \tau_i$  se  $\delta_i = 0$ , o valor esperado do logaritmo da verosimilhança completa é

$$\begin{aligned} \log L_E = & (n - m) \log q + m \log(1 - q) + (\log q - \log(1 - q)) \sum_{i=1}^m \tau_i + \\ & \lambda \sum_{i=1}^m \tau_i [1 - \exp(t_i^\beta)] + (n - m)(\log \lambda + \log \beta) + \\ & (\beta - 1) \sum_{i=m+1}^n \log t_i + \sum_{i=m+1}^n t_i^\beta + \lambda \sum_{i=m+1}^n [1 - \exp(t_i^\beta)], \end{aligned}$$

tendo em conta que os  $m$  tempos de vida censurados estão indexados de 1 a  $m$  e os  $n - m$  tempos de vida observados estão indexados de  $m + 1$  a  $n$ .

Quanto às derivadas parciais, temos

$$\begin{aligned} \frac{\partial \log L_E}{\partial q} &= \frac{n - m}{q} - \frac{m}{1 - q} + \left(\frac{1}{q} + \frac{1}{1 - q}\right) \sum_{i=1}^m \tau_i, \\ \frac{\partial \log L_E}{\partial \lambda} &= \sum_{i=1}^m \tau_i [1 - \exp(t_i^\beta)] + \frac{n - m}{\lambda} + \sum_{i=m+1}^n [1 - \exp(t_i^\beta)], \\ \frac{\partial \log L_E}{\partial \beta} &= -\lambda \sum_{i=1}^m \tau_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \frac{n - m}{\beta} + \sum_{i=m+1}^n \log t_i + \\ & \quad \sum_{i=m+1}^n (\log t_i) t_i^\beta - \lambda \sum_{i=m+1}^n (\log t_i) t_i^\beta \exp(t_i^\beta). \end{aligned}$$

A partir das igualdades anteriores conseguimos obter expressões explícitas para os estimadores de  $q$  e de  $\lambda$

$$\hat{q} = \frac{1}{n} \left[ (n - m) + \sum_{i=1}^m \tau_i \right], \quad (4.32)$$

$$\hat{\lambda} = \frac{n - m}{\sum_{i=1}^m \tau_i [\exp(t_i^\beta) - 1] + \sum_{i=m+1}^n [\exp(t_i^\beta) - 1]}, \quad (4.33)$$

mas não para o estimador de  $\beta$ . Por conseguinte, de modo a fazermos uso do método de N-R, determinámos ainda a segunda derivada parcial em ordem a  $\beta$ , a qual, após algumas simplificações, escreve-se na forma:

$$\begin{aligned} \frac{\partial^2 \log L_E}{\partial \beta^2} = & -\left[ \lambda \sum_{i=1}^m \tau_i (\log t_i)^2 t_i^\beta \exp(t_i^\beta) (1 + t_i^\beta) + \frac{n-m}{\beta^2} - \sum_{i=m+1}^n (\log t_i)^2 t_i^\beta + \right. \\ & \left. \lambda \sum_{i=m+1}^n (\log t_i)^2 t_i^\beta \exp(t_i^\beta) (1 + t_i^\beta) \right]. \end{aligned}$$

Para proceder à estimação dos parâmetros, ou seja, para usar o algoritmo EM, vamos substituir  $\tau_i$  pelo valor obtido em (4.31). Dada a sua importância, referimos apenas as relativas a (4.32) e (4.33):

$$\hat{q} = \frac{1}{n} \left[ (n - m) + q \sum_{i=1}^m \frac{\exp[\lambda(1 - \exp(t_i^\beta))]}{1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]} \right], \quad (4.34)$$

$$\hat{\lambda} = \frac{n - m}{q \sum_{i=1}^m \frac{\exp[\lambda(1 - \exp(t_i^\beta))]}{1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]} [\exp(t_i^\beta) - 1] + \sum_{i=m+1}^n [\exp(t_i^\beta) - 1]}. \quad (4.35)$$

#### 4.5.4 Aplicações

Uma consequência habitual do desenvolvimento de um novo modelo é a sua aplicação a dados, quer estes sejam reais ou simulados. A nossa opção foi a aplicação a dados reais, com duplo objectivo. Por um lado, naturalmente, a verificação da sua exequibilidade e adequabilidade. Por outro, averiguar se apresenta vantagens em relação a outros modelos previamente estudados. Com este objectivo em mente, ajustámos este modelo a dois conjuntos de dados distintos.

##### Dados de leucemia

De acordo com o trabalho já aqui desenvolvido, fomos usar os dados a que ajustámos anteriormente o modelo de cura com a distribuição log-logística, ou seja, aqueles que podem ser encontrados em Klein e Moeschberger (1998, p.10).

O valor inicial do parâmetro  $q$  é o mesmo do caso anterior, ou seja,  $q^{(0)} = 0.46786$ . Relativamente aos valores iniciais dos restantes dois parâmetros, poderíamos ter seguido o mesmo procedimento do caso anterior, mas não o foi o que fizemos. Na verdade, esse procedimento levaria à programação de um algoritmo para a distribuição de Chen, que embora não trouxesse dificuldades de maior, levou-nos a procurar um caminho mais simples e que se revelou tanto ou mais eficaz. A ideia é extremamente simples e básica: consiste em determinar o valor da estimativa de Kaplan-Meier da função de sobrevivência em dois pontos afastados e resolver em ordem a  $\lambda$  para dois valores fixos de  $\beta$ . O valor de  $\beta$  que leve a valores de  $\lambda$  mais próximos é a estimativa inicial de  $\beta$ . A estimativa inicial de  $\lambda$  é a média dos dois valores próximos (ver anexo C.5). Assim, os restantes valores iniciais que considerámos foram  $\lambda^{(0)} = 0.183395$  e  $\beta^{(0)} = 0.5$ . As estimativas que obtivemos foram  $\hat{q} = 0.47060$  (0.08809),  $\hat{\lambda} = 0.12508$  (0.02965) e  $\hat{\beta} = 0.42100$  (0.03058), onde o valor entre parêntesis indica o desvio padrão correspondente. Note-se a "proximidade" entre os valores iniciais e as estimativas dos parâmetros  $\lambda$  e  $\beta$ , abonando em favor do método utilizado. Ao representarmos a correspondente função de sobrevivência, constatámos que o ajustamento aos dados é quase perfeito, como está patente na Fig.4.6.

Apesar de parecer evidente a superioridade do nosso modelo, comparámos em seguida o valor da estatística  $-2 \log \hat{L}$  para o modelo de cura baseado na distribuição log-logística e para o novo modelo que propomos. De acordo

**Figura 4.6:** Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao novo modelo de cura.

com o esperado, no primeiro caso foi 386.4602 e no segundo apenas 106.2315. De referir ainda o valor ligeiramente inferior do desvio padrão de  $\hat{q}$  quando se utiliza a distribuição de Chen: 0.08809 em vez de 0.09087.

### Outros dados de leucemia

Apesar do quase perfeito ajustamento do modelo de cura com a distribuição de Chen aos dados anteriores, decidimos considerar outros dados para verificar se o modelo era suficientemente flexível para se ajustar bem aos novos dados. Para tal, consideramos os dados de leucemia de Kersey *et al.*, que podem ser encontrados em Maller e Zhou (1996, p.82). Neste caso, na verdade temos 2 grupos de indivíduos, de modo que realizamos o ajustamento para cada um dos grupos em separado.

Seguimos o mesmo procedimento do que com os dados anteriores, quer para obter os valores iniciais, quer no algoritmo utilizado.

Assim, para o grupo 1 os valores iniciais foram  $q^{(0)} = 0.73662$ ,  $\lambda^{(0)} = 0.9930455$  e  $\beta^{(0)} = 1$ , e as estimativas foram  $\hat{q} = 0.7285978$ ,  $\hat{\lambda} = 0.7611151$  e  $\hat{\beta} = 0.6139747$ . Com estes valores das estimativas, obtivemos o gráfico da função de sobrevivência ajustada (Fig. 4.7), a qual, uma vez mais, se ajustou

**Figura 4.7:** Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao modelo de cura com a distribuição de Chen (grupo1).

bastante bem aos dados.

Para o grupo 2, os valores iniciais foram  $q^{(0)} = 0.8134885$ ,  $\lambda^{(0)} = 0.8246753$  e  $\beta^{(0)} = 0.5$ , e as estimativas foram  $\hat{q} = 0.8181942$ ,  $\hat{\lambda} = 1.7420948$  e  $\hat{\beta} = 0.9199423$ . De modo análogo ao grupo 1, representámos a função de sobrevivência ajustada, patente na Fig. 4.8. Podemos verificar que, embora o ajustamento não seja tão bom como no grupo 1, ainda é bastante razoável.

Este mesmo tipo de diferença no ajustamento em relação aos dois grupos ocorreu ao ser considerado um modelo de cura baseado na distribuição de Weibull (Maller e Zhou, 1996, p.108 e p.117) e na distribuição de Burr de tipo *XII* (Shao e Zhou, 2004), o que reforça o facto de o nosso modelo ser uma alternativa a ter em conta.

**Figura 4.8:** Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao modelo de cura com a distribuição de Chen (grupo2).

# Capítulo 5

## Modelos de cura na presença de riscos competitivos

### 5.1 Introdução

A teoria de riscos competitivos é adequada para situações em que cada indivíduo está sujeito a duas ou mais causas de morte. De acordo com Gooley *et al.* (1999), definimos risco competitivo como um acontecimento cuja ocorrência ou impede a observação de outro acontecimento de interesse, ou altera a probabilidade da sua ocorrência.

Historicamente, o objectivo principal da análise de riscos competitivos consiste na estimação das distribuições marginais, ou seja, tenta-se inferir qual a acção de cada risco isoladamente a partir da análise de dados obtidos em situações onde os riscos actuam em conjunto. Frequentemente, o interesse reside num tipo de falha específico, de modo que o objectivo do estudo pode ser:

- a distribuição do tempo de vida associado à falha digamos de tipo 1, na ausência dos restantes tipos de falha;
- a comparação do tempo até à falha de tipo 1 em dois ou mais grupos de indivíduos com características diferentes relativamente aos outros tipos de falhas;
- o efeito da eliminação ou redução da falha de tipo 1, na distribuição marginal do tempo de vida.

Iremos referir três tipos de abordagem aos riscos competitivos, propostos por Moeschberger e David (1971), por Prentice *et al.* (1978) e por Larson e Dinse (1985).

Recentemente foram desenvolvidos modelos de cura que englobam riscos competitivos, tanto para o caso em que a cura é observável, (Betensky e Schoenfeld (2001)), como para quando isso não acontece, (Ng e McLachlan (1998, 2003), Phillips *et al.* (2002), Choi e Zhou (2002) e Maller e Zhou (2002)).

O nosso contributo é a abordagem paramétrica quando a cura é observável e algumas considerações sobre a definição de cura neste contexto. Além disso, iremos ainda propor um modelo de cura no contexto da censura parcialmente informativa.

## 5.2 Modelo de riscos competitivos usual

Uma das primeiras abordagens modernas à teoria dos riscos competitivos foi proposta por Moeschberger e David (1971). Os autores consideram os tempos de vida potenciais (ou latentes),  $T_1, \dots, T_r$ , associados a cada causa de morte de modo a que para cada indivíduo, observa-se apenas  $T = \min(T_1, \dots, T_r)$  e  $l = 1, \dots, r$ , tal que,  $T_l = T$ . Assim,  $P(T_l \geq t)$  representa a probabilidade do tempo de vida de um indivíduo ser igual ou superior a  $t$ , admitindo que ele apenas pode falhar devido à causa  $l$ , ou seja, admitindo que as restantes causas foram eliminadas. Trata-se pois de quantidades não observáveis.

A segunda abordagem, devida a Prentice *et al.* (1978), baseia-se na utilização de funções hazard específicas de cada causa, ou seja, nas também designadas funções sub-hazard. Neste caso, a cada indivíduo está associado o par  $(T, C)$ , onde  $T$  corresponde ao tempo de vida e  $C \in \{1, \dots, r\}$  corresponde à causa de morte. Pode ocorrer censura, no sentido usual, daí que também possa ser associada a variável indicatriz  $\delta$ , que toma o valor 1 se for observado algum tipo de falha e zero caso contrário. Além disso, também se define  $C = 0$  quando o tipo de falha é desconhecido devido à ocorrência de censura.

Um outro modo de encarar os riscos competitivos, proposto por Larson e Dinse (1985), consiste na utilização de modelos de mistura. Com algumas analogias ao caso anterior, tem a vantagem de se prestar mais facilmente ao uso do algoritmo EM para a maximização da função de verosimilhança.

Sem perda de generalidade, em qualquer das situações que iremos abor-

dar, consideramos que cada falha tem origem numa e numa só causa. Por outras palavras, consideramos que quando é observada a morte dum indivíduo, ela é provocada por apenas uma das  $r$  causas.

### 5.2.1 Tempos de vida potenciais

Na análise de riscos competitivos, admite-se que os indivíduos podem morrer de qualquer uma das  $r$  causas a que são vulneráveis. Uma abordagem habitual, consiste em associar à  $l$ -ésima causa de morte a variável  $T_l$ , que representa o tempo de vida associado à causa  $l$ , que seria observado se as restantes causas de morte fossem excluídas. Deste modo, embora o vector  $(T_1, \dots, T_r)$  não seja observável, o tempo de vida de um indivíduo é  $T = \min(T_1, \dots, T_r)$ , pois um indivíduo morre assim que é afectado por qualquer uma das  $r$  causas. Os tempos  $T_1, \dots, T_r$  designam-se por tempos de vida potenciais ou por tempos de vida latentes, no sentido em que apenas um deles poderá ser observado.

A função de sobrevivência conjunta do vector  $\mathbf{T} = (T_1, \dots, T_r)$  é

$$P(\mathbf{T} \geq \mathbf{t}) = P(T_1 \geq t_1, \dots, T_r \geq t_r) = S(t_1, \dots, t_r).$$

A função de distribuição correspondente é  $G(\mathbf{t}) = P(\mathbf{T} < \mathbf{t})$ . A função de sobrevivência marginal de  $T_l$  é  $S_{Gl}(t) = P(T_l \geq t)$ , que habitualmente se representa por  $S_{Gl}(t) = S(0, \dots, t_l, \dots, 0)$ , ou seja, significa que o efeito de eliminar as causas diferentes de  $l$  pode ser representado pela anulação do correspondente argumento. A função hazard marginal respectiva pode ser obtida a partir da relação  $\lambda_l(t) = -d \log S_{Gl}(t)/dt$ , a qual representa a função hazard associada à causa  $l$  na ausência das outras causas de morte. A função de sobrevivência de  $T$  é

$$S(t) = P(T \geq t) = S(t, \dots, t),$$

sendo  $\lambda(t) = -d \log S(t)/dt$ .

A independência dos tempos de vida latentes é definida por

$$S(\mathbf{t}) = \prod_{l=1}^r S_{Gl}(t_l).$$

Quando não existe independência, uma medida para quantificar o grau de dependência entre os tempos de vida latentes é definida por

$$R_S(\mathbf{t}) = \frac{S(\mathbf{t})}{\prod_{l=1}^r S_{Gl}(t_l)}. \quad (5.1)$$

$R_S(\mathbf{t}) < 1$  se e só se  $\forall \mathbf{t}$  existe uma dependência negativa entre as variáveis aleatórias  $T_1, \dots, T_r$  e  $R_S(\mathbf{t}) > 1$  significa que existe uma dependência positiva. Por exemplo, para uma distribuição bivariada,

$$R_S(\mathbf{t}) = \frac{P(T_1 > t_1, T_2 > t_2)}{P(T_1 > t_1)P(T_2 > t_2)} = \frac{P(T_1 > t_1 | T_2 > t_2)}{P(T_1 > t_1)} < 1$$

se e só se

$$P(T_1 > t_1 | T_2 > t_2) < P(T_1 > t_1),$$

ou seja, é menos provável que  $T_1$  exceda  $t_1$  se for conhecido que  $T_2 > t_2$ . Deste modo, como quando os riscos são independentes temos que  $P(\mathbf{T} \geq \mathbf{t}) = S(\mathbf{t}) = \prod_{l=1}^r S_{Gl}(t)$ , então  $R_S(\mathbf{t}) = 1$ .

Conforme já referimos, o tempo de vida  $T_l$  é entendido como o tempo até à falha devido à causa  $l$  que seria observado se as falhas devidas às restantes causas fossem eliminadas. Esta interpretação é algo controversa, na medida em que se pode questionar em que circunstâncias é que é possível, de facto, eliminar as restantes causas de falha. Verifica-se pois que a abordagem dos riscos competitivos através dos tempos de vida potenciais não é isenta de críticas. Tsiatis (1975) provou que um conjunto de funções de sobrevivência específicas de causa não tem correspondência com as respectivas funções de sobrevivência marginais, excepto se as variáveis latentes forem independentes, ou seja, se a função de sobrevivência conjunta for igual ao produto das funções de sobrevivência marginais. Contudo, o problema reside no facto de a hipótese de independência dos tempos de vida não poder ser testada directamente a partir dos dados. Sendo assim, colocam-se problemas de identificabilidade, uma vez que as funções de sobrevivência marginais não são observáveis.

### 5.2.2 Funções específicas de causa

Os dados relativos a cada indivíduo consistem no par  $(T, C)$ , onde  $T$  é uma v.a. não negativa que representa o tempo até à falha do indivíduo e  $C \in \{1, \dots, r\}$  corresponde à causa de morte. Se o período de *follow-up* de um indivíduo terminar antes de ocorrer qualquer falha, a observação será considerada incompleta; neste caso  $C = 0$  indica que o tipo de falha é desconhecido e o tempo de vida é censurado.

Este tipo de abordagem tem por base as funções hazard específicas de cada causa, definidas por

$$h_l(t) = \lim_{\Delta t \rightarrow 0} \frac{P(C = l, t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad l = 1, \dots, r, \quad (5.2)$$

as quais descrevem a probabilidade instantânea de morte devida à causa  $l$  no instante  $t$ , na presença das restantes  $r - 1$  causas.

A partir de (5.2), podemos obter a respectiva função hazard cumulativa

$$H_l(t) = \int_0^t h_l(u) du,$$

bem como a função

$$S_l^*(t) = \exp[-H_l(t)]. \quad (5.3)$$

No entanto, importa notar que, em geral, a função anterior não pode ser interpretada como função de sobrevivência, para  $r > 1$ . Na verdade, quando existe mais do que uma causa de morte, o facto de um indivíduo não falhar devido à causa  $l$  até ao instante  $t$  não significa que permanece livre de falhar de qualquer causa para além do instante  $t$ .

A função hazard global é definida por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad (5.4)$$

mas como se considera que as  $r$  causas são exaustivas e mutuamente exclusivas, tem-se

$$h(t) = \sum_{l=1}^r h_l(t). \quad (5.5)$$

A função hazard cumulativa global é dada por

$$\Lambda(t) = \int_0^t h(u) du = \sum_{l=1}^r H_l(t),$$

logo, quer os riscos sejam independentes ou não, a função de sobrevivência marginal de  $T$  é

$$S(t) = \exp \left[ - \sum_{l=1}^r H_l(t) \right] = \prod_{l=1}^r S_l^*(t). \quad (5.6)$$

Por outro lado, temos também

$$S(t) = P(T \geq t) = \sum_{l=1}^r P(T \geq t, C = l) = \sum_{l=1}^r S_l(t),$$

onde  $S_l(t)$  é designada por função de sobrevivência específica da causa  $l$ . Admitindo que  $T$  é uma v.a. contínua, como é mais usual, as correspondentes função de densidade e de distribuição (também designada por função de incidência cumulativa) são, respectivamente,

$$f_l(t) = h_l(t)S(t) \quad (5.7)$$

e

$$F_l(t) = P(T < t, C = l) = \int_0^t f_l(u) du. \quad (5.8)$$

Note-se que  $F_l(t) + S_l(t) = p_l$ , onde

$$p_l = P(C = l) = F_l(\infty) = S_l(0) \quad (5.9)$$

é a distribuição marginal de  $C$ , ou seja, é a probabilidade de um indivíduo falhar devido à causa  $l$ . Então,  $F_l(t)$  é uma função de distribuição imprópria dado que, quando  $t = \infty$ , o seu valor é  $p_l$  e não 1. Naturalmente, estamos a supor que  $p_l > 0$  e que  $\sum p_l = 1$ .

Algumas probabilidades condicionais de interesse neste contexto são:

$$P(\text{morte no instante } t \mid C = l) = \frac{f_l(t)}{p_l},$$

$$P(C = l \mid \text{morte no instante } t) = \frac{f_l(t)}{f(t)},$$

$$P(C = l \mid T \geq t) = \frac{S_l(t)}{S(t)}.$$

O risco relativo de falha devido à causa  $l$  no instante  $t$  é dado por  $h_l(t)/h(t)$ . Se esta razão for independente de  $t$  para cada  $l$ , então as funções hazard são proporcionais, ou seja,  $h_l(t) = p_l h(t)$ . Na verdade, é uma forma equivalente de indicar que  $T$  e  $C$  são independentes.

### Estimação paramétrica e não paramétrica para a probabilidade de falha

Suponhamos que foram especificados modelos paramétricos para as subdensidades  $f_l(t)$ . A situação mais usual consiste em admitir a mesma distribuição para cada uma das subdensidades  $f_l(t)$ , embora com diferentes valores dos parâmetros. No entanto, é possível considerar distribuições de diferentes famílias para cada uma das subdensidades.

Admitamos ainda que o mecanismo de censura é independente do mecanismo de falha. Consideremos uma amostra de  $n$  elementos,  $\{(l_i, t_i) : i = 1, \dots, n\}$ . A função de verosimilhança escreve-se na forma

$$L = \prod_o f_{l_i}(t_i) \prod_{\bar{o}} S(t_i), \quad (5.10)$$

onde  $\prod_O$  representa o produto para todos os casos em que a falha foi observada e  $\prod_{\bar{O}}$  denota o produto para as observações censuradas, ou seja, as observações correspondentes aos indivíduos para os quais não foi observado qualquer tipo de falha até ao instante  $t_i$ . Dependendo das funções empregues no cálculo da função de verosimilhança, existem algumas expressões equivalentes a (5.10). Por (5.7), obtemos

$$L = \prod_O h_{l_i}(t_i) \prod_{i=1}^n S(t_i).$$

Se associarmos a cada indivíduo a variável indicatriz de censura,  $\delta_i$ , então

$$L = \prod_{i=1}^n h_{l_i}(t_i)^{\delta_i} S(t_i). \quad (5.11)$$

Por outro lado, tendo em conta (5.6), a função de verosimilhança pode ser escrita apenas à custa das funções hazard específicas de cada causa:

$$L = \prod_O h_{l_i}(t_i) \prod_{i=1}^n \prod_{l=1}^r \exp \left[ - \int_0^{t_i} h_l(u) du \right].$$

Considerando

$$L_l = \prod_i h_l(t_i) \prod_{i=1}^n \exp \left[ - \int_0^{t_i} h_l(u) du \right],$$

onde  $\prod_i$  denota o produto em  $\{i : l_i = l\}$ , a equação (5.11) pode ser expressa por  $L = L_1 L_2 \dots L_r$ . Na prática, pode ser mais simples maximizar  $L_l$  separadamente dado que envolve menos parâmetros do que  $L$ .

Vejamos agora em que consiste a abordagem não paramétrica com vista a estimar a probabilidade de falha.

Notemos que é frequente (embora incorrecto) estimar a probabilidade de falha devido à causa  $l$ , na presença das outras falhas, como um menos o

estimador de Kaplan-Meier da função de sobrevivência calculado na maior observação, onde o tempo de vida dos indivíduos que falham devido a outra causa que não a  $l$  é considerado como uma observação censurada. Na verdade, o estimador assim obtido é um estimador enviesado pois, regra geral, sobrestima a probabilidade pretendida. Uma consequência deste erro é a soma das probabilidades marginais,  $p_l, l = 1, \dots, r$ , poder ser substancialmente maior do que um, como referem Gaynor *et al.* (1993).

Suponhamos que existem  $k$  instantes de morte distintos  $0 < t_{(1)} < \dots < t_{(k)}$  e que a função hazard específica da causa  $l$  é definida por

$$h_l(t) = \begin{cases} h_{lj} & \text{se } t = t_{(j)} \\ 0 & \text{c.c.} \end{cases} \quad (5.12)$$

onde  $l = 1, \dots, r, j = 1, \dots, k$  e  $t_{(0)} = 0$  denota o início do estudo. Neste caso,  $h_{lj}$  representa a probabilidade de um indivíduo falhar no instante  $t_{(j)}$  devido à causa  $l$ , dado que sobreviveu para além do instante  $t_{(j-1)}$ . Por Gaynor *et al.* (1993), a função de verosimilhança é escrita na forma

$$L = \prod_{j=1}^k \left[ \left( \prod_{l=1}^r h_{lj}^{d_{lj}} \right) \left( 1 - \sum_{l=1}^r h_{lj} \right)^{n_j - d_j} \right],$$

onde  $n_j$  denota o número de indivíduos em risco no instante  $t_{(j)}$ ,  $d_{lj}$  o número de indivíduos que falham em  $t_{(j)}$  devido à causa  $l$  e  $d_j = \sum_{l=1}^r d_{lj}$ . O estimador de máxima verosimilhança de  $h_{lj}$  é  $\hat{h}_{lj} = d_{lj}/n_j$ , donde, o estimador de máxima verosimilhança de  $H_l(t)$  é

$$\hat{H}_l(t) = \sum_{\{j:t_{(j)} \leq t\}} d_{lj}/n_j.$$

Quanto ao estimador de máxima verosimilhança de  $F_l(t)$ , é dado pela expressão

$$\hat{F}_l(t) = \sum_{\{j:t_{(j)} \leq t\}} \frac{d_{lj}}{n_j} \hat{S}(t_{(j-1)}), \quad (5.13)$$

em que

$$\hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} \left(1 - \frac{d_i}{n_i}\right)$$

é a estimativa de Kaplan-Meier da probabilidade de um indivíduo sobreviver a todas as causas de morte, pelo menos até ao instante  $t_{(j-1)}$ . Assim, o EMV da probabilidade de falha devida à causa  $l$  é

$$\hat{p}_l = \hat{F}_l(t_{(k)}). \quad (5.14)$$

Como  $\sum_{l=1}^r \hat{F}_l(t) = 1 - \hat{S}(t)$ , então  $\sum_{l=1}^r \hat{p}_l \leq 1$ , verificando-se a igualdade no caso de não existirem observações censuradas para além de  $t_{(k)}$ . Na ausência de censura,  $\hat{p}_l = \sum_{j=1}^k d_{lj}/n$ , onde  $n$  é a dimensão da amostra.

Considerando as igualdades (5.3) e (5.12), temos

$$S_l^*(t) = \prod_{\{j:t_{(j)} \leq t\}} (1 - h_{lj}),$$

donde, o correspondente EMV escreve-se na forma

$$\hat{S}_l^*(t) = \prod_{\{j:t_{(j)} \leq t\}} \left(1 - \frac{d_{lj}}{n_j}\right). \quad (5.15)$$

Note-se que  $\hat{S}_l^*(t)$  representa a estimativa de Kaplan-Meier da função de sobrevivência para a causa  $l$ , em que os indivíduos que falham devido a uma causa diferente de  $l$  dão origem a observações censuradas, isto é,  $d_j$  é substituído por  $d_{lj}$ .

Vemos assim que o uso de  $1 - \hat{S}_l^*(t)$  para estimar  $F_l(t)$  é baseado no pressuposto incorrecto de que a probabilidade de falha até ao instante  $t$ , devido a qualquer outra causa diferente de  $l$ , é nula. Isto conduz a um valor inflacionado do número de indivíduos em risco no instante  $t$ , que tem como consequência a sobrestimação de  $p_l$ . Mostra-se por indução que  $1 - \hat{S}_l^*(t) \geq \hat{F}_l(t)$ . Com uma representação alternativa dos estimadores  $\hat{F}_l(t)$  e  $\hat{S}_l^*(t)$ , Gooley *et al.* (1999) evidenciam de forma clara este facto.

### Modelos de regressão

A cada indivíduo pode ser associado um vector de covariáveis  $\mathbf{z}$ . Deste modo, podem ser construídos modelos de regressão de riscos competitivos, que são essencialmente generalizações do caso em que existe apenas uma causa de morte.

Por conseguinte, como para uma causa de morte o modelo de hazards proporcionais é  $h(t; \mathbf{z}) = h_0(t)\alpha(\mathbf{z})$ , uma generalização natural ocorre definindo

$$h_l(t; \mathbf{z}) = h_{0l}(t)\alpha_l(\mathbf{z}).$$

Se  $T$  e  $C$  forem independentes, então o modelo passa a ser dado por  $h_l(t; \mathbf{z}) = p_l h_0(t)\alpha_l(\mathbf{z})$ .

O modelo de tempo de vida acelerado, no caso de uma causa de morte, pode ser definido por  $S(t; \mathbf{z}) = S_0(t\alpha(\mathbf{z}))$  logo, o correspondente modelo de riscos competitivos, é dado por

$$S_l(t; \mathbf{z}) = S_{l0}(t\alpha_l(\mathbf{z})).$$

Por fim, para o modelo de possibilidades proporcionais, em que  $S(t; \mathbf{z})/[1 - S(t; \mathbf{z})] = \alpha(\mathbf{z})S_0(t)/[1 - S_0(t)]$ , a versão para riscos competitivos é

$$\frac{S_l(t; \mathbf{z})}{1 - S_l(t; \mathbf{z})} = \alpha_l(\mathbf{z}) \frac{S_{l0}(t; \mathbf{z})}{1 - S_{l0}(t; \mathbf{z})}.$$

### 5.2.3 Modelos de mistura

Uma alternativa à formulação anterior, proposta por Larson e Dinse (1985), consiste em adoptar um modelo de mistura que caracteriza a distribuição do par  $(T, C)$  através da distribuição marginal do tipo de falha e da distribuição do tempo de vida condicional ao tipo de falha. Assim, a função de sobrevivência de  $T$  é

$$S(t) = p_1 S_{1M}(t) + \dots + p_r S_{rM}(t) \tag{5.16}$$

onde  $p_l = P(C = l)$ ,  $\sum_{l=1}^r p_l = 1$  e  $S_{lM}(t) = P(T \geq t | C = l)$ , com

$$S_{lM}(t) = \exp \left[ - \int_0^t h_{lM}(u) du \right]$$

e

$$h_{lM}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, C = l)}{\Delta t}.$$

Na presença de covariáveis, é adoptado um modelo de regressão logística para a probabilidade de falha e, frequentemente, um modelo de hazards proporcionais para a distribuição do tempo de vida condicional ao tipo de falha.

Sob o modelo (5.16), a contribuição de um indivíduo para a função de verosimilhança é:

- $p_l h_{lM}(t) S_{lM}(t)$  se foi observada a falha de tipo  $l$  no instante  $t$ ,
- $\sum_{l=1}^r p_l S_{lM}(t)$  se não foi observado qualquer tipo de falha até ao instante  $t$ .

Deste modo, a contribuição para a função de verosimilhança de um tempo de vida observado corresponde à função de densidade conjunta de  $(T = t, C = l)$  e a contribuição de um tempo de vida censurado corresponde à probabilidade de  $T \geq t$ . Recuperamos desta forma a função de verosimilhança (5.10).

Um dos grandes méritos desta formulação consiste na fácil utilização do algoritmo EM. Para tal, basta atribuir uma causa de morte às observações censuradas com a seguinte probabilidade:

$$\tau_l(t) = P(C = l | T \geq t) = \frac{p_l S_{lM}(t)}{\sum_{j=1}^r p_j S_{jM}(t)}, \quad l = 1, \dots, r.$$

Para estimativa inicial dos parâmetros, Larson e Dinse (1985) propõem as obtidas quando se ignora as observações censuradas.

## 5.3 Modelo de cura

Os modelos de cura considerados até agora pressupõem que existia uma única causa de interesse. Como tal, a morte de um indivíduo durante o período de *follow-up* só podia ser devido a essa causa e os indivíduos imunes seriam encontrados entre aqueles cujo tempo de vida tinha sido censurado.

Vamos considerar uma generalização do modelo de cura, inserindo-o na teoria dos riscos competitivos. No entanto, a definição de imune já não é única e obriga a algumas considerações. Teremos essencialmente dois tipos de modelos: aqueles em que os tempos de cura são observados e aqueles em que não são (situação muito mais frequente).

### 5.3.1 Relação entre os modelos de cura e os riscos competitivos

No âmbito dos riscos competitivos, surgem várias possibilidades para a definição de indivíduo imune (ou curado). Assim, um indivíduo imune pode ser aquele que seja imune a todas as causas de morte ou aquele que seja imune à causa de interesse embora susceptível a qualquer uma das restantes causas de morte. Neste último caso, quando temos apenas duas causas de morte (a de interesse e a outra), um indivíduo imune é, portanto, aquele que é imune à causa de interesse.

Além disso, podemos ainda considerar a existência de indivíduos imunes para cada um dos tipos de falha, sendo então susceptíveis às restantes. Por exemplo, se tivermos apenas dois tipos de falha, podem existir indivíduos imunes à falha de tipo 1 mas susceptíveis à falha de tipo 2 e vice-versa.

Por outras palavras, isto significa que a definição de cura pode dizer respeito a

1. todas as causas;
2. uma causa em particular;
3. cada uma das causas.

Então surge o problema de saber afinal qual a definição de indivíduo imune no contexto de riscos competitivos. Parece-nos que a resposta não é única nem pacífica pois podem ser apresentados argumentos a favor e contra

qualquer uma das três opções. Mais relevante, a resposta terá que ser dada em função de alguns factores, nomeadamente a natureza dos dados e o objectivo do estudo.

Notemos que poderíamos pensar que um indivíduo que fosse susceptível à causa digamos de tipo 1, fosse imune às restantes  $(r - 1)$  causas, uma vez que estamos a admitir que quando é observada a morte de um indivíduo, ela é provocada por apenas uma das  $r$  causas, ou seja, que as  $r$  causas são mutuamente exclusivas. No entanto, porque a morte devido à causa 1 inviabiliza que se observe a morte devido a qualquer uma das restantes causas, não podemos afirmar, como é óbvio, que a morte devido à causa 1 seja equivalente à imunidade às restantes causas.

Uma outra questão extremamente importante, que surge implicitamente das várias definições possíveis para a imunidade, tem a ver com a relação entre o conceito de imune e o modo de encarar os riscos competitivos.

Num modelo de cura, a função de sobrevivência é imprópria, ou seja,  $S(\infty) > 0$ . Então, quando consideramos um modelo de riscos competitivos do tipo (5.16), se também tivermos  $S(\infty) > 0$ , isto significa que existem indivíduos imunes a todas as causas consideradas e, assim sendo, as causas não são exaustivas pois na vida real nenhum indivíduo é imortal. Pelo contrário, quando  $S(\infty) = 0$  então todos os indivíduos são susceptíveis a pelo menos uma das causas de morte consideradas. Assim, através do valor de  $S(\infty)$  ( $= 0$  ou  $> 0$ ) podemos saber se as causas de morte que estamos a considerar são exaustivas ou não.

Quando uma causa é definida como sendo *restantes causas*, do ponto de vista formal as causas indicadas são exaustivas, logo  $S(\infty) = 0$ . No entanto, quando pretendemos indicar cada causa explicitamente, por muito longa que seja a lista, dificilmente incluirá todas elas; teremos então  $S(\infty) > 0$ . Uma forte indicação desta última situação é a existência de muitas observações censuradas, pois para um indivíduo com tempo de vida censurado não é possível saber se ele iria morrer de alguma das causas em estudo.

### 5.3.2 Cura observável

Embora esta seja uma situação pouco frequente, a sua existência justifica um modelo próprio. Trata-se de uma situação mais simples, dado que a cura é encarada como um risco competitivo. Neste sentido, podem-se aplicar

directamente os resultados da teoria de riscos competitivos, embora a interpretação seja diferente. Na realidade, os modelos apresentados não são verdadeiros modelos de cura, pois  $S(\infty) = 0$ , mas podem ser interpretados como tal dado que também se pretende estimar a probabilidade de cura e a função de sobrevivência dos indivíduos susceptíveis.

Nesta secção, vamos apresentar a abordagem não paramétrica descrita em Betensky e Schoenfeld (2001) e propor uma possível abordagem paramétrica, através do uso da distribuição de Gompertz.

### Abordagem não paramétrica

A síndrome de dificuldade respiratória no adulto (SDRA) é uma condição aguda, potencialmente fatal, que surge algumas vezes após uma pneumonia ou uma cirurgia. Os pulmões endurecem, dificultando as trocas de oxigénio entre a hemoglobina e o ar alveolar, e o doente tem que ser ventilado artificialmente. Os pacientes que recuperam e que têm alta hospitalar, são considerados curados no instante em que saem do hospital. Neste contexto, Betensky e Schoenfeld (2001) consideram a cura como um risco competitivo ao acontecimento de interesse (morte por SDRA), pois não há dúvida que a sua ocorrência impede a observação do acontecimento de interesse. Deste modo, apenas é observado o mínimo entre o tempo até à cura e o tempo até à ocorrência do acontecimento de interesse. Os pressupostos subjacentes a este modelo são que os três tipos de resposta possível (morte, cura ou censura) são independentes e que ocorrerá a falha ou a cura com probabilidade um (apesar de qualquer uma delas poder não ser observada).

Neste caso, não vamos ter  $k$  instantes de morte distintos, pois alguns desses são de facto instantes de cura. Defina-se  $h_j$  como a probabilidade de morte no instante  $t_j$ , dado que o indivíduo não morreu nem ficou curado até então, e defina-se de modo análogo  $\gamma_j$  como sendo a probabilidade condicional de cura em  $t_j$ . Sejam ainda  $n_j$ ,  $d_j$  e  $m_j$  o número de indivíduos em risco, que morreram e que se curaram, respectivamente, no instante  $t_j$ . O logaritmo da função de verosimilhança é

$$\log \hat{L} = \sum_{j=1}^k [d_j \log(\lambda_j) + m_j \log(\gamma_j) + (n_j - d_j - m_j) \log(1 - \lambda_j - \gamma_j)], \quad (5.17)$$

donde  $\hat{\lambda}_j = d_j/n_j$  e  $\hat{\gamma}_j = m_j/n_j$ . A variável  $T$  representa o tempo até à morte e a variável  $U$  representa o tempo até à cura. A distribuição do tempo de vida,  $S(t) = P(T > t)$ , é estimada por

$$\hat{S}(t) = 1 - \sum_{j=1}^{v_t} \hat{\lambda}_j \prod_{v=1}^{j-1} (1 - \hat{\lambda}_v - \hat{\gamma}_v) \quad (5.18)$$

para  $t \in [t_{v_t}, t_{v_t+1})$ . Como a expressão anterior coincide com um menos a função de incidência cumulativa, por (5.9) e (5.14), a probabilidade de cura,  $\pi$ , pode ser estimada por

$$\hat{p}_1 = 1 - \hat{P}(T < \infty) = 1 - \sum_{j=1}^k \hat{\lambda}_j \prod_{v=1}^{j-1} (1 - \hat{\lambda}_v - \hat{\gamma}_v),$$

ou por

$$\hat{p}_2 = \hat{P}(U < \infty) = \sum_{j=1}^k \hat{\gamma}_j \prod_{v=1}^{j-1} (1 - \hat{\lambda}_v - \hat{\gamma}_v).$$

Note-se que  $\hat{p}_2$  vai coincidir com (5.14). Se a maior observação não for censurada,  $\hat{p}_1 = \hat{p}_2$ . No entanto, se a maior observação for censurada,  $\hat{p}_1$  e  $\hat{p}_2$  não são equivalentes, logo o EMV de  $\pi$  não é único. Mais precisamente,  $\pi \in (\hat{p}_2, \hat{p}_1)$ .

A formulação deste modelo em termos de um modelo de mistura requer a definição de mais algumas quantidades, nomeadamente,  $q_j = P(T = t_j | T < U)$ ,  $p_j = P(U = t_j | T > U)$ ,  $\pi = P(U < T)$ ,  $Q_a = \sum_{j=a}^k q_j$ , e  $P_a = \sum_{j=a}^k p_j$ . Desta forma,

$$P(T > t_j, U > t_j) = \pi P_{j+1} + (1 - \pi) Q_{j+1}. \quad (5.19)$$

As contribuições para a função de verosimilhança no instante  $t_j$  são  $(1 - \pi)q_j$  quando é observada uma morte,  $\pi p_j$  quando é observada a cura e  $\pi P_{j+1} + (1 - \pi)Q_{j+1}$  quando a observação é censurada. Assim, a função de verosimilhança pode ser expressa por

$$\prod_{j=1}^k [(1 - \pi)q_j]^{d_j} (\pi p_j)^{m_j} [\pi P_{j+1} + (1 - \pi)Q_{j+1}]^{n_j - n_{j+1} - d_j - m_j}.$$

Como acontece no caso anterior, se a maior observação é censurada, não existe unicidade do EMV de  $\pi$ .

A definição de  $\lambda_j$  e de  $\gamma_j$  em termos deste modelo é

$$\lambda_j = \frac{(1 - \pi)q_j}{\pi P_j + (1 - \pi)Q_j}$$

e

$$\gamma_j = \frac{\pi p_j}{\pi P_j + (1 - \pi)Q_j},$$

donde, usando a igualdade

$$\pi P_{j+1} + (1 - \pi)Q_{j+1} = \prod_{b=1}^j (1 - \lambda_b - \gamma_b),$$

a função de verosimilhança pode ser reescrita na forma

$$\prod_{j=1}^k \lambda_j^{d_j} \gamma_j^{m_j} (1 - \lambda_j - \gamma_j)^{-d_j - m_j} \left[ \prod_{b=1}^j (1 - \lambda_b - \gamma_b)^{n_j - n_{j+1}} \right],$$

cujos logaritmo é equivalente a (5.17).

### Abordagem paramétrica

Tal como no caso não paramétrico, admitamos que a cura é um risco competitivo; assim sendo, propomos a utilização de um modelo de mistura tal como em (5.16), com  $r = 2$ . Então o modelo é

$$S(t) = p_c S_{cM}(t) + p_d S_{dM}(t), \quad (5.20)$$

onde  $p_d$  e  $p_c$  representam a probabilidade de um indivíduo ser susceptível e a probabilidade de um indivíduo ser imune, respectivamente, e  $S_{dM}$  e  $S_{cM}$  as correspondentes funções de sobrevivência. Admitamos ainda uma distribuição contínua para o tempo de vida.

Consideremos uma amostra de dimensão  $n$  e designemos por  $t_1, \dots, t_n$  os tempos de vida. Sem perda de generalidade, suponhamos que os primeiros tempos de vida correspondem aos indivíduos curados, os seguintes aos indivíduos susceptíveis e os últimos aos indivíduos para os quais não foi observada a morte nem a cura, ou seja, aos indivíduos com tempo de vida censurado. Deste modo, temos

$$\sum_{i=1}^n t_i = \sum_{i=1}^{n_c} t_i + \sum_{i=n_c+1}^{n_d} t_i + \sum_{i=n_d+1}^n t_i,$$

onde  $n_c$  e  $n_d - n_c$  designam o número de indivíduos para os quais foi observada a cura ou a morte, respectivamente, e  $n - n_d$  o número de indivíduos com tempo de vida censurado. Sejam  $\delta_{ic}, \delta_{id}, \delta_{i0}$ , variáveis indicatrizes tais que  $\delta_{il} = 1$  se o  $i$ -ésimo indivíduo falhou devido à causa  $l$  ( $l = c, d$ ), no instante  $t_i$ , e  $\delta_{il} = 0$  caso contrário, e  $\delta_{i0} = 1$  se o  $i$ -ésimo indivíduo foi censurado no instante  $t_i$ . Note-se que  $\delta_{ic} + \delta_{id} + \delta_{i0} = 1, \forall i \in \{1, \dots, n\}$ . Consideremos ainda  $y_{1l}, \dots, y_{nl}, l = d, c$  tais que  $y_{il} = 1$  se o  $i$ -ésimo indivíduo falhou devido à causa  $l$  e  $y_{il} = 0$  se falhou devido à outra causa.

A função de verosimilhança completa para este caso é

$$L_C = \prod_{i=1}^n [p_c h_{cM}(t_i) S_{cM}(t_i)]^{\delta_{ic}} [p_d h_{dM}(t_i) S_{dM}(t_i)]^{\delta_{id}} S(t_i)^{\delta_{i0}},$$

onde  $h_{dM}(t)$  e  $h_{cM}(t)$  representam a função hazard associada à causa  $d$  e  $c$ , respectivamente. Se todos os  $y_{il}$  fossem observados teríamos a verosimilhança observada

$$L_O = \prod_{i=1}^n [[p_c h_{cM}(t_i) S_{cM}(t_i)]^{y_{ic}}]^{\delta_{ic}} [[p_d h_{dM}(t_i) S_{dM}(t_i)]^{y_{id}}]^{\delta_{id}} \prod_{l=c,d} [[p_l S_{lM}(t_i)]^{y_{il}}]^{\delta_{i0}},$$

a qual pode ser factorizada na forma  $L_O = L_{O_p} L_{O_c} L_{O_d}$ , onde

$$L_{O_p} = \prod_{i=1}^n p_c^{y_{ic}(\delta_{ic} + \delta_{i0})} p_d^{y_{id}(\delta_{id} + \delta_{i0})}, \quad (5.21)$$

$$L_{O_c} = \prod_{i=1}^n [h_{cM}(t_i) S_{cM}(t_i)]^{y_{ic} \delta_{ic}} S_{cM}(t_i)^{y_{ic} \delta_{i0}} \quad (5.22)$$

e

$$L_{O_d} = \prod_{i=1}^n [h_{dM}(t_i) S_{dM}(t_i)]^{y_{id} \delta_{id}} S_{dM}(t_i)^{y_{id} \delta_{i0}}. \quad (5.23)$$

Como  $p_d + p_c = 1$ , (5.21) tanto pode ser escrita à custa de  $p_d$

$$L_{O_p} = \prod_{i=1}^n p_d^{y_{id}(\delta_{id} + \delta_{i0})} (1 - p_d)^{y_{ic}(\delta_{ic} + \delta_{i0})},$$

como à custa de  $p_c$

$$L_{O_p} = \prod_{i=1}^n (1 - p_c)^{y_{id}(\delta_{id} + \delta_{i0})} p_c^{y_{ic}(\delta_{ic} + \delta_{i0})}.$$

Em geral, nem todos os  $y_{il}$  são observados. Então, cada  $y_{il}$  é substituído por  $\omega_{il}$ , onde

$$\omega_{il} = \begin{cases} 1 & \text{se } \delta_{id} = 1 \vee \delta_{ic} = 1 \\ \tau_{il} & \text{se } \delta_{i0} = 1 \end{cases}$$

e

$$\tau_{il} = E(Y|\mathcal{O}) = P(Y_{ij} = 1 | T > t_i, \delta_{i0} = 1) = \frac{p_l S_{lM}(t_i)}{S(t_i)}, \quad l = c, d.$$

Assim, o valor esperado do logaritmo da função de verosimilhança, no que concerne ao parâmetro  $p$ , pode ser escrito na forma

$$\log L_{E_p} = \sum_{i=1}^n [\omega_{id}(\delta_{id} + \delta_{i0}) \log p_d + \omega_{ic}(\delta_{ic} + \delta_{i0}) \log(1 - p_d)],$$

ou seja, simplificando

$$\log L_{E_p} = \left( (n_d - n_c) + \sum_{i=n_d+1}^n \tau_{id} \right) \log p_d + \left( n_c + \sum_{i=n_d+1}^n \tau_{ic} \right) \log(1 - p_d).$$

O estimador de  $p_d$  na  $(k + 1)$ -ésima iteração do algoritmo EM, obtido através do método da máxima verosimilhança é

$$\hat{p}_d^{(k+1)} = \frac{1}{n} \left[ (n_d - n_c) + \sum_{i=n_c+1}^n \tau_{id}^{(k)} \right]. \quad (5.24)$$

Note-se a semelhança com (4.15). De modo análogo, obteríamos o estimador de máxima verosimilhança de  $p_c$ .

Relativamente a (5.23), o valor esperado do seu logaritmo é

$$\log L_{E_d} = \sum_{i=n_c+1}^{n_d} [\log h_{dM}(t_i) + \log S_{dM}(t_i)] + \sum_{i=n_d+1}^n \tau_{id} \log S_{dM}(t_i).$$

Admitamos que o tempo de vida dos indivíduos susceptíveis tem distribuição de Gompertz, de parâmetros  $\lambda_d$  e  $\alpha_d$ , isto é,

$$S_{dM}(t) = \exp \left[ \frac{\alpha_d}{\lambda_d} [1 - \exp(\lambda_d t)] \right].$$

Analogamente, admitamos uma distribuição de Gompertz de parâmetros  $\lambda_c$  e  $\alpha_c$  para o tempo de vida dos indivíduos curados. Recorde-se que a identifiabilidade deste modelo foi provada por Gordon (1990). Então,

$\log L_{E_d} =$

$$\sum_{i=n_c+1}^{n_d} \left[ \log \alpha_d + \lambda_d t_i + \frac{\alpha_d}{\lambda_d} [1 - \exp(\lambda_d t_i)] \right] + \sum_{i=n_d+1}^n \tau_{id} \frac{\alpha_d}{\lambda_d} [1 - \exp(\lambda_d t_i)]. \quad (5.25)$$

É possível obter uma forma explícita para o EMV de  $\alpha_d$ ,

$$\hat{\alpha}_d = \frac{n_d \lambda_d}{\sum_{i=n_c+1}^{n_d} [\exp(\lambda_d t_i) - 1] + \sum_{i=n_d+1}^n \tau_{id} [\exp(\lambda_d t_i) - 1]}, \quad (5.26)$$

mas o mesmo já não é verdade para  $\lambda_d$ . Daí que, adicionalmente, seja necessário recorrer ao método de N-R para estimar  $\lambda_d$ .

### 5.3.3 Cura não observável

O modelo de cura na presença de riscos competitivos em que a cura não é observável é adequado quando, como o próprio nome indica, a cura não é observável e se admite que um indivíduo está sujeito a mais do que uma causa de morte. No entanto, quando se diz que um indivíduo está curado, a definição de cura tanto pode estar associada apenas a uma causa em particular (correspondente à causa de interesse), a todas as causas, ou a cada uma delas individualmente. Ao contrário do que sucedia no caso anterior, neste caso a cura não é considerada um risco competitivo.

#### Modelos já propostos

Começamos por abordar o modelo referido em Ng e McLachlan (1998). Trata-se de uma generalização do modelo (4.2), na medida em que também se adopta um modelo de mistura com duas componentes, mas em que uma delas representa a causa de interesse e a outra as restantes causas. Larson e Dinse (1985) tinham proposto este modelo, embora não no contexto dos modelos de cura. O objectivo principal continua a ser estimar a distribuição do tempo de vida dos indivíduos susceptíveis, bem como a probabilidade de cura. Apresentamos duas versões para a função de verosimilhança, total e

parcial, sendo a última mais vantajosa em termos práticos pois não envolve o conhecimento da distribuição dos tempos de vida dos indivíduos que falham devido a uma causa que não a causa de interesse.

As observações são da forma  $(t_i, D_i)$ , onde  $D_i = 2$  indica que o  $i$ -ésimo indivíduo falhou no instante  $t_i$  devido à causa de interesse,  $D_i = 1$  se o  $i$ -ésimo indivíduo falhou no instante  $t_i$  devido a uma causa competitiva e  $D_i = 0$  se a observação correspondente ao  $i$ -ésimo indivíduo é censurada em  $t_i$ . Seja  $T$  o tempo até à falha devido à causa de interesse ou devido a uma causa competitiva. Como vimos anteriormente, de acordo com Larson e Dinse (1985), a função de sobrevivência marginal de  $T$  pode ser dada por

$$S(t, \Psi) = pS_1(t; \theta_1) + (1 - p)S_2(t; \theta_2), \quad (5.27)$$

onde  $\Psi = (p, \theta'_1, \theta'_2)'$  designa o vector de parâmetros desconhecidos,  $S_1(t; \theta_1)$  é a função de sobrevivência associada aos indivíduos que falham devido a uma causa competitiva e  $S_2(t; \theta_2)$  é a função de sobrevivência associada à causa de interesse, na presença das restantes causas. De facto,  $S_k(t; \theta_k)$  é a função de sobrevivência condicional do tempo até à morte dado que a morte se ficou a dever à causa  $k$ ,  $k = 1, 2$ . A probabilidade de falhar devido à causa de interesse, na presença das restantes causas, é dada por  $1 - p$ , enquanto que  $p$  representa a probabilidade de falha devido a uma causa competitiva. O logaritmo da função de verosimilhança correspondente ao modelo (5.27) escreve-se na forma

$$\begin{aligned} \log L(\Psi) = & \sum_{i=1}^n [I(D_i = 0) \log S(t_i; \Psi) + I(D_i = 1) \log \{pf_1(t_i; \theta_1)\} \\ & + I(D_i = 2) \log \{(1 - p)f_2(t_i; \theta_2)\}], \end{aligned} \quad (5.28)$$

em que  $I(A)$  é a função indicatriz do acontecimento  $A$  e  $f_k(t_i; \theta_k)$  é a função densidade correspondente à função de sobrevivência  $S_k(t_i; \theta_k)$ ,  $k = 1, 2$ . A função de verosimilhança anterior – função de verosimilhança total – tem a desvantagem de requerer a especificação da distribuição do tempo de vida dos indivíduos que morrem devido a uma falha competitiva. A estimação de  $\theta_1$  torna-se ainda mais difícil de obter quando existem poucas falhas devido às causas competitivas.

De acordo com Ng e McLachlan (1998), uma forma de contornar a dificuldade anterior, consiste em construir uma função de verosimilhança parcial. Para tal, as observações provenientes de falhas devidas às causas competitivas são consideradas como sendo observações censuradas no fim do período de *follow-up*. Assim, uma falha registada em  $t_i$  com  $D_i = 1$ , contribui para a função de verosimilhança como uma observação censurada não no instante  $t_i$  mas sim no instante  $t_i^* = t_i + u_i$ , onde  $u_i$  designa o tempo decorrido desde a falha até ao fim do período de *follow-up*. Deste modo, o logaritmo da função de verosimilhança é

$$\begin{aligned} \log L = & \sum_{i=1}^n [I(D_i = 2) \log\{(1-p)f_2(t_i; \boldsymbol{\theta}_2)\} \\ & + I(D_i = 1) \log\{p + (1-p)S_2(t_i^*; \boldsymbol{\theta}_2)\} \\ & + I(D_i = 0) \log\{p + (1-p)S_2(t_i; \boldsymbol{\theta}_2)\}]. \end{aligned} \quad (5.29)$$

A partir da função anterior podemos obter as estimativas de máxima verosimilhança de  $p$  e de  $\boldsymbol{\theta}_2$  sem ter que especificar  $S_1(t; \boldsymbol{\theta}_1)$ . Ng e McLachlan (1998) referem que esta última abordagem origina estimadores consistentes mas menos eficientes do que a abordagem com a função de verosimilhança total.

No modelo (5.27), a quantidade  $1-p$  representa a probabilidade de falha devido à causa de interesse, na presença das outras causas. Então,  $p$  é a probabilidade de um indivíduo não falhar da causa de interesse, o que significa que irá falhar de uma das causas competitivas. Nalgumas aplicações, como na estimação da taxa de cura do cancro da mama, as doentes que morrem precocemente após o tratamento, aparentemente de uma causa competitiva, podem não ser consideradas curadas mesmo se não tiverem sintomas da doença aquando da sua morte. Na verdade, relativamente às mortes precoces, não é claro se as doentes teriam ou não uma recaída da doença, se tivessem vivido tempo suficiente até que essa recaída se manifestasse. Neste tipo de aplicações,  $p$  pode ser ajustado, por exemplo, excluindo dos doentes curados aqueles cuja morte ocorra antes de ter decorrido um determinado tempo  $T_0$ , após o tratamento, que deve ser determinado antes do início do estudo. Assim, considera-se que um doente está curado da doença em estudo quando morre devido a uma causa competitiva a partir de  $T_0$  após o trata-

mento e sem que a doença se manifeste. Por outras palavras, a taxa de cura é dada por

$$pS_1(T_0).$$

Mesmo sem especificar  $S_1$ , se a menor observação censurada for menor que  $T_0$ , pode-se estimar  $S_1(T_0)$  por

$$1 - \frac{n_{T_0}}{n\hat{p}},$$

onde  $n$  é a dimensão da amostra e  $n_{T_0}$  é o número de pacientes que morreram devido a uma causa competitiva antes de  $T_0$ . Então, a taxa de cura pode ser estimada por

$$\hat{p} - \frac{n_{T_0}}{n}.$$

Assim, embora os autores não o refiram explicitamente, o modelo que estão a considerar é

$$S(t) = pS_1(T_0)S_1(t) + pF_1(T_0)S_1(t) + (1 - p)S_2(t).$$

Com o intuito de estimar a prevalência do cancro, Phillips *et al.* (2002) estudam um outro modelo de mistura na presença de riscos competitivos, que referimos nesta secção, embora não seja considerada explicitamente uma variável que representa causas de morte. A ideia geral, consiste em estimar a prevalência a partir da proporção de indivíduos na população aos quais foi previamente diagnosticado cancro e que ainda não estão curados. Daí a necessidade de estimar a probabilidade de cura.

Assim, consideram a existência de duas variáveis latentes que representam o tempo até à morte devido à doença,  $T_d$ , e o tempo até à morte devido a outras causas,  $T_o$ . Seja  $T = \min\{T_d, T_o\}$ . Os autores consideram que só se observa  $T$  sem atribuição da causa de morte pois têm dúvidas quanto ao rigor na atribuição das causas de morte nas certidões de óbito. Um indivíduo será definido como curado,  $Y = 0$ , se  $T_d = \infty$  e não curado,  $Y = 1$ , caso contrário. Assumindo que a morte devido à doença é independente da morte devido a outras causas, temos

$$\begin{aligned} P\{T > t\} &= P(T_o > t)P(T_d > t) \\ &= P\{T_o > t|Y = 0\}P(Y = 0) + P\{T_o > t|Y = 1\}P\{T_d > t|Y = 1\}P(Y = 1) \end{aligned} \quad (5.30)$$

Quando apenas  $T$  é observado, é prática comum calcular a sobrevivência relativa, a qual ajusta a mortalidade devido a outras causas através de tabelas de mortalidade relativas à população em geral. Isto é equivalente a considerar que

$$P\{T_o > t\} = S_0(t). \quad (5.31)$$

No entanto, os indivíduos afectados por alguns tipos de cancro, têm taxas de mortalidade devido a outras causas diferentes da população em geral. Daí que a equação (5.31) seja alterada para que este facto seja tido em consideração. Assim, supõe-se que a função hazard correspondente a outras causas é proporcional à função hazard da população em geral,  $h_0(t)$ , donde

$$P\{T_o > t\} = [S_0(t)]^\tau.$$

Deste modo, a função de sobrevivência global (5.30), é

$$P\{T > t\} = p [S_0(t)]^\tau + (1 - p)S_d(t) [S_0(t)]^\tau \quad (5.32)$$

e, suprimindo das funções o argumento  $t_i$ , a função de verosimilhança pode ser escrita na forma

$$\prod_i [p\tau h_0(S_0)^\tau + (1 - p)(\tau h_0 + h_d)(S_0)^\tau S_d]^{\delta_i} [p(S_0)^\tau + (1 - p)(S_0)^\tau S_d]^{1 - \delta_i}, \quad (5.33)$$

onde  $\delta_i$  é a variável indicatriz do estado vital do  $i$ -ésimo indivíduo no último instante,  $t_i$ , do seu *follow-up*.

Choi e Zhou (2002) e Maller e Zhou (2002) propõem uma outra abordagem, na medida em que, para eles, um indivíduo imune será aquele que for imune não apenas à causa de interesse, mas a todas as causas em estudo. Sejam  $u_i$  o tempo de censura e  $t_i^*$  o verdadeiro tempo de vida do  $i$ -ésimo indivíduo,  $i = 1, \dots, n$ ,  $u_i$  e  $t_i^*$  independentes. O tempo de vida observado é designado por  $t_i = \min\{t_i^*, u_i\}$  e  $\delta_i$  é a variável indicatriz de censura, que

toma o valor 1 quando o tempo de vida é observado e 0 caso contrário. Adicionalmente, consideram uma outra variável indicatriz,  $\delta_{il}$ , que toma o valor 1 se o  $i$ -ésimo indivíduo morre da causa  $l$  e 0 caso contrário. Naturalmente,

$$\delta_i = \sum_{l=1}^r \delta_{il}.$$

Definem também um indicador de causa de morte,  $l(i)$ , em que  $l(i) = \eta$  se o  $i$ -ésimo indivíduo morre da causa  $\eta$ , donde a esse indivíduo corresponde um tempo de vida não censurado. Para  $l = 1, \dots, r$ ,  $i = 1, \dots, n$ , definem  $t_{il} = t_i$  se  $\delta_{il} = 1$ , para os indivíduos cujas observações não são censuradas. Consideram ainda a variável aleatória  $Y_i$ , não observável, que assume o valor  $l$  se o  $i$ -ésimo indivíduo morre da causa  $l$  e 0 se o  $i$ -ésimo indivíduo é imune a todas as  $r$  causas, com probabilidades  $P(Y_i = l) = p_{il}$  e  $P(Y_i = 0) = 1 - p_i$ ,  $p_i = \sum_{l=1}^r p_{il}$ . A variável aleatória  $T_i$  representa o tempo de vida do  $i$ -ésimo indivíduo. Temos que

$$P(T_i = \infty | Y_i = 0) = 1$$

e

$$P(T_i \leq t | Y_i = l) = F_{il}(t), l = 1, \dots, r. \quad (5.34)$$

em que as funções  $F_{il}(t)$  são funções de distribuição próprias. Então,

$$\begin{aligned} P(T_i \leq t) &= P(T_i \leq t | Y_i = 0)P(Y_i = 0) + \sum_{l=1}^r P(T_i \leq t | Y_i = l)P(Y_i = l) \\ &= \sum_{l=1}^r p_{il}F_{il}(t) = F_i(t). \end{aligned} \quad (5.35)$$

Supondo que os tempos de vida são independentes e que a cada  $F_{il}(t)$  corresponde a função densidade  $f_{il}(t)$ , a menos de uma constante, a função de verosimilhança é dada por

$$\begin{aligned}
L_n &= \prod_{i=1}^n [p_{il(i)} f_{il(i)}(t_i)]^{\delta_i} \left[ 1 - \sum_{l=1}^r p_{il} F_{il}(t_i) \right]^{1-\delta_i} \\
&= \prod_{i=1}^n \prod_{l=1}^r [p_{il} f_{il}(t_i)]^{\delta_{il}} [1 - F_i(t_i)]^{1-\delta_i}.
\end{aligned} \tag{5.36}$$

### 5.3.4 Interligação entre o modelo aditivo com fragilidade, o modelo de cura e o de riscos competitivos

Já verificamos neste capítulo a ligação existente entre os modelos de cura e os modelos de riscos competitivos. Além disso, no capítulo anterior, referimos que uma das formas de obter um modelo de cura de não-mistura envolve considerar uma variável fragilidade. Assim, mais interessante se torna estabelecer a ponte com os modelos aditivos com fragilidade, viabilizando o estudo de um problema sob diversos aspectos e em contextos variados.

De acordo com Rocha (1995), o modelo aditivo (2.25) em que a fragilidade tem distribuição  $\chi_0^2(\gamma)$ , cuja função de sobrevivência populacional é dada por (2.29), pode ser interpretado como um modelo de riscos competitivos independentes em que existem duas causas de morte, sendo a imunidade admitida apenas para uma delas. Apesar de  $S(\infty) = 0$ , o que significa que todos os indivíduos acabam por morrer, podemos dizer que existem indivíduos imunes na população, só que essa imunidade é definida apenas para a causa 2. De facto, considerando duas causas de morte, supondo que os riscos são independentes, temos

$$S(t) = S_1(t)S_2(t).$$

$S_1(t)$  é a função de sobrevivência associada à causa 1 e, de acordo com Rocha (1995), é a função de sobrevivência correspondente aos indivíduos imunes à causa 2.  $S_2(t)$  é a função de sobrevivência associada à causa 2. Notemos que em relação à causa 2, existem indivíduos imunes e indivíduos susceptíveis.

Assim, temos

$$S(t) = \exp \left[ -\frac{\gamma}{2} (1 - (1 + 2t)^{-1}) \right] S_0(t),$$

onde  $S_2(t) = \exp\left[-\frac{\gamma}{2}(1 - (1 + 2t)^{-1})\right]$  e  $S_1(t) = S_0(t)$ , donde verificamos que  $S_2(\infty) = \exp(-\gamma/2)$ .

O facto de  $S_2(\infty) > 0$ , permite que  $S_2(t)$  seja escrita na forma de um modelo de cura de mistura, concretamente

$$S_2(t) = p + (1 - p)S_{2d}(t),$$

onde  $p = \exp(-\gamma/2)$  representa a proporção de indivíduos imunes à causa 2 e  $S_{2d}(t) = (S_2(t) - p)/(1 - p)$  é a função de sobrevivência dos indivíduos susceptíveis à causa 2. Então, a função de sobrevivência populacional também pode ser escrita do modo que se segue

$$S(t) = pS_1(t) + (1 - p)S_1(t)S_{2d}(t).$$

Note-se que  $S_{2d}(t)$  é uma função de sobrevivência relativa (Chao, 1998) e que  $S_1(t)S_{2d}(t)$  é a função de sobrevivência dos indivíduos susceptíveis às causas 1 e 2.

No entanto, como a imunidade é considerada apenas para a causa 2, isto significa que todos os indivíduos são susceptíveis à causa 1, donde  $S_1(\infty) = 0$ , logo  $S(\infty) = 0$ , o que coincide com o resultado referido inicialmente. Concluimos então que a partir de um modelo aditivo com fragilidade, podemos ter um modelo de riscos competitivos, o qual pode ser interpretado à luz dos modelos de cura, apesar de não ser um verdadeiro modelo de cura dado que  $S(\infty) = 0$ .

## 5.4 Modelo de cura com censura parcialmente informativa

### 5.4.1 Motivação

No contexto dos modelos de cura, a censura tem um papel extremamente importante e é uma fonte de informações úteis. Por um lado, o padrão de censura é a chave para podermos suspeitar da presença de indivíduos imunes na população. De facto, se a maioria das observações censuradas corresponder às maiores observações, é admissível que existam indivíduos que nunca vão experimentar o acontecimento de interesse. Por outro lado, embora, em geral, o mecanismo de censura seja não informativo, existem situações concretas no âmbito da prática clínica, em que a censura informativa é um facto

que não é passível de ser ignorado. Um exemplo típico refere-se ao estudo do tempo de vida de pacientes após um tratamento, onde supomos que os indivíduos susceptíveis estão sujeitos a mais do que uma causa de morte: a de interesse e outra com ela relacionada. Assim sendo, podemos admitir que a distribuição do tempo de vida dos indivíduos que chegam vivos ao fim do estudo não vai conter informação que possa contribuir para o conhecimento do tempo de vida dos indivíduos que morrem da causa de interesse, e que é perfeitamente plausível que o mesmo não aconteça relativamente à distribuição do tempo de vida dos indivíduos que morrem de outra causa. Em resumo, consideramos o tempo até à morte pela outra causa como variável de censura informativa e o tempo de vida dos indivíduos que chegam vivos ao fim do estudo como variável de censura não-informativa.

Neste cenário, julgamos relevante explorar uma nova abordagem, que consiste em considerar um modelo de cura com censura parcialmente informativa, o que significa, em termos gerais, que o tempo de vida dos indivíduos vai estar sujeito a dois tipos de censura: informativa e não-informativa.

Até à data, a experiência na aplicação de métodos que contemplem censura informativa é muito limitada, porventura devido às dificuldades associadas quer à modelação quer à computação. Conforme já referimos anteriormente, um modelo simples com censura informativa é o designado modelo de Koziol-Green, onde a função de sobrevivência correspondente à variável tempo de censura é uma potência da função de sobrevivência da variável tempo de vida.

Sejam  $T$  e  $C$  variáveis aleatórias que representam o tempo de vida e o tempo de censura, respectivamente. Em termos formais, o modelo de Koziol-Green é definido por

$$\bar{G}_1(t) = (S(t))^\beta, \quad t \geq 0,$$

para algum  $\beta > 0$ , onde  $\bar{G}_1$  e  $S$  são as funções de sobrevivência de  $C$  e  $T$ , respectivamente. O parâmetro  $\beta$  pode ser interpretado como parâmetro de censura e se  $\beta = 0$  significa que não existe censura.

Fazendo uso deste modelo, têm surgido recentemente alguns trabalhos com censura parcialmente informativa, nomeadamente Gather e Pawlitschko (1998), Braekers e Veraverbeke (2001) e Zhang e Rao (2004), mas não no contexto dos modelos de cura.

A nossa proposta é um modelo no qual, além de se admitir que parte da censura é informativa, também se admite a presença de indivíduos curados. Neste sentido, pode ser considerado uma generalização do modelo proposto por Gather e Pawlitschko (1998). Noutra perspectiva, pode ser considerado uma generalização do modelo de cura de não-mistura proposto por Yakovlev *et al.* (1993), pois inclui a censura informativa.

### 5.4.2 O modelo de Koziol-Green parcial

Este modelo, também designado por modelo de Koziol-Green generalizado, foi proposto por Gather e Pawlitschko (1998) e é adequado para situações em que existem dois tipos de censura, informativa e não-informativa. Assim, além da variável que representa o tempo de vida,  $T$ , temos não só a variável correspondente à censura não-informativa,  $D$ , como ainda a variável correspondente à censura informativa,  $C$ .

Consideremos as variáveis aleatórias  $T$ ,  $C$  e  $D$ , independentes e não negativas, com funções de distribuição  $F$ ,  $G_1$  e  $G_2$ , respectivamente, e correspondentes funções de sobrevivência  $S$ ,  $\overline{G}_1$  e  $\overline{G}_2$ . As variáveis aleatórias observáveis são  $(Z, \Delta)$ , onde  $Z = \min\{T, C, D\}$  e  $\Delta = 1$  se  $T \leq \min\{C, D\}$ ,  $\Delta = 0$  se  $C \leq \min\{T, D\}$  e  $\Delta = -1$  se  $D \leq \min\{T, C\}$ . Para que não haja ambiguidade na definição de  $\Delta$ , são adoptadas algumas convenções. Se  $\min\{T, C, D\} = T$ , ainda que  $T = C = D$ , considera-se  $\Delta = 1$ . Se  $\min\{T, C, D\} = C$ , ainda que  $C = D$ , considera-se  $\Delta = 0$ .

Resumindo, o modelo de Koziol-Green parcial é tal que:

1.  $\overline{G}_1(t) = (S(t))^\beta$ ,  $t \geq 0$ , para algum  $\beta > 0$ , donde as funções hazard das variáveis  $T$  e  $C$  são proporcionais (hipótese de Koziol-Green);  $\beta$  pode ser interpretado como parâmetro de censura informativa;
2. a variável  $T$ , correspondente ao tempo de vida, é parcialmente observável, podendo ocorrer censura de dois tipos: informativa, representada pela variável  $C$ , e não-informativa, representada pela variável  $D$ . Deste modo, o que se observa é  $Z = \min\{T, C, D\}$  e

$$\Delta = \begin{cases} 1 & \text{se } T \leq \min\{C, D\} \\ 0 & \text{se } C \leq \min\{T, D\} \\ -1 & \text{se } D \leq \min\{T, C\} \end{cases}$$

Note-se que, pelo facto de a variável  $D$  representar a censura não-informativa, a sua distribuição não inclui quaisquer parâmetros da distribuição do tempo de vida.

Admitindo o modelo de Koziol-Green, Abdushukurov (1984) e Cheng e Lin (1984, 1987) propuseram um estimador não-paramétrico para a função de sobrevivência do tempo de vida. Note-se que, neste caso, temos apenas as variáveis  $Z = \min\{T, C\}$  e

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

Assim, considerando uma amostra de dimensão  $n$ , o estimador de Abdushukurov-Cheng-Lin (ACL) da função de sobrevivência de  $T$  é igual à função de sobrevivência empírica dos  $Z_i$ 's elevada à média do número de mortes observadas, isto é, elevada a  $\frac{1}{n} \sum_{i=1}^n \delta_i$ .

Sejam  $(Z_1, \Delta_1), \dots, (Z_n, \Delta_n)$   $n$  realizações independentes de  $(Z, \Delta)$ ,  $Z_{(1)} \leq \dots \leq Z_{(n)}$  as estatísticas ordinais de  $Z_1, \dots, Z_n$ , e  $\Delta_{[1]}, \dots, \Delta_{[n]}$  os concomitantes valores de  $\Delta$  relativamente aos  $Z_i$ 's ordenados. O estimador não-paramétrico da função de sobrevivência do tempo de vida, proposto por Gather e Pawlitschko (1998) e Csörgö (1998), análogo ao estimador de Abdushukurov-Cheng-Lin, é dado por

$$\hat{S}_{PACL}(t) = (\hat{K}(t))^{\hat{P}_{PACL}}, \quad t \geq 0 \quad (5.37)$$

onde o índice  $PACL$  indica que se trata do estimador ACL adaptado ao caso em que a censura é parcialmente informativa (e não apenas informativa),  $\hat{K}(t)$  é dado por

$$\hat{K}(t) = \prod_{i=1}^n \left( 1 - \frac{\varepsilon_{[i]}}{n - i + 1} \right)^{I[Z_{(i)} \leq t]}, \quad t \geq 0, \quad (5.38)$$

em que  $I[A]$  representa a função indicatriz do acontecimento  $A$ , e  $\hat{P}_{PACL}$ , um possível estimador de  $P = P(\Delta = 1 | \Delta \neq -1) = 1/(1 + \beta)$ , é dado por

$$\hat{P}_{PACL} = \begin{cases} \frac{\sum_{i=1}^n I[\Delta_i=1]}{\sum_{i=1}^n I[\Delta_i \neq -1]} & \text{se } \exists_{i \in \{1, \dots, n\}} : I[\Delta_i \neq -1] \neq 0 \\ 0 & \text{se } \forall_{i \in \{1, \dots, n\}}, I[\Delta_i \neq -1] = 0 \end{cases}$$

Neste contexto,  $\varepsilon_i = I[\Delta_i \neq -1]$  e  $\varepsilon_{[1]}, \dots, \varepsilon_{[n]}$  denotam os valores concomitantes de  $\varepsilon$  relativamente aos  $Z'_i$ s.

Relativamente ao estimador de  $\beta$ , temos

$$\hat{\beta}_{PACL} = \frac{1 - \hat{P}_{PACL}}{\hat{P}_{PACL}},$$

que é o mesmo do que ter

$$\hat{\beta}_{PACL} = \frac{\sum_{i=1}^n I[\Delta_i = 0]}{\sum_{i=1}^n I[\Delta_i = 1]}$$

se  $\exists_{i \in \{1, \dots, n\}} : I[\Delta_i \neq -1] \neq 0$ .

### 5.4.3 Modelo de cura com censura parcialmente informativa

#### Formulação do modelo

Estando num contexto de existência de indivíduos curados, vamos propor uma generalização do modelo anterior. Tal como em 5.4.2, consideremos as variáveis aleatórias  $T$ ,  $C$  e  $D$ , independentes e não negativas, com funções de distribuição  $F$ ,  $G_1$  e  $G_2$ , respectivamente, e correspondentes funções de sobrevivência  $S$ ,  $\bar{G}_1$  e  $\bar{G}_2$ . Admitamos para o tempo de vida,  $T$ , o modelo de cura de não-mistura

$$S(t) = \exp[-\theta G(t)], \quad t \geq 0,$$

onde  $G(t)$  é uma função de distribuição própria de uma variável aleatória não negativa. Admitamos ainda que:

1. pode ocorrer censura de dois tipos: informativa, representada pela variável  $C$ , e não-informativa, representada pela variável  $D$ . Deste modo, o que se observa é  $Z = \min\{T, C, D\}$  e

$$\Delta = \begin{cases} 1 & \text{se } T \leq \min\{C, D\} \\ 0 & \text{se } C \leq \min\{T, D\} \\ -1 & \text{se } D \leq \min\{T, C\} \end{cases}$$

2. a distribuição de  $C$  satisfaz a igualdade que se segue,

$$\overline{G}_1(t) = (S(t))^\beta, t \geq 0$$

para algum  $\beta > 0$ , ou seja, admitimos o modelo de Koziol-Green para a censura informativa; trata-se também de um modelo de cura de não-mistura pois  $\overline{G}_1(t) = \exp[-\beta\theta G(t)]$ ;

3. a distribuição de  $D$  não inclui quaisquer dos parâmetros de interesse, isto é, quaisquer dos parâmetros da distribuição de  $T$  e é uma distribuição própria.

Para que não haja ambiguidade na definição de  $\Delta$ , são adoptadas as mesmas convenções do que no modelo de Koziol-Green parcial. Para efeitos de simplicidade de escrita, vamos designar este modelo de cura com censura parcialmente informativa por MCCPI.

É de salientar que o modelo de Koziol-Green não pode ser directamente generalizado de modo a permitir a existência de indivíduos curados. De facto, existindo apenas censura informativa, teríamos indivíduos com tempos de vida observados infinitos (uma vez que a distribuição da censura seria imprópria) o que, naturalmente, não pode acontecer. Com o modelo de Koziol-Green parcial, esse problema não ocorre, dado que a distribuição de  $D$  é própria. Note-se ainda que estamos a permitir a existência de indivíduos curados quanto à causa de interesse e também quanto a causas com ela relacionadas, pois as distribuições de  $T$  e  $C$  são impróprias.

### Função de verosimilhança

Admitamos que as variáveis aleatórias  $T$ ,  $C$  e  $D$  são contínuas. Consideremos uma amostra de dimensão  $n$  e designemos por  $t_1, \dots, t_n$  os tempos de vida.

Sem perda de generalidade, suponhamos que as observações estão indexadas de modo a que os primeiros  $m_1$  ( $m_1 < n$ ) tempos de vida são os censurados informativamente, os compreendidos entre  $m_1$  e  $m_2$  ( $m_2 < n$ ) correspondem a tempos de vida observados, e os restantes  $n - m_2$  são os censurados não-informativamente. Então,  $\Delta_1, \dots, \Delta_n$  são tais que

$$\Delta = \begin{cases} 0 & \text{se } 1 \leq i \leq m_1 \\ 1 & \text{se } m_1 + 1 \leq i \leq m_2 \\ -1 & \text{se } m_2 + 1 \leq i \leq n \end{cases}$$

Com base nos resultados apresentados na secção 1.7, vamos construir a função de verosimilhança, omitindo os factores dispensáveis em termos da estimação dos parâmetros de interesse. Assim, embora a contribuição para a função de verosimilhança do  $i$ -ésimo indivíduo seja

- $f(t_i)\overline{G}_1(t_i)\overline{G}_2(t_i)$  se  $\Delta_i = 1$ ,
- $S(t_i)g_1(t_i)\overline{G}_2(t_i)$  se  $\Delta_i = 0$ ,
- $S(t_i)\overline{G}_1(t_i)g_2(t_i)$  se  $\Delta_i = -1$

onde  $f$ ,  $g_1$  e  $g_2$  são as funções densidade de  $T$ ,  $C$  e  $D$ , respectivamente, basta-nos considerar

- $f(t_i)\overline{G}_1(t_i)$  se  $\Delta_i = 1$ ,
- $S(t_i)g_1(t_i)$  se  $\Delta_i = 0$ ,
- $S(t_i)\overline{G}_1(t_i)$  se  $\Delta_i = -1$

Então, a função de verosimilhança para o modelo considerado é da forma

$$L = \prod_{\Delta_i=1} f(t_i)\overline{G}_1(t_i) \prod_{\Delta_i=0} S(t_i)g_1(t_i) \prod_{\Delta_i=-1} S(t_i)\overline{G}_1(t_i).$$

Por outro lado, como estamos a admitir o modelo de Koziol-Green para a censura informativa, temos que

$$g_1(t_i) = \beta f(t_i)[S(t_i)]^{\beta-1}.$$

Donde, substituindo, temos

$$L = \prod_{\Delta_i=1} f(t_i)[S(t_i)]^\beta \prod_{\Delta_i=0} \beta f(t_i)[S(t_i)]^\beta \prod_{\Delta_i=-1} [S(t_i)]^{\beta+1}.$$

Notando que  $f(t_i) = h(t_i)S(t_i)$ , temos ainda que

$$L = \prod_{\Delta_i=0} \beta \prod_{\Delta_i \neq -1} h(t_i) \prod_{i=1}^n [S(t_i)]^{\beta+1}.$$

O logaritmo da função de verosimilhança escreve-se na forma

$$\log L = \sum_{\Delta_i=0} \log \beta + \sum_{\Delta_i \neq -1} \log h(t_i) + (\beta + 1) \sum_{i=1}^n \log S(t_i) \quad (5.39)$$

Admitindo um modelo paramétrico para a distribuição do tempo de vida, vamos agora ilustrar a obtenção dos estimadores de máxima verosimilhança dos parâmetros envolvidos, considerando para tal a distribuição de Gompertz modificada. Em seguida, fazemos uma aplicação dos resultados a dados reais.

Uma extensão natural do modelo MCCPI, que não será aqui abordada, consiste na inclusão de covariáveis. Assim sendo, o exemplo apresentado poderá também ser utilizado uma vez que contém informação acerca de covariáveis, tais como, sexo, idade, etc.

### Estimação dos parâmetros

Admitamos que o tempo de vida segue uma distribuição de Gompertz modificada,

$$S(t) = \exp \left[ \frac{\alpha}{\lambda} (1 - e^{\lambda t}) \right], \quad \alpha > 0, \quad \lambda < 0.$$

Então, o logaritmo da função de verosimilhança é

$$\log L = \sum_{\Delta_i=0} \log \beta + \sum_{\Delta_i \neq -1} \log(\alpha e^{\lambda t_i}) + (\beta + 1) \sum_{i=1}^n \frac{\alpha}{\lambda} (1 - e^{\lambda t_i})$$

o qual, após algumas simplificações, escreve-se na forma

$$\log L = m_1 \log \beta + m_2 \log \alpha + \lambda \sum_{\Delta_i \neq -1} t_i + (\beta + 1) \frac{\alpha}{\lambda} \sum_{i=1}^n (1 - e^{\lambda t_i}).$$

Relativamente às primeiras derivadas parciais em ordem aos parâmetros, temos

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= \frac{m_1}{\beta} + \frac{\alpha}{\lambda} \sum_{i=1}^n (1 - e^{\lambda t_i}), \\ \frac{\partial \log L}{\partial \alpha} &= \frac{m_2}{\alpha} + \frac{\beta + 1}{\lambda} \sum_{i=1}^n (1 - e^{\lambda t_i}) \end{aligned}$$

e

$$\frac{\partial \log L}{\partial \lambda} = \sum_{\Delta_i \neq -1} t_i - (\beta + 1) \frac{\alpha}{\lambda} \left[ \frac{1}{\lambda} \sum_{i=1}^n (1 - e^{\lambda t_i}) + \sum_{i=1}^n t_i e^{\lambda t_i} \right]$$

das quais resultam expressões explícitas para os estimadores de máxima verosimilhança de  $\beta$  e de  $\alpha$

$$\begin{aligned} \hat{\beta} &= \frac{-\hat{\lambda} m_1}{\hat{\alpha} \sum_{i=1}^n (1 - e^{\hat{\lambda} t_i})}, \\ \hat{\alpha} &= \frac{-\hat{\lambda} m_2}{(\hat{\beta} + 1) \sum_{i=1}^n (1 - e^{\hat{\lambda} t_i})}, \end{aligned}$$

mas não para  $\lambda$ . Assim, é necessário recorrer a um método numérico para determinar uma estimativa de  $\lambda$ . Como habitualmente, optámos pelo método de N-R, pelo que determinámos também a segunda derivada parcial em ordem a  $\lambda$ . Temos então,

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -(\beta + 1) \frac{\alpha}{\lambda} \sum_{i=1}^n t_i^2 e^{\lambda t_i} + (\beta + 1) \frac{2\alpha}{\lambda^2} \left[ \frac{1}{\lambda} \sum_{i=1}^n (1 - e^{\lambda t_i}) + \sum_{i=1}^n t_i e^{\lambda t_i} \right].$$

Para iniciar o processo de estimação, temos que obter valores iniciais dos parâmetros. Em relação a  $\beta^{(0)}$ , podemos usar o seu estimador não

paramétrico, ou seja,  $\hat{\beta}_{PACL}$ , à semelhança do procedimento que adoptámos em relação à estimativa inicial da proporção de susceptíveis num modelo de cura de mistura. Relativamente aos parâmetros da distribuição, de acordo com a distribuição escolhida, assim será o procedimento adoptado. No caso da Gompertz modificada, não é imediato qual o processo a seguir. Das alternativas possíveis, procurámos uma que nos desse as estimativas de forma bastante célere. Para tal, com base na aproximação  $e^x \approx 1 + x$  a distribuição passa "a ser" exponencial de parâmetro  $\alpha$ , logo podemos utilizar a estimativa de máxima verosimilhança de  $\alpha$  para obter um seu valor inicial. Por outro lado, como a estimativa da probabilidade de cura, neste modelo, é dada por  $\exp(\hat{\alpha}/\hat{\lambda})$ , utilizando a estimativa não paramétrica da probabilidade de cura, deduzimos uma expressão que nos permite obter uma estimativa inicial de  $\lambda$ .

### Uma aplicação

Braekers e Veraverbeke (2001) e Zhang e Rao (2004), no contexto da censura parcialmente informativa, estudaram uma amostra de 205 doentes com melanoma, onde o tempo de vida é definido como o tempo desde a completa remoção do tumor até à morte pela doença. Os dados, que podem ser encontrados em detalhe em Andersen *et al.* (1993), são adequados neste contexto pois, além de existirem doentes que morrem do melanoma (57), também são observadas mortes devido a outras causas (14), que são presumivelmente consequência indirecta do melanoma. O tempo de vida dos indivíduos que morrem das outras causas é uma variável que representa a censura informativa e o tempo de vida dos doentes que permanecem vivos no fim do estudo (134), é uma variável que representa a censura não-informativa.

Note-se que o estimador  $\hat{K}$  definido em (5.38) corresponde, neste caso, ao estimador de Kaplan-Meier calculado com base numa amostra em que as observações não censuradas correspondem a todas as mortes observadas, quer estas sejam devidas à causa de interesse quer às causas com ela relacionadas.

O facto de, a partir de determinado instante (3338 dias), todas as observações serem censuradas não-informativamente, o que é característico de populações com indivíduos curados, levou a que este exemplo se enquadrasse perfeitamente na situação em estudo e, como tal, fosse escolhido para ilustrar a metodologia anterior.

Assim, partindo dos valores iniciais  $\beta^{(0)} = 0.2456140$ ,  $\alpha^{(0)} = 0.0001292154$  e  $\lambda^{(0)} = -0.0002945433$ , chegámos às estimativas  $\hat{\beta} = 0.2455999$  (0.07325815),  $\hat{\alpha} = 0.0001479468$  ( $3.045428e - 05$ ) e  $\hat{\lambda} =$

$-0.0001026668$  ( $0.0001241030$ ), onde os valores entre parêntesis indicam o desvio padrão das estimativas. Os dados e os detalhes do ajustamento constam do anexo D. Com este modelo, obtivemos uma estimativa da probabilidade de cura para o melanoma dada por  $\exp(\hat{\alpha}/\hat{\lambda}) \approx 0.237$ . Relativamente à outra causa de morte, a taxa de cura é estimada por  $\exp(\hat{\beta}\hat{\alpha}/\hat{\lambda}) \approx 0.702$ .

Na Fig. 5.1, encontram-se representadas a estimativa não paramétrica  $\hat{S}_{PACL}(t)$  definida em (5.37) e a estimativa correspondente ao modelo com a distribuição de Gompertz modificada, da função de sobrevivência do tempo até à morte por melanoma. Podemos constatar um bom ajustamento aos dados, excepto para valores onde já não são observadas mortes pela doença.

**Figura 5.1:** Estimativa  $\hat{S}_{PACL}(t)$  (linha quebrada) e estimativa da f.s. correspondente ao modelo proposto.

# Capítulo 6

## Fragilidade no contexto dos modelos de cura

### 6.1 Introdução

No contexto dos modelos de cura, as observações censuradas vão corresponder a indivíduos que constituem um grupo heterogéneo pois este é formado por indivíduos curados e por indivíduos susceptíveis. Como nesse grupo não conseguimos identificar individualmente quais os indivíduos curados e quais os indivíduos susceptíveis, podemos afirmar que estamos perante uma situação em que a heterogeneidade é não observável, logo em condições de poder considerar uma variável fragilidade.

Por outro lado, também é razoável admitirmos que, mesmo entre os indivíduos susceptíveis, haja uma heterogeneidade não observável. Na verdade, constatamos frequentemente que nem todos os indivíduos têm igual grau de susceptibilidade ao acontecimento de interesse. Por isso, dado que esse grau de susceptibilidade não é mensurável, pode ser caracterizado através de uma variável fragilidade. Vamos assim procurar relacionar estes dois conceitos: fragilidade e modelos de cura.

Num tipo de abordagem um pouco diferente da que vamos considerar, Aalen (1988, 1992) admitiu a possibilidade da existência de indivíduos com susceptibilidade nula (indivíduos imunes), a par de outros com grau variável de susceptibilidade. Assim, considerou uma variável fragilidade mista, com massa de probabilidade não nula em zero e com distribuição contínua na semi-recta positiva.

Hougaard *et al.* (1994), no contexto dos modelos com fragilidade e Sy e Taylor (2000) no contexto dos modelos de cura, observaram que a variável  $Y$  que classifica os indivíduos como sendo susceptíveis ( $Y = 1$ ) ou imunes ( $Y = 0$ ) pode ser encarada como uma variável fragilidade, se for adoptado um modelo multiplicativo com fragilidade. Deste modo, é possível definir uma variável fragilidade com características muito particulares: ao contrário do que é habitual, é uma variável discreta e não se trata de uma fragilidade no sentido literal, dado que é parcialmente observável.

Numa abordagem mais geral, Longini e Halloran (1996), no contexto dos modelos com fragilidade, admitem a existência de indivíduos imunes. Assim, para modelar a heterogeneidade individual na susceptibilidade ao acontecimento de interesse, consideram uma variável aleatória mista com massa de probabilidade não nula em zero e uma distribuição contínua para a parte positiva da variável (mais especificamente, a distribuição gama). Price e Manatunga (2001), também no contexto dos modelos de cura com fragilidade apresentam várias simulações baseadas no modelo de Longini e Halloran (1996) e no de Aalen (1992), entre outros. Tsodikov *et al.* (1998), no contexto dos modelos de cura de não-mistura, também fazem referência a uma fragilidade definida como em Aalen (1992). A nossa abordagem não é esta última, uma vez que vamos considerar a fragilidade na classe dos modelos de cura de mistura.

Os modelos que propomos são um pouco mais gerais do que o modelo referido por Sy e Taylor (2000), no sentido em que além de admitirmos que existem indivíduos susceptíveis e não susceptíveis, consideramos que os indivíduos que são susceptíveis não o são em igual grau. Assim sendo, como para cada indivíduo não é possível quantificar o seu grau de susceptibilidade, vamos definir uma variável fragilidade que é nula para os indivíduos imunes e é estritamente positiva para os indivíduos susceptíveis. Deste modo, a variável fragilidade é-o no sentido literal. Além disso, consideramos não só o modelo multiplicativo para a fragilidade como ainda o modelo de tempo de vida acelerado com fragilidade.

## 6.2 Modelo com pseudo-fragilidade

De acordo com Sy e Taylor (2000), o modelo de cura de hazards proporcionais referido na secção 4.2.1 é um caso especial do modelo multiplicativo com fragilidade, no qual a função hazard no instante  $t$  de um indivíduo com

vector de covariáveis  $\mathbf{z}$ , condicional a  $Y$ , pode ser escrita como

$$\mu(t|Y; \mathbf{z}) = Y\lambda(t|Y = 1; \mathbf{z}). \quad (6.1)$$

Esta afirmação continua válida ainda que não haja covariáveis observadas, visto que o resultado proveio da estrutura do modelo de cura e não da hipótese de hazards proporcionais que se admitiu para os indivíduos susceptíveis.

Note-se que  $Y$  não é inteiramente não observável, uma vez que, quando um indivíduo é classificado como sendo susceptível, isto quer dizer que foi observado o acontecimento de interesse, ou seja, observou-se  $Y = 1$ . Assim sendo,  $Y$  não é uma variável fragilidade, no sentido literal. Digamos que é uma pseudo-fragilidade.

Vamos então considerar o modelo

$$\mu(t|Y) = Y\lambda(t|Y = 1), \quad (6.2)$$

onde  $\lambda(t|Y = 1)$  é a função hazard da distribuição log-logística e  $Y$  tem distribuição de Bernoulli com parâmetro  $1 - p$ , ou seja, tem função massa probabilística dada por

$$f(y) = (1 - p)^y p^{1-y}, y = 0, 1. \quad (6.3)$$

Com a ressalva de que  $Y$  não é completamente não observável, o modelo (6.2) tem a mesma forma do que o modelo (2.1), onde  $\lambda(t)$  coincide com  $\lambda(t|Y = 1)$ .

Habitualmente, a fragilidade é representada por uma variável aleatória contínua. No nosso caso,  $Y$  é uma variável aleatória discreta que assume os valores 0 e 1. Assim sendo, o valor médio da variável fragilidade é  $1 - p$  e não 1, como é usual considerar. Além disso,  $\lambda(t)$  nem vai ser uma função comum a todos os indivíduos (mas apenas aos indivíduos susceptíveis), nem vai poder ser encarada como uma função hazard individual *média* visto que não é o valor esperado da função hazard individual  $\mu(t|Y)$ . No entanto, pode ser interpretada como força de mortalidade padrão, como referem Vaupel *et al.* (1979) e Manton *et al.* (1986).

Tendo em atenção que, para  $a = 0, 1$ , temos  $P(Y = a|T \geq t) = \frac{P(Y=a)P(T \geq t|Y=a)}{P(T \geq t)}$ , então por (4.16) e (4.17), a distribuição da fragilidade entre os sobreviventes no instante  $t$  é dada por:

$$P(Y = 1|T \geq t) = \frac{1 - p}{1 + p\lambda t^\alpha}, \quad P(Y = 0|T \geq t) = \frac{p + p\lambda t^\alpha}{1 + p\lambda t^\alpha}.$$

Alternativamente, podemos obter a expressão geral da distribuição da fragilidade entre os sobreviventes no instante  $t$  utilizando os resultados referidos no capítulo 2, embora com as devidas adaptações, visto que  $Y$  é uma variável discreta. Assim, considerando que a variável  $W$  é discreta e que toma  $\tau + 1$  valores no conjunto  $\{w_0, w_1, \dots, w_\tau\}$ ,  $S(t|w)$  tem a mesma expressão. Em relação a (2.3), obviamente que

$$S(t) = \sum_{k=0}^{\tau} S(t|w_k)f(w_k) = \sum_{k=0}^{\tau} \exp[-w_k\Lambda(t)]f(w_k) = L[\Lambda(t)], \quad (6.4)$$

onde  $f$  é a função massa probabilidade de  $W$  e  $L(s)$  é a transformada de Laplace da distribuição da fragilidade  $W$ . No nosso caso, por (4.16),

$$\Lambda(t) = \log(1 + \lambda t^\alpha) \quad (6.5)$$

visto que os indivíduos que têm fragilidade  $w = 1$  são os indivíduos susceptíveis. A transformada de Laplace de uma variável aleatória de Bernoulli com parâmetro  $1 - p$  é dada por  $L(s) = p + (1 - p)e^{-s}$ , logo,

$$L[\Lambda(t)] = p + (1 - p) \exp[-\log(1 + \lambda t^\alpha)] = \frac{1 + p\lambda t^\alpha}{1 + \lambda t^\alpha}. \quad (6.6)$$

É de salientar que (6.6) coincide com (4.17), como seria de esperar.

Com base em (2.5), e usando as igualdades (6.3) e (6.6), temos

$$g(y|T \geq t) = \frac{(1 - p)^y p^{1-y}}{(1 + \lambda t^\alpha)^y} \left( \frac{1 + p\lambda t^\alpha}{1 + \lambda t^\alpha} \right)^{-1},$$

que é equivalente a

$$g(y|T \geq t) = \frac{(1-p)^y p^{1-y} (1 + \lambda t^\alpha)^{1-y}}{1 + p\lambda t^\alpha}. \quad (6.7)$$

Se reescrevermos a expressão anterior na forma

$$g(y|T \geq t) = \left( \frac{1-p}{1+p\lambda t^\alpha} \right)^y \left[ \frac{p(1 + \lambda t^\alpha)}{1 + p\lambda t^\alpha} \right]^{1-y},$$

é imediato observar que a distribuição da fragilidade entre os sobreviventes no instante  $t$  é uma distribuição de Bernoulli de parâmetro  $(1-p)/(1+p\lambda t^\alpha)$ . Assim, a família das distribuições de Bernoulli é fechada para a seleção induzida pela mortalidade, visto que a distribuição da fragilidade em qualquer instante é ainda uma distribuição de Bernoulli, embora com parâmetro diferente. Além disso, o valor do parâmetro decresce com o tempo, o que significa que a probabilidade de ser susceptível entre os sobreviventes vai se tornando cada vez mais pequena.

Temos então que a fragilidade média entre os sobreviventes no instante  $t$  é

$$E(Y|T \geq t) = \frac{1-p}{1+p\lambda t^\alpha}, \quad (6.8)$$

que coincide com (2.7). Por outro lado, a função que representa esta fragilidade média além de ser uma função decrescente em  $t$  (conforme já referido), tem limite zero, isto é, quando  $t \rightarrow \infty$ ,  $E(Y|T \geq t) \rightarrow 0$ . Pela expressão de  $h(t)$  dada em (4.18) e por (6.8), temos

$$h(t) = E(Y|T \geq t)\lambda(t|Y = 1), \quad (6.9)$$

embora  $h(t)$  não represente aqui o valor esperado entre os sobreviventes no instante  $t$  da verdadeira função hazard, uma vez que admitimos o modelo multiplicativo apenas para os indivíduos susceptíveis. Note-se que a distribuição da fragilidade à *nascença* coincide com a distribuição da fragilidade, ou seja, para  $t = 0$ ,  $g(y|T \geq 0) = f(y)$ .

Quanto à distribuição da fragilidade entre os indivíduos que morrem no instante  $t$ , é óbvio que  $P(Y = 1|t) = 1$ , visto que esses indivíduos são necessariamente susceptíveis.

O facto de estarmos a lidar com uma pseudo-fragilidade e de esta ter distribuição de Bernoulli, tem como consequência não só a simplificação no processo de estimação dos parâmetros, como algumas relações interessantes.

Conforme referimos no capítulo 2, ao procedermos à estimação dos parâmetros os valores da fragilidade são substituídos pelos valores esperados condicionais aos dados. Ora neste caso, isso só se aplica às observações censuradas; às restantes atribui-se o próprio valor uma vez que se trata de valores observados.

Por outro lado, como estamos a considerar uma variável de Bernoulli, existe mais do que uma forma de determinar o valor esperado desta variável entre os sobreviventes: tanto podemos usar uma abordagem via modelos com fragilidade como uma abordagem via modelos de cura pois

$$E(Y|T \geq t) = P(Y = 1|T \geq t) = \frac{(1-p)S_d(t)}{S(t)} = \tau_i.$$

### 6.3 Modelo de cura com fragilidade

No contexto de um modelo de cura de mistura, o modelo que agora propomos resulta de considerar que a parte positiva da variável fragilidade, correspondente aos indivíduos susceptíveis, segue uma distribuição contínua. Deste modo a variável fragilidade é-o no sentido literal.

#### 6.3.1 Formulação e resultados gerais

Defina-se a variável fragilidade como uma v.a. não negativa,  $W$ , tal que  $P(W = 0) = p$  e  $P(W > 0) = 1 - p$ . Os indivíduos cuja fragilidade é nula são os indivíduos imunes, enquanto que aqueles que têm fragilidade estritamente positiva são os indivíduos susceptíveis.

Seja  $f_X(\cdot)$  a função densidade de  $W|_{W>0} \equiv X$ , onde  $X$  é uma variável aleatória contínua com suporte na semi-recta positiva. Então,

$$W = \begin{cases} 0 & \text{com probabilidade } p \\ X & \text{com probabilidade } 1 - p \end{cases}$$

ou seja,  $W$  tem distribuição de Bernoulli composta.

Longini e Halloran (1996), num estudo sobre a eficácia da vacina do sarampo, fizeram uso de um modelo semelhante, em que a variável  $X$  segue uma distribuição gama. No entanto, esse modelo não foi formulado no contexto dos modelos de cura de mistura e os autores limitaram-se à obtenção das estimativas de máxima verosimilhança dos parâmetros do modelo.

Por razões de identificabilidade do modelo, considera-se que  $E(X) = 1$ . Além disso, seja  $\text{var}(X) = \gamma$ . Deste modo, obtém-se  $E(W) = 1 - p$  e  $\text{var}(W) = (1 - p)(\gamma + p)$ .

Os coeficientes de variação das distribuições das v.a.'s  $W$  e  $X$  são, respectivamente,  $(\frac{\gamma+p}{1-p})^{1/2}$  e  $\gamma^{1/2}$ . Assim, verifica-se facilmente que, como seria de esperar, a heterogeneidade é maior quando se considera toda a população do que quando se considera apenas os indivíduos susceptíveis, pois  $(\frac{\gamma+p}{1-p})^{1/2} > \gamma^{1/2}$ . Além disso, a heterogeneidade populacional é tanto maior quanto maior for a proporção de imunes na população, uma vez que o coeficiente de variação de  $W$  é uma função crescente de  $p$ .

Os indivíduos susceptíveis constituem também um grupo heterogéneo, cuja susceptibilidade (não mensurável) é representada pela variável  $X$ . Para os indivíduos susceptíveis, considere-se então o modelo multiplicativo

$$\mu_d(t|x) = x\lambda(t),$$

onde  $\lambda(t)$  é uma função do tempo comum aos indivíduos susceptíveis. Se considerarmos que  $E(X) = 1$ , conforme refere Aalen (1988),  $\lambda(t)$  pode ser encarada como uma função hazard individual média visto que é o valor esperado da função hazard individual.

A função de sobrevivência condicional de  $T$  dado  $X = x$  é

$$S_d(t|x) = \exp[-x\Lambda(t)],$$

enquanto que a função de sobrevivência e a função hazard marginais de  $T$  para os indivíduos susceptíveis são dadas, respectivamente, por

$$S_d(t) = L_X[\Lambda(t)]$$

e

$$h_d(t) = -\frac{L'_X[\Lambda(t)]}{L_X[\Lambda(t)]}\lambda(t).$$

Quanto à função de sobrevivência populacional (indivíduos susceptíveis e não susceptíveis), temos

$$S(t) = P(T \geq t|W = 0)P(W = 0) + P(T \geq t|W > 0)P(W > 0). \quad (6.10)$$

A expressão anterior corresponde ao modelo de cura de mistura usual (4.1), uma vez que  $P(T \geq t|W = 0) = 1$  para qualquer  $t$  e que  $P(T \geq t|W > 0) = S_d(t)$ , logo

$$S(t) = p + (1 - p)L_X[\Lambda(t)]. \quad (6.11)$$

Por outro lado, a transformada de Laplace da distribuição de  $W$  é dada por

$$L_W(s) = p + (1 - p)L_X(s). \quad (6.12)$$

Então, uma vez que  $S_d(t) = L_X[\Lambda(t)]$ , por (6.22) e (6.23) conclui-se que  $S(t) = L_W[\Lambda(t)]$ .

Notemos que da relação entre a função de sobrevivência e a função hazard, ou da forma como se escreve a função hazard num modelo de cura de mistura, vem

$$h(t) = -\frac{(1 - p)L'_X[\Lambda(t)]}{S(t)}\lambda(t) = -\frac{(1 - p)L'_X[\Lambda(t)]}{p + (1 - p)L_X[\Lambda(t)]}\lambda(t). \quad (6.13)$$

Consideremos agora a distribuição da fragilidade entre os mortos, no instante  $t$ . Dado que apenas os indivíduos susceptíveis podem morrer, esta distribuição coincide com a distribuição de  $X|_{T=t}$ , ou seja, é dada por

$$g(x|t) = \frac{\mu_d(t|x)S_d(t|x)f_X(x)}{h_d(t)S_d(t)}, \quad (6.14)$$

donde

$$g(x|t) = -\frac{x \exp[-x\Lambda(t)]f_X(x)}{L'_X[\Lambda(t)]}.$$

Uma vez que

$$P(W = 0|T \geq t) = \frac{P(T \geq t|W = 0)P(W = 0)}{S(t)}$$

e que

$$P(W > 0|T \geq t) = \frac{P(T \geq t|W > 0)P(W > 0)}{S(t)},$$

a distribuição de  $W$  entre os sobreviventes no instante  $t$  é definida por

$$W|_{T \geq t} = \begin{cases} 0 & \text{com probabilidade } p/[p + (1-p)L_X[\Lambda(t)]] \\ X_1 & \text{com probabilidade } (1-p)L_X[\Lambda(t)]/[p + (1-p)L_X[\Lambda(t)]] \end{cases}$$

onde  $X_1 \equiv X|_{T \geq t}$ . Note-se que, para  $t = 0$ ,  $P(W = 0|T \geq 0) = p$  e  $P(W > 0|T \geq 0) = 1 - p$ . Além disso,  $P(W = 0|T \geq t)$  é uma função crescente de  $t$ , logo a proporção de indivíduos imunes entre os sobreviventes num dado instante aumenta com o decorrer do tempo. A distribuição da fragilidade entre os sobreviventes susceptíveis, no instante  $t$ , é dada pela expressão  $S_d(t|x)f_X(x)/S_d(t)$ , donde

$$g(x|T \geq t) = \frac{\exp[-x\Lambda(t)]f_X(x)}{L_X[\Lambda(t)]}.$$

O valor médio da fragilidade entre os sobreviventes, no instante  $t$ , pode ser calculado do modo que se segue

$$E(W|T \geq t) = \frac{(1-p)L_X[\Lambda(t)]}{p + (1-p)L_X[\Lambda(t)]}E(X|T \geq t), \quad (6.15)$$

logo como

$$E(X|T \geq t) = -\frac{L'_X[\Lambda(t)]}{L_X[\Lambda(t)]},$$

temos que

$$E(W|T \geq t) = -\frac{(1-p)L'_X[\Lambda(t)]}{p + (1-p)L_X[\Lambda(t)]}. \quad (6.16)$$

Note-se que, para os indivíduos susceptíveis a função hazard populacional pode ser escrita na forma

$$h_d(t) = E(X|T \geq t)\lambda(t),$$

o que está de acordo com os resultados referentes ao modelo multiplicativo com fragilidade usual. Para o modelo de mistura com fragilidade, continua a ser verdade que

$$h(t) = E(W|T \geq t)\lambda(t),$$

como se pode constatar por (6.13) e (6.28). No entanto, não podemos neste caso afirmar que  $h(t)$  seja o valor esperado entre os sobreviventes no instante  $t$  da verdadeira função hazard. Tal facto deve-se a termos assumido o modelo multiplicativo apenas para os indivíduos susceptíveis. Por (6.15), o que podemos concluir é que  $h(t) \leq h_d(t)$ .

### 6.3.2 Distribuição Gaussiana inversa como distribuição para a parte contínua da fragilidade

Várias distribuições têm sido propostas como distribuição para a fragilidade, sendo a facilidade do tratamento matemático um factor importante de decisão. Deste ponto de vista, a distribuição Gaussiana inversa é uma escolha possível. Por outro lado, o facto de o seu coeficiente de variação decrescer com o decorrer do tempo (o que significa que a população sobrevivente torna-se mais homogénea), pode ser uma razão para a escolha desta distribuição em detrimento da distribuição gama, cujo coeficiente de variação é constante (logo o grau de heterogeneidade é constante ao longo do tempo). Além disso, à semelhança do que acontece com a distribuição gama, esta distribuição é fechada para a selecção induzida pela mortalidade, ou seja, a distribuição condicional da variável fragilidade, dada a sobrevivência até determinado instante, pertence ainda à família de distribuições da fragilidade.

Seja  $X$  uma v.a. com distribuição Gaussiana inversa de parâmetros  $\mu$  e  $\theta$ , isto é,  $X \sim GI(\mu, \theta)$ . Então, a correspondente f.d.p. é dada por

$$f_X(x) = (\mu/\pi)^{1/2} \exp[(4\mu\theta)^{1/2}]x^{-3/2} \exp(-\theta x - \mu/x), \quad x > 0,$$

em que  $\theta \geq 0$  e  $\mu > 0$ . O valor médio da distribuição Gaussiana inversa é  $(\mu/\theta)^{1/2}$  para  $\theta > 0$  e não existe para  $\theta = 0$ ; a variância é  $(1/2)\mu^{1/2}\theta^{-3/2}$  para

$\theta > 0$ . Como estamos a admitir que  $E(X) = 1$ , consideramos que  $\mu = \theta$ , para  $\theta > 0$ , logo

$$f_X(x) = (\theta/\pi)^{1/2} \exp(2\theta)x^{-3/2} \exp(-\theta x - \theta/x).$$

Temos ainda que  $\text{var}(X) = (2\theta)^{-1}$  e que o coeficiente de variação é  $(2\theta)^{-1/2}$ .

Quanto à variável  $W$ , como referimos inicialmente, tem valor médio igual a  $1 - p$ . A variância é  $(1 - p)[(2\theta)^{-1} + p]$  e o correspondente coeficiente de variação é  $(1 + 2\theta p)^{1/2}[2\theta(1 - p)]^{-1/2}$ .

Com base nos resultados referidos no capítulo 2,

$$S_d(t) = L_X[\Lambda(t)] = \frac{\phi(-1/2, \theta + \Lambda(t))}{\phi(-1/2, \theta)}$$

e

$$h_d(t) = -\frac{L'_X[\Lambda(t)]}{L_X[\Lambda(t)]} \lambda(t) = \frac{\phi(1/2, \theta + \Lambda(t))}{\phi(-1/2, \theta + \Lambda(t))} \lambda(t).$$

Então, quando consideramos a distribuição Gaussiana inversa com um parâmetro para a fragilidade, a função de sobrevivência e a função hazard dos susceptíveis são dadas, respectivamente, por

$$S_d(t) = \exp[-2[\theta(\theta + \Lambda(t))]^{1/2} + 2\theta],$$

$$h_d(t) = [\theta/(\theta + \Lambda(t))]^{1/2} \lambda(t).$$

Relativamente à função de sobrevivência e à função hazard populacionais temos, respectivamente,

$$S(t) = p + (1 - p) \exp[-2[\theta(\theta + \Lambda(t))]^{1/2} + 2\theta], \quad (6.17)$$

$$h(t) = \frac{(1 - p)[\theta/(\theta + \Lambda(t))]^{1/2}}{p \exp[2[\theta(\theta + \Lambda(t))]^{1/2} - 2\theta] + (1 - p)} \lambda(t). \quad (6.18)$$

Price e Manatunga (2001), num estudo sobre o tempo de vida de pacientes com leucemia, referem os modelos de mistura com fragilidade (do tipo (6.11)). Trata-se essencialmente de uma aplicação deste tipo de modelos onde, para

a distribuição da fragilidade dos indivíduos susceptíveis, os autores indicam duas possibilidades (as mais usuais para a fragilidade): a distribuição gama e a distribuição Gaussiana inversa. No entanto, relativamente ao modelo com a distribuição Gaussiana inversa, apresentam apenas a expressão da função de sobrevivência (para além dos resultados da estimação), ou seja, a expressão (6.17).

Por (6.14), a distribuição da fragilidade entre os mortos, no instante  $t$ , é dada por

$$g(x|t) = \frac{x\lambda(t) \exp[-x\Lambda(t)](\theta/\pi)^{1/2} \exp(2\theta)x^{-3/2} \exp(-\theta x - \theta/x)}{[\theta/(\theta + \Lambda(t))]^{1/2}\lambda(t) \exp[-2[\theta(\theta + \Lambda(t))]^{1/2} + 2\theta]},$$

ou seja, simplificando

$$g(x|t) = \frac{x^{-1/2}[(\theta + \Lambda(t))/\pi]^{1/2} \exp[-(\theta + \Lambda(t))x - \theta/x]}{\exp[-2[\theta(\theta + \Lambda(t))]^{1/2}]},$$

logo obtemos a f.d.p. da distribuição Gaussiana inversa recíproca,  $GIG(1/2, \theta, \theta + \Lambda(t))$ .

A distribuição da fragilidade entre os sobreviventes susceptíveis, no instante  $t$ , é  $D(-1/2, \theta + \Lambda(t))$ , ou seja,

$$g(x|T \geq t) = (\theta/\pi)^{1/2} \exp[2[\theta(\theta + \Lambda(t))]^{1/2}]x^{-3/2} \exp[-(\theta + \Lambda(t))x - \theta/x],$$

que também é Gaussiana inversa, mas com parâmetros  $\theta$  e  $\theta + \Lambda(t)$ . Então,  $X|_{T \geq t} \sim GIG(-1/2, \theta, \theta + \Lambda(t))$  logo  $X|_{T=t}$  designa-se por recíproca complementar de  $X|_{T \geq t}$ .

A fragilidade média entre os sobreviventes susceptíveis, no instante  $t$ , é  $E(X|T \geq t) = [\theta/[\theta + \Lambda(t)]]^{1/2}$ , donde a fragilidade média entre os sobreviventes, no instante  $t$ , é

$$E(W|T \geq t) = \frac{(1-p)[\theta/(\theta + \Lambda(t))]^{1/2}}{p \exp[2[\theta(\theta + \Lambda(t))]^{1/2} - 2\theta] + (1-p)},$$

o que está de acordo com a expressão (6.32).

### 6.3.3 Estimação dos parâmetros e aplicação do algoritmo EM

Consideremos o modelo

$$S(t) = p + (1 - p)L_X[\Lambda(t)],$$

onde  $L_X[\Lambda(t)] (= S_d(t))$  representa a transformada de Laplace da distribuição de  $X$ , sendo  $X$  uma v.a. contínua. Vejamos como, através do algoritmo EM, se podem estimar os vários parâmetros envolvidos neste modelo:  $p$  (ou  $q = 1 - p$ ), o(s) parâmetro(s) da distribuição do tempo de vida dos susceptíveis e o(s) parâmetro(s) da distribuição da fragilidade.

Consideremos uma amostra de dimensão  $n$  e designemos por  $t_1, \dots, t_n$  os tempos de vida e por  $w_1, \dots, w_n$  as fragilidades associadas. Note-se que  $w_i = x_i$ ,  $i = 1, \dots, n$  sempre que  $w_i > 0$ . Sem perda de generalidade, suponhamos que os primeiros  $m$  ( $m < n$ ) tempos de vida são censurados. Sejam  $\delta_1, \dots, \delta_n$  tais que

$$\delta_i = \begin{cases} 0 & \text{se } 1 \leq i \leq m \\ 1 & \text{se } m + 1 \leq i \leq n \end{cases}$$

e  $y_1, \dots, y_n$  tais que

$$y_i = \begin{cases} 0 & \text{se o indivíduo é imune} \\ 1 & \text{se o indivíduo é susceptível} \end{cases}$$

No modelo de cura sem fragilidade, se todos os  $y_i$ 's fossem observados, a verosimilhança completa seria

$$\begin{aligned} L_C &= \prod_{i=1}^n [(1 - p)f_d(t_i)]^{y_i \delta_i} [p^{1-y_i} [(1 - p)S_d(t_i)]^{y_i}]^{1-\delta_i} \\ &= \prod_{i=1}^n (1 - p)^{y_i} p^{1-y_i} \prod_{i=1}^n h_d(t_i)^{y_i \delta_i} S_d(t_i)^{y_i} \\ &= \prod_{i=1}^n q^{y_i} (1 - q)^{1-y_i} \prod_{i=1}^n h_d(t_i)^{y_i \delta_i} S_d(t_i)^{y_i} \end{aligned}$$

pois  $q = 1 - p$ .

Por outro lado, no modelo multiplicativo com fragilidade usual, a função de verosimilhança conjunta para  $T$  e  $W$  é

$$L = \prod_{i=1}^n f(t_i, w_i)^{\delta_i} S(t_i, w_i)^{1-\delta_i},$$

onde  $S(t_i, w_i) = P(T \geq t_i, W = w_i)$ , a qual pode ser factorizada na forma

$$L_C = \prod_{i=1}^n f(w_i) \prod_{i=1}^n \mu(t_i|w_i)^{\delta_i} S(t_i|w_i),$$

que designámos por verosimilhança completa dado ser construída como se os valores da fragilidade fossem observados.

Assim, a função de verosimilhança correspondente ao modelo de cura com fragilidade é

$$\begin{aligned} L_C &= \prod_{i=1}^n [[qf_d(t_i, x_i)]^{y_i}]^{\delta_i} [(1-q)^{1-y_i} [qS_d(t_i, x_i)]^{y_i}]^{1-\delta_i} \\ &= \prod_{i=1}^n q^{y_i} (1-q)^{1-y_i} \prod_{i=1}^n f_X(x_i)^{y_i} \prod_{i=1}^n \mu_d(t_i|x_i)^{y_i \delta_i} S_d(t_i|x_i)^{y_i} \quad (6.19) \\ &= L_{C_Y} L_{C_{XT}} \end{aligned}$$

e o correspondente logaritmo é

$$\begin{aligned} \log L_C &= \sum_{i=1}^n [y_i \log q + (1-y_i) \log(1-q)] + \\ &\quad \sum_{i=1}^n y_i \log f_X(x_i) + \sum_{i=1}^n [y_i \delta_i \log \mu_d(t_i|x_i) + y_i \log S_d(t_i|x_i)] \end{aligned}$$

Considerando o logaritmo de  $L_{C_Y}$ , com base nos resultados do capítulo 4, ao substituir cada um dos  $y_i$ 's não observados pelo seu valor esperado, obtemos

$$\log L_{E_Y} = \sum_{i=1}^n [\omega_i \log q + (1 - \omega_i) \log(1 - q)], \quad (6.20)$$

onde

$$\omega_i = \begin{cases} 1 & \text{se } \delta_i = 1 \\ \tau_i & \text{se } \delta_i = 0 \end{cases}$$

e

$$\tau_i = E(Y|\mathcal{O}) = P(Y_i = 1|T > t_i, \delta_i = 0) = \frac{qS_d(t_i)}{S(t_i)}.$$

Então, a expressão do estimador de  $q$  na  $(k + 1)$ -ésima iteração é

$$q^{(k+1)} = \frac{1}{n} \left[ \sum_{i=1}^m \tau_i^{(k)} + (n - m) \right]. \quad (6.21)$$

Para podermos obter  $\tau_i$ , temos que começar por escolher uma distribuição quer para o tempo de vida dos indivíduos susceptíveis quer para a fragilidade desses mesmos indivíduos. No primeiro caso, optámos pela distribuição exponencial devido à sua simplicidade, isto é, considerámos  $\lambda(t) = \lambda$ . No que concerne à distribuição para a fragilidade, escolhemos a distribuição Gaussiana inversa de parâmetro  $\theta$ ,  $GI(\theta, \theta)$ , pelas razões já referidas.

Assim,

$$S_d(t) = \exp[-2[\theta(\theta + \lambda t)]^{1/2} + 2\theta],$$

e

$$S(t) = 1 - q + q \exp[-2[\theta(\theta + \lambda t)]^{1/2} + 2\theta].$$

Então, para  $i = 1, \dots, m$ , temos

$$\tau_i = \frac{q \exp[-2[\theta(\theta + \lambda t_i)]^{1/2} + 2\theta]}{1 - q + q \exp[-2[\theta(\theta + \lambda t_i)]^{1/2} + 2\theta]}, \quad (6.22)$$

donde, por (6.21), o estimador de  $q$  em cada iteração é

$$\hat{q} = \frac{1}{n} \left[ \sum_{i=1}^m \frac{q \exp[-2[\theta(\theta + \lambda t_i)]^{1/2} + 2\theta]}{1 - q + q \exp[-2[\theta(\theta + \lambda t_i)]^{1/2} + 2\theta]} + (n - m) \right]. \quad (6.23)$$

Relativamente a  $\log L_{C_{XT}}$ , temos

$$\begin{aligned} \log L_{C_{XT}} = & \sum_{i=1}^n y_i \log \left[ (\theta/\pi)^{1/2} \exp(2\theta) x_i^{-3/2} \exp(-\theta x_i - \theta/x_i) \right] + \\ & \sum_{i=1}^n [y_i \delta_i \log(x_i \lambda) + y_i \log(\exp(-x_i \lambda t_i))] \end{aligned}$$

Simplificando,

$$\begin{aligned} \log L_{C_{XT}} = & \sum_{i=1}^n y_i \left[ \frac{1}{2} (\log \theta - \log \pi) + 2\theta + (\delta_i - \frac{3}{2}) \log x_i - \theta x_i - \theta/x_i \right] + \\ & \sum_{i=1}^n y_i (\delta_i \log \lambda - x_i \lambda t_i) \\ = & \log L_{C_X} + \log L_{C_T}. \end{aligned} \quad (6.24)$$

Notemos que (6.24) é uma função linear em  $y_i \log x_i$ ,  $y_i x_i$  e  $y_i \frac{1}{x_i}$ . Admitamos a independência entre as variáveis  $X$  e  $Y$ . Sejam  $(x_i)^* = E(X_i | \mathcal{O})$ ,  $(\frac{1}{x_i})^* = E(\frac{1}{X_i} | \mathcal{O})$  e  $(\log x_i)^* = E(\log X_i | \mathcal{O})$ . Ao substituir em (6.24) cada um dos valores não observados pelo correspondente valor esperado, seguindo no essencial a abordagem referida em Klein e Moeschberger (1998), p. 410-413, obtemos

$$\begin{aligned}
 \log L_{EXT} &= \sum_{i=1}^n \omega_i \left[ \frac{1}{2}(\log \theta - \log \pi) + 2\theta + (\delta_i - \frac{3}{2})(\log x_i)^* - \theta(x_i)^* - \theta(\frac{1}{x_i})^* \right] \\
 &\quad + \sum_{i=1}^n \omega_i (\delta_i \log \lambda - (x_i)^* \lambda t_i) \\
 &= \log L_{EX} + \log L_{ET}.
 \end{aligned} \tag{6.25}$$

No passo M do algoritmo EM, vamos maximizar  $\log L_{EXT}$  com respeito a  $\theta$  e a  $\lambda$ , considerando cada valor médio condicional como um valor fixo. Portanto, embora as expressões de  $\omega_i$ ,  $(x_i)^*$  e  $(\frac{1}{x_i})^*$  incluam os parâmetros  $\theta$  e  $\lambda$ , para efeitos de derivação, esse facto não é tido em conta, uma vez que os valores  $\omega_i$ ,  $(x_i)^*$  e  $(\frac{1}{x_i})^*$  são encarados como constantes, seja qual for a sua expressão. Assim, temos

$$\frac{\partial \log L_{EX}}{\partial \theta} = \sum_{i=1}^n \omega_i \left( \frac{1}{2\theta} + 2 - (x_i)^* - (\frac{1}{x_i})^* \right)$$

e

$$\frac{\partial \log L_{ET}}{\partial \lambda} = \sum_{i=1}^n \omega_i \left( \frac{\delta_i}{\lambda} - (x_i)^* t_i \right),$$

donde

$$\hat{\theta} = \frac{\sum_{i=1}^n \omega_i}{2 \left( \sum_{i=1}^n \omega_i (x_i)^* + \sum_{i=1}^n \omega_i (\frac{1}{x_i})^* - 2 \sum_{i=1}^n \omega_i \right)} \tag{6.26}$$

e

$$\hat{\lambda} = \frac{\sum_{i=1}^n \omega_i \delta_i}{\sum_{i=1}^n \omega_i (x_i)^* t_i}. \tag{6.27}$$

É agora necessário determinar os valores esperados condicionais das variáveis  $X$  e  $(1/X)$ , o que faremos com recurso às propriedades da distribuição gaussiana inversa generalizada, referidas na subsecção 2.1.1., de modo a podermos obter as expressões explícitas de  $(x_i)^*$  e  $(\frac{1}{x_i})^*$ . Assim, como  $X|_{T=t} \sim GIG(1/2, \theta, \theta + \lambda t)$  e  $X|_{T \geq t} \sim GIG(-1/2, \theta, \theta + \lambda t)$ , temos

$$\begin{aligned} E(X|T = t) &= \frac{1 + [\theta(\theta + \lambda t)]^{1/2}}{\theta + \lambda t}, & E(X|T \geq t) &= \left(\frac{\theta}{\theta + \lambda t}\right)^{1/2}, \\ E\left(\frac{1}{X}|T = t\right) &= \left(\frac{\theta + \lambda t}{\theta}\right)^{1/2}, & E\left(\frac{1}{X}|T \geq t\right) &= \frac{1 + [\theta(\theta + \lambda t)]^{1/2}}{\theta}. \end{aligned}$$

Vejamos então a que irá corresponder  $\sum_{i=1}^n \omega_i (x_i)^*$ . Notando que

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^m \tau_i + (n - m) \quad (6.28)$$

e definindo

$$E(X_i|\mathcal{O}) = \begin{cases} E(X_i|T = t_i) & \text{se } \delta_i = 1 \\ E(X_i|T \geq t_i) & \text{se } \delta_i = 0 \end{cases}$$

temos

$$\sum_{i=1}^n \omega_i E(X_i|\mathcal{O}) = \sum_{i=1}^m \tau_i E(X_i|T \geq t_i) + \sum_{i=m+1}^n E(X_i|T = t_i).$$

Usando os resultados relativos ao valor médio, obtemos

$$\sum_{i=1}^n \omega_i E(X_i|\mathcal{O}) = \sum_{i=1}^m \tau_i \left(\frac{\theta}{\theta + \lambda t_i}\right)^{1/2} + \sum_{i=m+1}^n \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta + \lambda t_i}. \quad (6.29)$$

Por último, falta-nos considerar  $\sum_{i=1}^n \omega_i (\frac{1}{x_i})^*$ . Ora,

$$\sum_{i=1}^n \omega_i E\left(\frac{1}{X_i} \mid \mathcal{O}\right) = \sum_{i=1}^m \tau_i E\left(\frac{1}{X_i} \mid T \geq t_i\right) + \sum_{i=m+1}^n E\left(\frac{1}{X_i} \mid T = t_i\right)$$

donde, novamente fazendo uso dos resultados relativos ao valor médio, obtemos

$$\sum_{i=1}^n \omega_i E\left(\frac{1}{X_i} \mid \mathcal{O}\right) = \sum_{i=1}^m \tau_i \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta} + \sum_{i=m+1}^n \left(1 + \frac{\lambda t_i}{\theta}\right)^{1/2}. \quad (6.30)$$

Por (6.28), (6.29) e (6.30), a expressão para o estimador de  $\theta$  usada em cada iteração é dada por

$$\hat{\theta} = \frac{\frac{1}{2} \left[ \sum_{i=1}^m \tau_i + (n-m) \right]}{\sum_{i=1}^m \tau_i \left[ \left(\frac{\theta}{\theta + \lambda t_i}\right)^{1/2} + \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta} - 2 \right] + \sum_{i=m+1}^n \left[ \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta + \lambda t_i} + \left(1 + \frac{\lambda t_i}{\theta}\right)^{1/2} \right] - 2(n-m)}, \quad (6.31)$$

Quanto a (6.27), vamos seguir o mesmo tipo de raciocínio. Assim, a  $\sum_{i=1}^n \omega_i \delta_i$  irá corresponder

$$\sum_{i=1}^n \omega_i \delta_i = \sum_{i=m+1}^n \omega_i = n - m, \quad (6.32)$$

enquanto que  $\sum_{i=1}^n \omega_i (x_i)^* t_i$  se escreverá na forma

$$\sum_{i=1}^n \omega_i E(X_i \mid \mathcal{O}) t_i = \sum_{i=1}^m \tau_i E(X_i \mid T \geq t_i) t_i + \sum_{i=m+1}^n E(X_i \mid T = t_i) t_i.$$

Então, de modo análogo a (6.29), obtemos

$$\sum_{i=1}^n \omega_i E(X_i \mid \mathcal{O}) t_i = \sum_{i=1}^m \tau_i t_i \left(\frac{\theta}{\theta + \lambda t_i}\right)^{1/2} + \sum_{i=m+1}^n \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta + \lambda t_i} t_i. \quad (6.33)$$

Por (6.32) e (6.33), a expressão para o estimador de  $\lambda$  usado em cada iteração é

$$\hat{\lambda} = \frac{n - m}{\sum_{i=1}^m \tau_i t_i \left(\frac{\theta}{\theta + \lambda t_i}\right)^{1/2} + \sum_{i=m+1}^n \frac{1 + [\theta(\theta + \lambda t_i)]^{1/2}}{\theta + \lambda t_i} t_i}. \quad (6.34)$$

No anexo E, implementámos um algoritmo EM para o cálculo das estimativas destes parâmetros, com recurso a dados provenientes de Maller e Zhou, (1996, p. 82), já referidos neste trabalho (p. 112), mas considerando agora apenas para o grupo 1. Escolhemos estes dados em particular, pelo facto de, dado o nivelamento da estimativa de Kaplan-Meier da função de sobrevivência, podermos suspeitar da existência de indivíduos curados.

Para além da expressão dos estimadores (6.23), (6.31) e (6.34), usámos o logaritmo da função de verosimilhança esperada ( $\log L_{E_Y} + \log L_{E_X} + \log L_{E_T}$ ), ou seja, aquela que obteríamos se todos os  $y_i$ 's,  $x_i$ 's,  $(1/x_i)$ 's e  $\log x_i$ 's fossem observados, de modo a controlar a convergência do algoritmo. Substituindo em (6.20),  $\omega_i$  pelo respectivo valor, obtemos

$$\log L_{E_Y} = \left[ -m + \sum_{i=1}^m \tau_i \right] \log \frac{q}{1 - q} + n \log q.$$

No que diz respeito a  $\theta$  e a  $\lambda$ , é necessário proceder ao cálculo prévio de  $E(\log X|T = t)$  e de  $E(\log X|T \geq t)$ . Por Jørgensen (1982), sabemos que, quando  $X \sim GIG(\lambda, \mu, \theta)$ ,

$$E(\log X) = \log \eta + \frac{\partial}{\partial \lambda} \log K_\lambda(\zeta),$$

onde  $\eta = (\mu/\theta)^{1/2}$ ,  $\zeta = (\mu\theta)^{1/2}$  e  $K_\lambda(\zeta)$  designa a função de Bessel modificada de terceira espécie com índice  $\lambda \in R$ . Para  $\zeta$  suficientemente grande, é válida uma expansão assintótica para  $K_\lambda(\zeta)$ , donde temos a seguinte aproximação de primeira ordem:

$$E(\log X) = \log \eta + \frac{\lambda}{\eta}.$$

Então, como  $X|_{T=t} \sim GIG(1/2, \theta, \theta + \lambda t)$  e  $X|_{T \geq t} \sim GIG(-1/2, \theta, \theta + \lambda t)$ , temos

$$E(\log X|T = t) = \frac{1}{2} [\log \theta - \log(\theta + \lambda t) + [\theta(\theta + \lambda t)]^{-1/2}]$$

e

$$E(\log X|T \geq t) = \frac{1}{2} [\log \theta - \log(\theta + \lambda t) - [\theta(\theta + \lambda t)]^{-1/2}]$$

donde, por (6.25) e após algumas simplificações, obtemos

$$\begin{aligned} \log L_{E_{XT}} = & \\ & \sum_{i=1}^m \left[ -\frac{1}{4} \log \theta - \frac{1}{2} \log \pi + 2\theta - 1 + \frac{3}{4} \log(\theta + \lambda t) - \theta \left(\frac{\theta}{\theta + \lambda t}\right)^{1/2} - [\theta(\theta + \lambda t)]^{1/2} + \right. \\ & \left. \frac{3}{4} [\theta(\theta + \lambda t)]^{-1/2} - \frac{1 + [\theta(\theta + \lambda t)]^{1/2}}{\theta + \lambda t} \right] + (n - m) \left( \frac{1}{4} \log \theta - \frac{1}{2} \log \pi + 2\theta + \log \lambda - 1 \right) + \\ & \sum_{i=m+1}^n \left[ \frac{1}{4} \log(\theta + \lambda t) - \frac{1}{4} [\theta(\theta + \lambda t)]^{-1/2} - \theta \left(\frac{\theta + \lambda t}{\theta}\right)^{1/2} - [\theta(\theta + \lambda t)]^{1/2} \right]. \end{aligned}$$

Ao aplicar o algoritmo EM aos dados, necessitamos de valores iniciais para os parâmetros  $q$ ,  $\lambda$  e  $\theta$ . Para  $q$ , de acordo com Maller e Zhou (1996), consideramos  $q^{(0)} = 1 - \hat{S}(t_{(r)})$ , onde  $t_{(r)}$  é a maior observação não censurada e  $\hat{S}(t)$  é a estimativa de Kaplan-Meier da função de sobrevivência. Para  $\lambda$ , optamos por ajustar uma distribuição exponencial mas apenas com os dados até  $t_{(r)}$ . Assim,  $q^{(0)} = 0.736622$ ,  $\lambda^{(0)} = 0.472837$  e  $\theta^{(0)} = 1$ , e os resultados a que chegamos foram  $\hat{q} = 0.9708967$ ,  $\hat{\lambda} = 0.5606802$  e  $\hat{\theta} = 0.03277456$ .

O exemplo considerado foi incluído apenas para ilustrar a metodologia apresentada. Como tal, não é de estranhar que o ajustamento deste modelo aos dados não possa ser considerado um bom ajustamento, eventualmente devido à inclusão da variável fragilidade ou mesmo devido à escolha da distribuição para o tempo de vida.

## 6.4 Modelo de tempo de vida acelerado

Seja  $X$  uma v.a. discreta estritamente positiva, que assume os valores  $\{x_1, \dots, x_\tau\}$ . Para simplificar a notação, vamos admitir que os acontecimentos são equiprováveis, logo a f.m.p. de  $X$  é apenas  $f_X(x_k) = 1/\tau$ ,  $k = 1, \dots, \tau$ . No entanto, a passagem dos resultados que vamos apresentar para o caso geral, não apresenta qualquer dificuldade.

Defina-se a variável fragilidade como uma v.a. não negativa,  $W$ , tal que  $P(W = 0) = p$  e  $P(W = x_k) = (1 - p)/\tau$ ,  $k = 1, \dots, \tau$ . Tal como na secção anterior, os indivíduos imunes são aqueles que têm fragilidade nula, enquanto que os susceptíveis são aqueles que têm fragilidade estritamente positiva. Note-se que para cada indivíduo não é possível observar qual o  $x_k$  que lhe corresponde.

Por razões de identificabilidade do modelo, considera-se que  $E(X) = 1$ , logo  $E(W) = 1 - p$ . Além disso, como  $\text{var}(X) = E(X^2) - 1$ , temos  $\text{var}(W) = (1 - p)[E(X^2) - (1 - p)]$ .

Os coeficientes de variação das distribuições das v.a.'s  $W$  e  $X$  são, respectivamente,  $[E(X^2)/(1 - p) - 1]^{1/2}$  e  $[E(X^2) - 1]^{1/2}$ . Como  $[E(X^2)/(1 - p) - 1]^{1/2} > [E(X^2) - 1]^{1/2}$ , concluímos que a heterogeneidade é maior quando se considera toda a população do que quando se considera apenas os indivíduos susceptíveis, como seria de esperar. Ainda através dos coeficientes de variação, observamos que a heterogeneidade populacional é tanto maior quanto maior for a proporção de imunes na população, uma vez que o coeficiente de variação de  $W$  é uma função crescente de  $p$ .

Para os indivíduos susceptíveis, considere-se o modelo de tempo de vida acelerado com fragilidade

$$\mu_a(t|x) = x\lambda(xt),$$

onde  $\lambda(t)$  é uma função do tempo comum aos indivíduos susceptíveis.

A função de sobrevivência condicional de  $T$  dado  $X = x$  é

$$S_a(t|x) = \exp[-\Lambda(xt)],$$

enquanto que a função de sobrevivência e a função hazard marginais de  $T$  para os indivíduos susceptíveis são dadas, respectivamente, por

$$S_d(t) = \sum_{k=1}^{\tau} S_d(t|x_k) f_X(x_k) = \frac{1}{\tau} \sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)]$$

e

$$h_d(t) = \frac{\sum_{k=1}^{\tau} \mu_d(t|x_k) S_d(t|x_k) f_X(x_k)}{\sum_{k=1}^{\tau} S_d(t|x_k) f_X(x_k)} = \frac{\sum_{k=1}^{\tau} x_k \lambda(x_k t) \exp[-\Lambda(x_k t)]}{\sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)]}.$$

Quanto à função de sobrevivência populacional (indivíduos susceptíveis e não susceptíveis), notando que

$$S(t) = P(T \geq t | W = 0)P(W = 0) + P(T \geq t | W > 0)P(W > 0),$$

temos

$$S(t) = p + (1 - p)S_d(t),$$

ou seja,

$$S(t) = p + \frac{(1 - p)}{\tau} \sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)].$$

Da relação entre a função de sobrevivência e a função hazard, obtém-se

$$h(t) = \frac{(1 - p) \sum_{k=1}^{\tau} \mu_d(t|x_k) S_d(t)}{p + (1 - p)S_d(t)},$$

logo

$$h(t) = \frac{(1 - p) \sum_{k=1}^{\tau} x_k \lambda(x_k t) \exp[-\Lambda(x_k t)]}{p\tau + (1 - p) \sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)]}.$$

Consideremos agora a distribuição da fragilidade entre os mortos, no instante  $t$ . Dado que apenas os indivíduos susceptíveis podem morrer, esta distribuição coincide com a distribuição de  $X|_{T=t}$ , ou seja, é dada por

$$g(x|t) = \frac{\mu_d(t|x)S_d(t|x)f_X(x)}{h_d(t)S_d(t)},$$

donde

$$g(x|t) = \frac{x\lambda(xt)\exp[-\Lambda(xt)]}{\sum_{k=1}^{\tau} x_k\lambda(x_k t)\exp[-\Lambda(x_k t)]}.$$

A distribuição de  $W$  entre os sobreviventes no instante  $t$  é definida por

$$P(W > 0|T \geq t) = \frac{(1-p)S_d(t)}{S(t)} = \frac{(1-p)\sum_{k=1}^{\tau}\exp[-\Lambda(x_k t)]}{p\tau + (1-p)\sum_{k=1}^{\tau}\exp[-\Lambda(x_k t)]}$$

e por

$$P(W = 0|T \geq t) = \frac{p}{S(t)} = \frac{p\tau}{p\tau + (1-p)\sum_{k=1}^{\tau}\exp[-\Lambda(x_k t)]}.$$

Note-se que também neste caso, para  $t = 0$ ,  $P(W = 0|T \geq 0) = p$  e  $P(W > 0|T \geq 0) = 1 - p$ . Além disso,  $P(W = 0|T \geq t)$  continua a ser uma função crescente de  $t$ .

A distribuição da fragilidade entre os sobreviventes susceptíveis, no instante  $t$ , é dada por

$$g(x|T \geq t) = \frac{S_d(t|x)f_X(x)}{S_d(t)} = \frac{\exp[-\Lambda(xt)]}{\sum_{k=1}^{\tau}\exp[-\Lambda(x_k t)]},$$

donde o valor médio da fragilidade entre os sobreviventes susceptíveis, no instante  $t$ , escreve-se na forma

$$E(X|T \geq t) = \frac{\sum_{k=1}^{\tau} x_k \exp[-\Lambda(x_k t)]}{\sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)]}.$$

Como o valor médio entre os sobreviventes, no instante  $t$ , calcula-se do modo que se segue

$$E(W|T \geq t) = \frac{(1-p)S_d(t)}{p + (1-p)S_d(t)} E(X|T \geq t),$$

resulta que

$$E(W|T \geq t) = \frac{(1-p) \sum_{k=1}^{\tau} x_k \exp[-\Lambda(x_k t)]}{p\tau + (1-p) \sum_{k=1}^{\tau} \exp[-\Lambda(x_k t)]}.$$

Note-se que, para os indivíduos susceptíveis a função hazard populacional pode ser escrita na forma

$$h_d(t) = E(X\lambda(Xt)|T \geq t),$$

o que está de acordo com os resultados referentes ao modelo de tempo de vida acelerado com fragilidade usual.



# ANEXOS



# Anexo A

## Algoritmo EM

### A.1 Introdução

O algoritmo EM, proposto por Dempster *et al.* (1977), também designado por algoritmo de Maximização de Expectativas (Cabral, 1994), é um método iterativo para encontrar as estimativas de máxima verossimilhança dos parâmetros do modelo em situações em que existem observações omis-sas, ou seja, dados incompletos. Talvez devido à relativa facilidade de im-plementação e à sua estabilidade numérica, no sentido em que cada iteração aumenta o valor da função de verossimilhança, goza de uma enorme popula-ridade. Meng e Pedlow (1992) efectuaram uma pesquisa bibliográfica sobre este algoritmo de 1977 a 1991 e encontraram mais de 1000 artigos em quase 300 revistas, quer da área estatística quer de outras áreas. Da vasta gama de aplicações encontrada nestes artigos, temos como exemplos a medicina, a genética, a engenharia, a psicologia e a economia.

Sem dúvida que o aparecimento deste algoritmo fez desenvolver cer-tas áreas da Estatística, nomeadamente a análise de mistura finita de dis-tribuições, que acaba por englobar parte dos modelos de cura e parte dos modelos com riscos competitivos, no que refere à área da inferência.

Neste anexo, vamos apresentar resumidamente o algoritmo EM e algumas das suas variantes, pois precisaremos destes resultados neste trabalho.

## A.2 Formulação

Seja  $\mathbf{X}$  um vector aleatório que representa os dados observados  $\mathbf{x}$ , ou seja, os dados incompletos, com f.d.p.  $g(\mathbf{x}; \Psi)$ , onde  $\Psi = (\Psi_1, \dots, \Psi_d)^T$  é o vector dos parâmetros desconhecidos, pertencentes ao espaço de parâmetros  $\Omega$ . Sejam  $\mathbf{y}$  o vector dos dados completos e  $\mathbf{z}$  o vector que contém os dados não observados. Designemos por  $g_c(\mathbf{y}; \Psi)$  a f.d.p. do vector aleatório  $\mathbf{Y}$  correspondente ao vector  $\mathbf{y}$ .

O logaritmo da função de verosimilhança para  $\Psi$  construído à custa dos dados observados é

$$\log L(\Psi) = l(\Psi) = \log g(\mathbf{x}; \Psi), \quad (\text{A.1})$$

e o logaritmo da função de verosimilhança para  $\Psi$  que se obtém se  $\mathbf{y}$  fosse observado escreve-se na forma

$$\log L_c(\Psi) = l_c(\Psi) = \log g_c(\mathbf{y}; \Psi). \quad (\text{A.2})$$

Em termos gerais, o objectivo na utilização do algoritmo EM é determinar o EMV de (A.1) à custa de (A.2). Os dados que não são observados são substituídos pela sua esperança condicional dado  $\mathbf{x}$ , usando o valor actual de  $\Psi$ .

Mais concretamente, seja  $\Psi^{(0)}$  um valor inicial para  $\Psi$ . Na primeira iteração, na etapa E procede-se ao cálculo de

$$Q(\Psi; \Psi^{(0)}) = E_{\Psi^{(0)}} \{ \log L_c(\Psi) | \mathbf{x} \},$$

ou seja, determina-se o valor esperado de  $\log L_c(\Psi)$  condicional aos valores observados  $\mathbf{x}$ , usando  $\Psi^{(0)}$  para  $\Psi$ ; na etapa M efectua-se a maximização de  $Q(\Psi; \Psi^{(0)})$  em relação a  $\Psi$ , sobre o espaço de parâmetros  $\Omega$ . Posto de outra forma, escolhemos  $\Psi^{(1)}$  tal que

$$Q(\Psi^{(1)}; \Psi^{(0)}) \geq Q(\Psi; \Psi^{(0)})$$

para todo o  $\Psi \in \Omega$ . As etapas E e M são calculadas novamente mas agora com  $\Psi^{(0)}$  substituído por  $\Psi^{(1)}$ . Na  $(k+1)$ -ésima iteração, as etapas E e M são definidas como se segue.

- **Etapa E:** Calcular  $Q(\Psi; \Psi^{(k)})$ , onde  $Q(\Psi; \Psi^{(k)}) = E_{\Psi^{(k)}} \{ \log L_c(\Psi) | \mathbf{x} \}$

- **Etapa M:** Escolher  $\Psi^{(k+1)}$  de modo a que seja um valor de  $\Omega$  que maximize  $Q(\Psi; \Psi^{(k)})$ , ou seja, tal que

$$Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi; \Psi^{(k)}), \forall \Psi \in \Omega.$$

As etapas E e M são sucessivamente repetidas até que a diferença

$$L(\Psi^{(k+1)}) - L(\Psi^{(k)}) \tag{A.3}$$

seja arbitrariamente pequena no caso de convergência da sucessão  $\{L(\Psi^{(k)})\}$ . Dempster *et al.* (1977) mostraram que a função de verosimilhança  $L(\Psi)$  (dos dados observados, logo incompletos) não decresce após cada iteração EM, ou seja,

$$L(\Psi^{k+1}) \geq L(\Psi^k)$$

para  $k = 0, 1, 2, \dots$ . Daí que a convergência tenha que ser obtida com uma sucessão de valores da verosimilhança que sejam limitados superiormente.

### A.3 Vantagens e desvantagens

Quando comparado com outros algoritmos utilizados para maximizar funções, tal como o associado ao método de Newton-Raphson, o algoritmo EM apresenta vantagens e desvantagens. Naturalmente que a escolha do algoritmo (ou eventualmente de uma combinação de dois) é feita de acordo com o tipo de situação.

#### Vantagens:

- ↗ Numericamente estável, em que cada iteração EM aumenta a verosimilhança (excepto num ponto fixo do algoritmo).
- ↗ Sob condições bastante gerais, converge globalmente. Assim, começando a partir de um ponto arbitrário  $\Psi^{(0)}$  do espaço de parâmetros, a convergência é quase sempre para um maximizante local, excepto se houver muito azar na escolha de  $\Psi^{(0)}$  ou se houver alguma patologia local do logaritmo da função de verosimilhança.

- ↗ Tipicamente é fácil de implementar porque se baseia nos dados completos: a etapa E de cada iteração envolve o cálculo do valor esperado da variável associada aos dados completos condicional aos dados observados e a etapa M resume-se à obtenção do EMV para dados completos que, frequentemente, ou se consegue obter de forma explícita ou recorrendo a algum *software* estatístico tradicional, no caso de não haver forma explícita.
- ↗ É fácil de programar pois não envolve nem o cálculo da verosimilhança, nem de alguma das suas derivadas.
- ↗ Como não é necessário guardar grandes quantidades de informação em cada iteração (como, por exemplo, a matriz de informação observada e a sua inversa), pode ser efectuado mesmo em computadores menos potentes.
- ↗ O esforço analítico necessário é muito menor uma vez que só é preciso maximizar o valor esperado condicional do logaritmo da verosimilhança dos dados completos. Embora possam ter que ser efectuados alguns cálculos na etapa E, em muitas aplicações não são complicados.
- ↗ Geralmente, o custo por iteração é pequeno, o que pode compensar o grande número de iterações necessárias neste algoritmo quando comparado com outros.
- ↗ Pode ser usado para fornecer valores estimados das observações omissas.

### **Desvantagens:**

- ↘ Não fornece simultaneamente uma estimativa da matriz de covariâncias das estimativas dos parâmetros. No entanto, o seu cálculo é possível com a metodologia adequada.
- ↘ Pode convergir lentamente, mesmo em situações consideradas simples, ou se houver demasiada informação incompleta.
- ↘ Nalguns casos, a etapa E pode ser analiticamente intratável, apesar de essa situação poder ser resolvida por simulações de Monte Carlo.

- ↘ À semelhança do que acontece com o método de Newton-Raphson, o algoritmo EM não garante a convergência para um máximo global quando existem vários máximos locais. Neste caso, a estimativa obtida depende do valor inicial considerado. No entanto, em geral, nenhum algoritmo de otimização garante a convergência para um máximo global ou local, e o algoritmo EM não é exceção.

## A.4 O método de Newton-Raphson

Como vários outros métodos para determinar o EMV, o algoritmo EM é um método para encontrar os zeros de uma função. Em Análise Numérica existem muitas técnicas para encontrar os zeros de uma função específica, incluindo o método de Newton-Raphson (abreviadamente, método de N-R), donde a sua abordagem neste anexo. Na verdade, além de ser uma das principais alternativas para o cálculo do EMV, também pode ser usado conjuntamente com o algoritmo EM, com benefícios mútuos.

Seja

$$\mathbf{S}(\mathbf{x}; \Psi) = \partial \log \mathbf{L}(\Psi) / \partial \Psi$$

o vector gradiente do logaritmo da função de verosimilhança, isto é, a estatística *score*. Denote-se por

$$\mathbf{I}(\Psi; \mathbf{x}) = -\partial^2 \log \mathbf{L}(\Psi) / \partial \Psi \partial \Psi^T,$$

a matriz simétrica da matriz das segundas derivadas parciais do logaritmo da verosimilhança em ordem aos elementos de  $\Psi$ .

A solução da equação de verosimilhança

$$\partial L(\Psi) / \partial \Psi = 0,$$

ou, de modo equivalente, da equação

$$\partial \log L(\Psi) / \partial \Psi = 0,$$

é a estimativa de máxima verosimilhança,  $\hat{\Psi}$ , de  $\Psi$ .

O método de N-R para resolver a equação de máxima verosimilhança

$$\mathbf{S}(\mathbf{x}; \Psi) = \mathbf{0},$$

aproxima o vector gradiente  $\mathbf{S}(\mathbf{x}; \Psi)$  do logaritmo da função de verosimilhança,  $L(\Psi)$ , por uma expansão em série de Taylor, de primeira ordem, em torno de  $\Psi^{(k)}$ . Assim,

$$\mathbf{S}(\mathbf{x}; \Psi) \approx \mathbf{S}(\mathbf{x}; \Psi^{(k)}) - \mathbf{I}(\Psi^{(k)}; \mathbf{x})(\Psi - \Psi^{(k)}). \quad (\text{A.4})$$

Ao determinar o zero da expressão do lado direito de (A.4), obtém-se  $\Psi^{(k+1)}$ , que é dado por

$$\Psi^{(k+1)} = \Psi^{(k)} + \mathbf{I}^{-1}(\Psi^{(k)}; \mathbf{x})\mathbf{S}(\mathbf{x}; \Psi^{(k)}). \quad (\text{A.5})$$

Se o logaritmo da função de verosimilhança for côncavo e unimodal, então a sucessão de iterações  $\{\Psi^{(k)}\}$  converge para o EMV de  $\Psi$ . Quando não for côncavo, não é garantido que o método de N-R convirja a partir de um valor inicial arbitrário.

Quando  $-2 \log L(\Psi)$  se alterar menos do que, digamos 0.025, entre duas iterações consecutivas, o processo iterativo termina e obtém-se a convergência. A razão de ser desta regra de paragem, Harrel Jr (2001), advém do facto de que as estimativas do parâmetro que afectem  $-2 \log L(\Psi)$  por menos do que o valor referido, não afectam a inferência estatística uma vez que  $-2 \log L(\Psi^{(k+1)}) - (-2 \log L(\Psi^{(k)}))$  tem distribuição  $\chi^2$ .

## A.5 Algumas variantes do algoritmo EM

Devido a certas dificuldades na implementação do algoritmo EM, nomeadamente quando a solução da etapa M não é explícita, surgiram algumas variantes para resolver este tipo de situação.

### A.5.1 Algoritmo EM generalizado

No caso de a solução da etapa M não ser explícita, pode não ser plausível tentar encontrar o valor de  $\Psi$  que maximize globalmente a função  $Q(\Psi; \Psi^{(k)})$ . Assim, Dempster *et al.* (1977) definiram o algoritmo EM generalizado (abreviadamente designado por algoritmo GEM) de modo que na etapa M,  $\Psi^{(k+1)}$  é escolhido tal que seja válida a desigualdade

$$Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi^{(k)}; \Psi^{(k)}).$$

Por outras palavras,  $\Psi^{(k+1)}$  é escolhido de forma a aumentar o valor da função  $Q(\Psi; \Psi^{(k)})$ , mas onde se fixa  $\Psi = \Psi^{(k)}$ , em vez de maximizar para qualquer  $\Psi \in \Omega$ . A condição anterior sobre  $\Psi^{(k+1)}$  é suficiente para garantir que

$$L(\Psi^{(k+1)}) \geq L(\Psi^{(k)}).$$

Assim, também neste caso, a verosimilhança  $L(\Psi)$  não decresce após cada iteração GEM. Então, uma sucessão de valores da verosimilhança resultantes do algoritmo GEM converge se for limitada superiormente.

### A.5.2 Algoritmo EM gradiente

Este algoritmo envolve não só o algoritmo EM, como ainda o método de N-R. Foi proposto por Lange (1995) e a diferença reside no facto de na etapa M ser efectuada uma iteração do método de N-R. Esta aproximação tem a vantagem de acelerar o processo de convergência.

Seja

$$\mathcal{I}_c(\Psi^{(k)}; \mathbf{x}) = - \left[ \partial Q(\Psi; \Psi^{(k)}) / \partial \Psi \partial \Psi^T \right]_{\Psi = \Psi^{(k)}}. \quad (\text{A.6})$$

Na prática, muitas vezes a matriz (A.6) é definida positiva (ainda que possa ser necessário efectuar uma reparametrização dos parâmetros), o que evita problemas de não convergência do algoritmo.

Recorde-se que na  $(k+1)$ -ésima iteração da etapa M pretende-se maximizar a função  $Q(\Psi; \Psi^{(k)})$ . Então, por (A.5), escolhe-se  $\Psi^{(k+1)}$  tal que

$$\Psi^{(k+1)} = \Psi^{(k)} + \mathcal{I}_c^{-1}(\Psi^{(k)}; \mathbf{x}) \mathbf{S}(\mathbf{x}; \Psi^{(k)}). \quad (\text{A.7})$$

Se o método de N-R for aplicado directamente aos dados incompletos, a respectiva  $(k+1)$ -ésima iteração escreve-se na forma

$$\Psi^{(k+1)} = \Psi^{(k)} + \mathbf{I}^{-1}(\Psi^{(k)}; \mathbf{x}) \mathbf{S}(\mathbf{x}; \Psi^{(k)}). \quad (\text{A.8})$$

Então, por (A.7) e (A.8), podemos afirmar que o algoritmo EM gradiente usa o método de N-R para encontrar o máximo da função de verosimilhança (dos dados incompletos) com a aproximação

$$\mathbf{I}^{-1}(\Psi^{(k)}; \mathbf{x}) \approx \mathcal{I}_c^{-1}(\Psi^{(k)}; \mathbf{x}).$$

### A.5.3 Um algoritmo EM híbrido

Já afirmamos antes que o algoritmo EM tem a vantagem de, sob condições bastante gerais, convergir globalmente. No entanto, essa convergência pode ser atingida apenas após um grande número de iterações, pois habitualmente a velocidade de convergência deste algoritmo é pequena. Por outro lado, o método de N-R apresenta uma velocidade de convergência mais rápida, apesar de não fornecer tantas garantias de convergência a partir de um valor inicial arbitrário. Assim, se conseguirmos juntar as vantagens destes dois métodos, concerteza que em muitos casos vamos obter um algoritmo melhor.

De facto, no contexto dos modelos de mistura finita, Aitkin e Aitkin (1996) sugeriram que as primeiras 5 iterações fossem de acordo com o algoritmo EM usual, mas que depois se optasse por aplicar o método de N-R em cada etapa M. Esta escolha baseou-se no trabalho relatado por Redner e Walker (1984), que afirmaram que, na sua experiência, 95% das alterações que se verificam na verosimilhança a partir do valor inicial até ao máximo, geralmente ocorrem nas primeiras 5 iterações. O critério de paragem que adoptaram foi o de que (A.3) fosse  $10^{-5}$ .

### A.5.4 Algoritmo ECM

Meng e Rubin (1993) propuseram que a etapa M do algoritmo EM fosse substituída por uma série de maximizações condicionais a alguma função dos parâmetros sob estimação, e assim criaram o que chamaram o algoritmo ECM. Esta substituição só tem razão de ser quando a etapa M é complicada pois a ideia ao aplicar a série de maximizações condicionais (abreviadamente designadas por etapas CM) é a simplificação a nível computacional. O preço a pagar é um maior número de iterações mas que não correspondem necessariamente a um maior tempo de execução global. Além disso, e ainda mais importante, é o facto deste algoritmo preservar as propriedades do algoritmo EM, nomeadamente a convergência monótona. O algoritmo ECM é um caso particular do algoritmo GEM.

Cada etapa CM pode ter uma solução explícita ou requerer, ela própria, iteração. Todavia, como as maximizações CM são sobre espaços de pequena dimensão, frequentemente são mais simples, rápidas e mais estáveis do que se procedessemos à maximização à custa de uma etapa M.

Suponhamos que cada etapa M é substituída por  $R > 1$  etapas. Seja  $\Psi^{(k+r/R)}$  o valor de  $\Psi$  na  $r$ -ésima etapa CM da  $(k+1)$ -ésima iteração.  $\Psi^{(k+r/R)}$  é escolhido de modo a maximizar a função  $Q(\Psi; \Psi^{(k)})$  mas sujeito à restrição

$$g_r(\Psi) = g_r(\Psi^{(k+(r-1)/R)}).$$

$C = \{g_r(\Psi), r = 1, \dots, R\}$  representa o conjunto formado pelas  $R$  funções vectoriais previamente fixadas e tem que ser "space filling". Assim,  $\Psi^{(k+r/R)}$  satisfaz a desigualdade

$$Q(\Psi^{(k+r/R)}; \Psi^{(k)}) \geq Q(\Psi; \Psi^{(k)}), \forall \Psi \in \Omega_r(\Psi^{(k+(r-1)/R)}),$$

onde

$$\Omega_r(\Psi^{(k+(r-1)/R)}) \equiv \{\Psi \in \Omega : g_r(\Psi) = g_r(\Psi^{(k+(r-1)/R)})\}.$$

O valor de  $\Psi$  na última etapa CM,  $\Psi^{(k+r/R)} = \Psi^{(k+1)}$ , será o valor a considerar para o início da  $(k+2)$ -ésima iteração.

### A.5.5 Algoritmo ECM multiciclos

Em algumas situações, o cálculo de uma etapa E pode ser bastante mais simples do que o cálculo das etapas CM. Neste caso, pode ser vantajoso efectuar uma etapa E antes de cada ou de algumas etapas CM.

Para clareza de exposição, suponhamos que efectuamos uma etapa E antes de cada etapa CM. Define-se um ciclo como sendo uma etapa E seguida de uma etapa CM. Meng e Rubin (1993) designaram este algoritmo por algoritmo ECM multiciclos.

A título de exemplo, consideremos  $R = 2$ . Então, após a primeira etapa CM, teremos uma etapa E, onde a função

$$Q(\Psi; \Psi^{(k)}) \tag{A.9}$$

é actualizada para

$$Q(\Psi; \Psi^{(k+1/2)}). \quad (\text{A.10})$$

Na segunda etapa CM,  $\Psi^{(k+2/R)} = \Psi^{(k+1)}$  é escolhido de modo a maximizar condicionalmente (A.10) e não (A.9).

Os resultados de convergência do algoritmo ECM aplicam-se também a este algoritmo. Uma desvantagem óbvia deste último é o acréscimo na quantidade de cálculos em cada iteração. Embora intuitivamente se possa esperar maiores diferenças na função de verosimilhança em cada iteração (uma vez que a função  $Q$  é actualizada mais vezes), isso não é necessariamente verdade.

### A.5.6 Outras variantes

Para nós, o algoritmo EM é um instrumento extremamente útil, mas não o objecto do nosso estudo. A razão da escolha das variantes que referimos até agora, é não só a sua importância mas também o seu uso no decorrer deste trabalho.

Por isso, para os potenciais interessados nesta matéria, remetemos para McLachlan e Krishnan (1997), uma excelente obra, onde se pode encontrar não só uma vasta teoria, como uma série de exemplos e uma numerosa lista de referências.

# Anexo B

## Estimação no modelo multiplicativo com fragilidade

Os dados que usámos, referidos em Cox e Oakes (1984, p. 169), são os que se seguem.

Tempos de vida:

tempo < - c(30, 67, 79, 82, 95, 148, 170, 171, 176, 193, 200, 221, 243, 261, 262, 263, 399, 414, 446, 446, 464, 777)

Indicadores de censura (1-tempo observado; 0-tempo censurado):

delta < - c(1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1)

Estimativa inicial para lambda:

sum(delta)/sum(tempo)  
[1] 0.003388621

No que se segue, optámos por dividir os dados iniciais em dois grupos, visto que as expressões dos estimadores usadas no algoritmo a desenvolver incluem somas que envolvem separadamente os dados censurados ( $i = 1, \dots, m$ ) e os dados observados ( $i = m + 1, \dots, n$ ). Assim, os dados observados são

*tempoo* < - c(30, 67, 95, 148, 170, 171, 176, 193, 200, 221, 243, 261, 262, 263, 399, 414, 446, 464, 777)

## Anexo B. Estimação no modelo multiplicativo com fragilidade

---

e os dados censurados são

$$tempoc < - c(79,82,446)$$

Algoritmo para estimar  $\lambda$ :

```
emlamb <- function(tempoc, tempoo, delta, lambda, theta)
{
  lambda <- - sum(delta)/(sum(tempoc/(1 + (1/theta) *
  lambda * tempoc)) + sum((tempoo * (1 + (theta * (theta + lambda *
  tempoo)) ^ (1/2))/(theta + lambda * tempoo))))
  return(lambda)
}
```

Algoritmo para estimar  $\theta$ :

```
emtheta <- function(tempoc, tempoo, lambda, theta)
{
  theta <- -1/(2 * ((1/22) * ((3/theta) + 2 * sum(((theta ^ (1/2)) +
  (theta ^ (-1/2)) * lambda * tempoc)/(theta + lambda * tempoc) ^ (1/2)) +
  sum((1 + (theta * (theta + lambda * tempoo)) ^ (1/2))/(theta + lambda *
  tempoo) + (1 + (1/theta) * lambda * tempoo) ^ (1/2))) - 2))
  return(theta)
}
```

As 2 funções anteriores (emlamb e emtheta) são usadas no algoritmo final, que envolve todas as iterações até à convergência.

```
emgi <- function(tempoc, tempoo, lambda, theta, n)
{
  for (i in 1 : n)
  {
    verI <- - 11 * (log(theta) - log(pi)) - 44 * theta - (3/4) * sum(log(theta) -
    log(theta + lambda * tempoc)) - theta * sum((1 + (1/theta) * lambda *
    tempoc) ^ (-1/2)) - theta * sum((1 + (theta * (theta + lambda *
    tempoc)) ^ (1/2))/theta) - (1/4) * sum(log(theta) - log(theta + lambda *
    tempoo)) - theta * sum((1 + (theta * (theta + lambda * tempoo)) ^ (1/2))/(theta +
    lambda * tempoo)) - theta * sum((1 + (1/theta) * lambda * tempoo) ^ (1/2)) +
    19 * log(lambda) - lambda * (sum(tempoc * (1 + (1/theta) * lambda *
    tempoc)) + sum(tempoo * (1 + (theta * (theta + lambda * tempoo)) ^ (1/2)))))
  }
}
```

```

tempoc) ^ (-1/2)) + sum(tempoo * (1 + (theta * (theta + lambda *
tempoo)) ^ (1/2))/(theta + lambda * tempoo)))
lambda1 <- emlamb(tempoc, tempoo, delta, lambda, theta)
theta1 <- emtheta(tempoc, tempoo, lambda, theta)
lambda <- lambda1
theta <- theta1
verF <- - 11 * (log(theta) - log(pi)) - 44 * theta - (3/4) * sum(log(theta) -
log(theta + lambda * tempoc)) - theta * sum((1 + (1/theta) * lambda *
tempoc) ^ (-1/2)) - theta * sum((1 + (theta * (theta + lambda *
tempoc)) ^ (1/2))/theta - (1/4) * sum(log(theta) - log(theta + lambda *
tempoo)) - theta * sum((1 + (theta * (theta + lambda * tempoo)) ^ (1/2))/(theta +
lambda * tempoo)) - theta * sum((1 + (1/theta) * lambda * tempoo) ^ (1/2)) +
19 * log(lambda) - lambda * (sum(tempoc * (1 + (1/theta) * lambda *
tempoc) ^ (-1/2)) + sum(tempoo * (1 + (theta * (theta + lambda *
tempoo)) ^ (1/2))/(theta + lambda * tempoo))))
dif <- - 2*verF+2*verI
final <- c(i, lambda1, theta1)
if(dif <= 0.00001 & dif >= -0.00001) return(final)
}
return(-1)
}

emgi(tempoc, tempoo, 0.003388621, 1, 50)
[1] 42 0.002259851 0.167939506

```

Experimentamos diferentes estimativas iniciais para  $\theta$  e o resultado foi:

```

emgi(tempoc, tempoo, 0.003388621, 2, 50)
[1] 43 0.002259851 0.167939536

emgi(tempoc, tempoo, 0.003388621, 0.5, 50)
[1] 41 0.00225985 0.16793943

emgi(tempoc, tempoo, 0.003388621, 10, 50)
[1] 46 0.002259850 0.167939421

```

Significa que o algoritmo no   dependente da estimativa inicial uma vez que no h grandes alteraces nem no valor dos parmetros nem no n<sup>o</sup> de iteraes.



## Anexo C

### Anexos do capítulo 4

## C.1 Funções hazard da distribuição Gompertz/Weibull modificada

Exibiremos neste anexo algumas funções hazard, de forma a revelar o modo como variam de acordo com o valor dos parâmetros.

O parâmetro gama é aquele que induz maiores alterações na função, pois é o responsável pelo sentido de crescimento (Figs C.1, C.2 e C.3). O parâmetro alfa é o parâmetro de escala (Fig. C.4) e o parâmetro lambda é um parâmetro indicador do valor de  $t$  a partir do qual a função hazard estabiliza num valor próximo de zero. Concretamente, quanto maior o valor de lambda (mais próximo de zero), menor o valor de  $t$  a partir do qual esperamos que os indivíduos que sobrevivem até esse instante estejam curados (Fig.C.5).

**Figura C.1:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\lambda = -2$  e  $\alpha = 2$ .

**Figura C.2:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\lambda=-2$  e  $\alpha=0.5$ .

**Figura C.3:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\lambda=-0.5$  e  $\alpha=0.5$ .

**Figura C.4:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\text{gama}=1.4$  e  $\text{lambda}=-1$ .

**Figura C.5:** Funções hazard da distribuição Gompertz/Weibull modificada, com  $\text{gama}=1.4$  e  $\text{alfa}=2$ .

## C.2 Estimação no modelo de cura baseado na distribuição log-logística

Os dados que usamos, referidos em Klein e Moeschberger (1998, p. 10), são os que se seguem.

Tempos de vida:

tempoallo < - c(0.030, 0.493, 0.855, 1.184, 1.283, 1.480, 1.776, 2.138, 2.500, 2.763, 2.993, 3.224, 3.421, 4.178, 4.441, 5.691, 5.855, 6.941, 6.941, 7.993, 8.882, 8.882, 9.145, 11.480, 11.513, 12.105, 12.796, 12.993, 13.849, 16.612, 17.138, 20.066, 20.329, 22.368, 26.776, 28.717, 28.717, 32.928, 33.783, 34.211, 34.770, 39.539, 41.118, 45.033, 46.053, 46.941, 48.289, 57.401, 58.322, 60.625)

Indicadores de censura (1-tempo observado; 0-tempo censurado):

deltaallo < - c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

Para obter a estimativa inicial da proporção de susceptíveis ( $1 - p = q$ ) procedemos da forma que se segue.

```
Sallo <- Surv(tempoallo,deltaallo)
fallo <- survfit(Sallo~1)
edit(fallo)
```

(as instruções anteriores fornecem-nos uma série de informações, nomeadamente a estimativa de Kaplan-Meier da função de sobrevivência)

Como o maior tempo de vida observado é 20.066, temos que  $fallo(20.066) = 0.532142477405635$ , logo  $q^{(0)} = 1 - 0.532142477405635 = 0.467857523$ .

Quanto aos valores iniciais dos parâmetros da distribuição log-logística, fizemos o ajustamento mas com os dados apenas até  $t_{(r)}$ , isto é, as observações são

tallot < - c(0.030, 0.493, 0.855, 1.184, 1.283, 1.480, 1.776, 2.138, 2.500, 2.763, 2.993, 3.224, 3.421, 4.178, 4.441, 5.691, 5.855, 6.941, 6.941, 7.993,

8.882, 8.882, 9.145, 11.480, 11.513, 12.105, 12.796, 12.993, 13.849, 16.612, 17.138, 20.066)

e

$\text{deltaalt} < -c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1)$

Assim, temos:

```
survreg(Surv(tallot,deltaalt)~ 1,dist="loglogistic")
```

```
Call: survreg(formula = Surv(tallot, deltaalt) ~1, dist = "loglogistic")
```

```
Coefficients:
```

```
(Intercept) 1.840259
```

```
Scale= 0.9101956
```

Note-se que  $\alpha = 1/\text{Scale}$  e que  $\lambda = \exp(-\text{Intercept}/\text{Scale})$ , donde  $\alpha^{(0)} = 1.098664946$  e  $\lambda^{(0)} = 0.132413185$ .

No que se segue, optámos por dividir os dados iniciais em dois grupos, visto que as expressões das derivadas parciais necessárias ao algoritmo a desenvolver incluem somas que envolvem separadamente os dados censurados ( $i = 1, \dots, m$ ) e os dados observados ( $i = m + 1, \dots, n$ ). Assim, os dados observados são

$\text{tao} < -c(0.030, 0.493, 0.855, 1.184, 1.283, 1.480, 1.776, 2.138, 2.500, 2.763, 2.993, 3.224, 3.421, 4.178, 5.691, 6.941, 8.882, 8.882, 11.480, 11.513, 12.796, 20.066)$

e os dados censurados são

$\text{tac} < -c(4.441, 5.855, 6.941, 7.993, 9.145, 12.105, 12.993, 13.849, 16.612, 17.138, 20.329, 22.368, 26.776, 28.717, 28.717, 32.928, 33.783, 34.211, 34.770, 39.539, 41.118, 45.033, 46.053, 46.941, 48.289, 57.401, 58.322, 60.625)$

### C.2.1 Algoritmo EM

Cada iterao   actualizada   custa dos valores da iterao anterior.

Para estimar  $q$ :

```
emq <- function(tac, lambda, alpha, q)
{
q <- (1/50) * (22+sum(q/(1 + (1 - q) * lambda * tac ^ alpha)))
return(q)
}
```

```
emq(tac, 0.132413185, 1.098664946, 0.467857523)
[1] 0.5261763
```

Para estimar  $lambda$ :

```
eml <- function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verosl <- - 22 * (log(alpha) + log(lambda)) + (alpha - 1) * sum(log(tao)) -
2 * sum(log(1 + lambda * tac ^ alpha)) - sum(q * log(1 + lambda * tac ^ alpha) / (1 +
(1 - q) * lambda * tac ^ alpha))
d1lamb <- - (22/lambda) - 2 * sum((tao ^ alpha) / (1 + lambda * tao ^ alpha)) -
q * sum((tac ^ alpha) / ((1 + (1 - q) * lambda * tac ^ alpha) * (1 + lambda *
tac ^ alpha)))
d2lamb <- - - (22/(lambda ^ 2)) + 2 * sum((tao ^ (2 * alpha)) / (1 + lambda *
tao ^ alpha) ^ 2) + q * sum((tac ^ (2 * alpha)) / ((1 + (1 - q) * lambda * tac ^ alpha) *
(1 + lambda * tac ^ alpha) ^ 2))
lambda <- - lambda - d1lamb/d2lamb
veroslF <- - 22 * (log(alpha) + log(lambda)) + (alpha - 1) * sum(log(tao)) -
2 * sum(log(1 + lambda * tac ^ alpha)) - sum(q * log(1 + lambda * tac ^ alpha) / (1 +
(1 - q) * lambda * tac ^ alpha))
Difl <- - - 2 * veroslF + 2 * verosl
# Crit rio de paragem
if(Difl <= 0.00001 & Difl >= -0.00001) return(lambda)
}
# Para saber quando ainda n o h  converg ncia
```

```
return(-1)
}
```

```
eml(tao, tac, 0.132413185, 1.098664946, 0.467857523, 8)
[1] -1
eml(tao, tac, 0.132413185, 1.098664946, 0.467857523, 20)
[1] 0.2060834
```

Para estimar  $\alpha$ :

```
ema <- function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verosaI <- - 22*(log(alpha)+log(lambda)) + (alpha - 1)*sum(log(tao)) -
2*sum(log(1+lambda*tac^alpha)) - sum(q*log(1+lambda*tac^alpha)/(1+
(1 - q) * lambda * tac^alpha))
d1alpha <- - (22/alpha)+sum(log(tao)) - 2 * lambda*sum((tao^alpha) *
log(tao)/(1 + lambda * tao ^ alpha)) - lambda * q*sum((tac ^ alpha) *
log(tac)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda * tac^alpha)))
d2alpha <- - - (22/(alpha ^ 2)) - 2 * lambda*sum(((tao ^ alpha) *
(log(tao))^2)/(1 + lambda * tao^alpha)^2) - lambda * q*sum(((tac ^ alpha) *
(log(tac))^2)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda * tac^alpha)^2))
alpha <- - alpha - d1alpha/d2alpha
verosaF <- - 22*(log(alpha)+log(lambda)) + (alpha - 1)*sum(log(tao)) -
2*sum(log(1+lambda*tac^alpha)) - sum(q*log(1+lambda*tac^alpha)/(1+
(1 - q) * lambda * tac^alpha))
Difa <- - - 2*verosaF + 2*verosaI
# Critério de paragem
if(Difa <= 0.00001 & Difa >= -0.00001) return(alpha)
}
# Para saber quando ainda não há convergência
return(-2)
}
```

```
ema(tao, tac, 0.132413185, 1.098664946, 0.467857523, 5)
[1] -2
ema(tao, tac, 0.132413185, 1.098664946, 0.467857523, 20)
```

[1] 1.335504

As 3 funoes anteriores (emq, eml e ema) so usadas no algoritmo final, que envolve todas as iteraoes at a convergncia.

```

emlogis <- function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verostI <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
q1 <- emq(tac, lambda, alpha, q)
lambda1 <- eml(tao, tac, lambda, alpha, q, n)
alpha1 <- ema(tao, tac, lambda, alpha, q, n)
q <- q1
lambda <- lambda1
alpha <- alpha1
verostF <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
Dif <- - 2*verostF+2*verostI
final <- c(i, q1, lambda1, alpha1)
if(Dif<= 0.00001&Dif>= -0.00001) return(final)
}
return(-3)
}

```

emlogis(tao, tac, 0.132413185, 1.098664946, 0.467857523, 20)

[1] -3

emlogis(tao, tac, 0.132413185, 1.098664946, 0.467857523, 60)

[1] 52 0.5092644 0.1780919 1.1934109

## C.2.2 Algoritmo ECM

**R=2**

O parâmetro  $q$  é actualizado à custa do valor da iteração anterior mas  $lambda$  e  $alpha$  já têm em conta o valor de  $q$  actualizado nessa mesma iteração. Assim, considera-se  $R=2$ , onde se fixa ora  $q$ , ora  $alpha$  e  $lambda$ .

```

emlogis1 <- function(tao, tac, lambda, alpha, q, n)
{
  for (i in 1 : n)
  {
    verostI <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac ^ alpha))-sum(q *
log(1 + lambda * tac ^ alpha)/(1 + (1 - q) * lambda * tac ^ alpha))
    q1 <- - emq(tac, lambda, alpha, q)
    q <- - q1
    lambda1 <- - eml(tao, tac, lambda, alpha, q, n)
    alpha1 <- - ema(tao, tac, lambda, alpha, q, n)
    lambda <- - lambda1
    alpha <- - alpha1
    verostF <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac ^ alpha))-sum(q *
log(1 + lambda * tac ^ alpha)/(1 + (1 - q) * lambda * tac ^ alpha))
    Dif <- - 2*verostF+2*verostI
    final <- - c(i, q1, lambda1, alpha1)
    if(Dif<= 0.00001&Dif>= -0.00001) return(final)
  }
  return(-3)
}

emlogis1(tao, tac, 0.132413185, 1.098664946, 0.467857523, 60)
[1] 24 0.5092644 0.1780918 1.1934110

```

Uma outra hipótese neste caso, seria estimar conjuntamente  $alpha$  e  $lambda$ , ou seja, em vez de usar `eml` e `ema`, usar apenas um algoritmo que estimasse simultaneamente  $alpha$  e  $lambda$ . Claro que a opção é novamente

o método de N-R mas com 2 parâmetros (esta opção também podia ter sido utilizada no algoritmo EM). Então, temos:

```

emla <- function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verosaI <- - 22*(log(alpha)+log(lambda))+(alpha-1)*sum(log(tao))-
2*sum(log(1+lambda*tac^alpha))-sum(q*log(1+lambda*tac^alpha)/(1+
(1-q)*lambda*tac^alpha))
d1alpha <- - (22/alpha)+sum(log(tao))- 2*lambda*sum((tao^alpha)*
log(tao)/(1+lambda*tao^alpha))-lambda*q*sum((tac^alpha)*
log(tac)/((1+(1-q)*lambda*tac^alpha)*(1+lambda*tac^alpha)))
d2alpha <- - (22/(alpha^2))- 2*lambda*sum(((tao^alpha)*
(log(tao))^2)/(1+lambda*tao^alpha)^2)-lambda*q*sum(((tac^alpha)*
(log(tac))^2)/((1+(1-q)*lambda*tac^alpha)*(1+lambda*tac^alpha)^2))
d1lamb <- - (22/lambda)-2*sum((tao^alpha)/(1+lambda*tao^alpha))-
q*sum((tac^alpha)/((1+(1-q)*lambda*tac^alpha)*(1+lambda*
tac^alpha)))
d2lamb <- - (22/(lambda^2))+2*sum((tao^(2*alpha))/(1+lambda*
tao^alpha)^2)+q*sum((tac^(2*alpha))/((1+(1-q)*lambda*tac^alpha)*
(1+lambda*tac^alpha)^2))
d2la <- - 2*sum((tao^alpha)*log(tao)/(1+lambda*tao^alpha)^2)-
q*sum((tac^alpha)*log(tac)/((1+(1-q)*lambda*tac^alpha)*(1+lambda*
tac^alpha)^2))
# Vector score
Ula <- matrix(c(d1alpha,d1lamb),1,2)
# Matriz de informação observada
Ila <- matrix(c(-d2alpha,-d2la,-d2la,-d2lamb),2,2)
# Inversa da matriz de informação observada
InvIla <- solve(Ila)
# Vector solução
Mla <- matrix(c(alpha,lambda),1,2)+Ula%%InvIla
alpha <- Mla[,1]
lambda <- Mla[,2]
verosaF <- - 22*(log(alpha)+log(lambda))+(alpha-1)*sum(log(tao))-
2*sum(log(1+lambda*tac^alpha))-sum(q*log(1+lambda*tac^alpha)/(1+
(1-q)*lambda*tac^alpha))

```

```

Difa < - - 2*verosaF+2*verosaI
if(Difa<= 0.00001&Difa>= -0.00001) return(Mla)
}
return(-2)
}

```

```

emla(tao, tac, 0.132413185, 1.098664946, 0.467857523, 20)
      [,1]      [,2]
[1,] 1.234903  0.1795981

```

```

emlogis01 < - function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verostI < - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac ^ alpha))-sum(q *
log(1 + lambda * tac ^ alpha)/(1 + (1 - q) * lambda * tac ^ alpha))
q1 < - emq(tac, lambda, alpha, q)
q < - q1
lambda1 < - emla(tao, tac, lambda, alpha, q, n)[,2]
alpha1 < - emla(tao, tac, lambda, alpha, q, n)[,1]
lambda < - lambda1
alpha < - alpha1
verostF < - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac ^ alpha))-sum(q *
log(1 + lambda * tac ^ alpha)/(1 + (1 - q) * lambda * tac ^ alpha))
Dif < - - 2*verostF+2*verostI
final < - c(i, q1, lambda1, alpha1)
if(Dif<= 0.00001&Dif>= -0.00001) return(final)
}
return(-3)
}

```

```

emlogis01(tao, tac, 0.132413185, 1.098664946, 0.467857523, 30)
[1] 15  0.5092645  0.1780919  1.1934109

```

**R=3**

O parâmetro  $q$  é actualizado à custa do valor da iteraço anterior mas  $lambda$  já tem em conta o valor de  $q$  actualizado nessa mesma iteraço e  $alpha$  tem em conta os valores de  $q$  e de  $lambda$  actualizados nessa mesma iteraço. Assim, considera-se R=3, onde se fixa ora  $q$ , ora  $alpha$ , ora  $lambda$ .

```

emlogis2 <- function(tao, tac, lambda, alpha, q, n)
{
  for (i in 1 : n)
  {
    verostI <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
    q1 <- - emq(tac, lambda, alpha, q)
    q <- q1
    lambda1 <- eml(tao, tac, lambda, alpha, q, n)
    lambda <- lambda1
    alpha1 <- ema(tao, tac, lambda, alpha, q, n)
    alpha <- alpha1
    verostF <- - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
    Dif <- - 2*verostF+2*verostI
    final <- c(i, q1, lambda1, alpha1)
    if(Dif<= 0.00001&Dif>= -0.00001) return(final)
  }
  return(-3)
}

emlogis2(tao, tac, 0.132413185, 1.098664946, 0.467857523, 30)
[1] 14 0.5092644 0.1780919 1.1934108

```

Observaço: Sem dvida que o algoritmo ECM acelerou o processo de convergncia.

### C.2.3 Algoritmo híbrido

Recorde-se que neste caso fazem-se 5 iterações usando o algoritmo EM e depois recorre-se ao método de N-R. Então começamos com um algoritmo baseado em A.1 e só depois é que temos um novo algoritmo.

```
emlogish5 <- function(tao, tac, lambda, alpha, q, n)
{
  for (i in 1 : n)
  {
    q1 <- emq(tac, lambda, alpha, q)
    lambda1 <- eml(tao, tac, lambda, alpha, q, n)
    alpha1 <- ema(tao, tac, lambda, alpha, q, n)
    q <- q1
    lambda <- lambda1
    alpha <- alpha1
    final <- c(i, q1, lambda1, alpha1)
    print(final)
  }
}
```

```
emlogish5(tao, tac, 0.132413185, 1.098664946, 0.467857523, 20)
[1] 1    0.5261763    0.2060834    1.3355041
[1] 2    0.4907917    0.1515533    1.1293645
[1] 3    0.5211279    0.1942607    1.2659873
[1] 4    0.4993536    0.1633242    1.1528012
[1] 5    0.5163527    0.1877413    1.2321157
...
```

Note-se que apesar de só querermos o valor dos parâmetros na 5<sup>a</sup> iteração, escolhemos  $n = 20$ . A razão é evitar problemas de convergência nos algoritmos eml e ema, que podem surgir se o número de iterações for pequeno.

O algoritmo para o método de N-R é:

```
NRlogish5 <- function(tao, tac, lambda, alpha, q, n)
{
  for (i in 1 : n)
  {
```

```

verostI < - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
d1q < - (22/q) - (28/(1 - q)) + q*sum(1/(q * (1 - q) * (1 + (1 - q) *
lambda * tac^alpha)))
d2q < - - (22/(q^2)) - (28/((1 - q)^2)) + q*sum(((2 * q - 1)/((q^2) *
((1 - q)^2) * (1 + (1 - q) * lambda * tac^alpha)))
d1alpha < - (22/alpha)+sum(log(tao)) - 2 * lambda*sum((tao^alpha) *
log(tao)/(1 + lambda * tao^alpha)) - lambda * q*sum((tac^alpha) *
log(tac)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda * tac^alpha)))
d2alpha < - - (22/(alpha^2)) - 2 * lambda*sum(((tao^alpha) *
(log(tao))^2)/(1 + lambda * tao^alpha)^2) - lambda * q*sum(((tac^alpha) *
(log(tac))^2)/((1+(1-q)*lambda*tac^alpha)*(1+lambda*tac^alpha)^2))
d1lamb < - (22/lambda)-2*sum((tao^alpha)/(1+lambda*tao^alpha))-
q*sum((tac^alpha)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda *
tac^alpha)))
d2lamb < - - (22/(lambda^2)) + 2*sum((tao^(2*alpha))/(1 + lambda *
tao^alpha)^2) + q*sum((tac^(2*alpha))/((1 + (1 - q) * lambda * tac^alpha) *
(1 + lambda * tac^alpha)^2))
d2la < - - 2*sum((tao^alpha) * log(tao)/(1 + lambda * tao^alpha)^2) -
q*sum((tac^alpha) * log(tac)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda *
tac^alpha)^2))
d2ql < - 0
d2qa < - 0
Uqla < - matrix(c(d1q,d1alpha,d1lamb),1,3)
Iqla < - matrix(c(-d2q,-d2qa,-d2ql,-d2qa,-d2alpha,-d2la,-d2ql,-d2la,-
d2lamb),3,3)
InvIqla < - solve(Iqla)
Mqla < - matrix(c(q, alpha, lambda),1,3)+Uqla%*%InvIqla
q < - Mqla[,1]
alpha < - Mqla[,2]
lambda < - Mqla[,3]
verostF < - 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
Dif < - - 2*verostF+2*verostI

```

```

final < - c(i, q, lambda, alpha)
if(Dif<= 0.00001&Dif>= -0.00001) return(final)
}
return(-3)
}

```

```

NRlogish5(tao, tac, 0.1877413, 1.2321157, 0.5163527, 30)
[1] 29    0.5092644    0.1780919    1.1934110

```

Sem dúvida que em relação ao algoritmo EM, este algoritmo é mais rápido, mas o mesmo já não podemos dizer em relação ao algoritmo ECM (neste caso).

### C.2.4 Algoritmo EM gradiente

Agora em cada etapa M aplica-se uma iteração do método de N-R.

```

NRglogis < - function(tao, tac, lambda, alpha, q, n)
{
d1q < - (22/q) - (28/(1 - q)) + q*sum(1/(q * (1 - q) * (1 + (1 - q) *
lambda * tac^alpha)))
d2q < - - (22/(q^2)) - (28/((1 - q)^2)) + q*sum((2 * q - 1)/((q^2) *
((1 - q)^2) * (1 + (1 - q) * lambda * tac^alpha)))
d1alpha < - (22/alpha)+sum(log(tao)) - 2 * lambda*sum((tao^alpha) *
log(tao)/(1 + lambda * tao^alpha)) - lambda * q*sum((tac^alpha) *
log(tac)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda * tac^alpha)))
d2alpha < - - (22/(alpha^2)) - 2 * lambda*sum(((tao^alpha) *
(log(tao))^2)/(1 + lambda * tao^alpha)^2) - lambda * q*sum(((tac^alpha) *
(log(tac))^2)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda * tac^alpha)^2))
d1lamb < - (22/lambda)-2*sum((tao^alpha)/(1+lambda*tao^alpha))-
q*sum((tac^alpha)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda *
tac^alpha)))
d2lamb < - - (22/(lambda^2)) + 2*sum((tao^(2*alpha))/(1 + lambda *
tao^alpha)^2) + q*sum((tac^(2*alpha))/((1 + (1 - q) * lambda * tac^alpha) *
(1 + lambda * tac^alpha)^2))
d2la < - - 2*sum((tao^alpha) * log(tao)/(1 + lambda * tao^alpha)^2) -
q*sum((tac^alpha) * log(tac)/((1 + (1 - q) * lambda * tac^alpha) * (1 + lambda *
tac^alpha)^2))

```

```

d2ql <- 0
d2qa <- 0
Uqla <- matrix(c(d1q,d1alpha,d1lamb),1,3)
Iqla <- matrix(c(-d2q,-d2qa,-d2ql,-d2qa,-d2alpha,-d2la,-d2ql,-d2la,-
d2lamb),3,3)
InvIqla <- solve(Iqla)
Mqla <- matrix(c(q, alpha, lambda),1,3)+Uqla%*%InvIqla
q <- Mqla[,1]
alpha <- Mqla[,2]
lambda <- Mqla[,3]
return(Mqla)
}

```

```

NRglogis(tao, tac, 0.132413185, 1.098664946, 0.467857523)
      [,1]      [,2]      [,3]
[1,] 0.5253112  1.163415  0.1571377

```

```

emglogis <- function(tao, tac, lambda, alpha, q, n)
{
for (i in 1 : n)
{
verostI <- 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
q1 <- NRglogis(tao, tac, lambda, alpha, q)[,1]
lambda1 <- NRglogis(tao, tac, lambda, alpha, q)[,3]
alpha1 <- NRglogis(tao, tac, lambda, alpha, q)[,2]
q <- q1
lambda <- lambda1
alpha <- alpha1
verostF <- 22 * log(q/(1 - q)) + 50 * log(1 - q) + (log(q/(1 - q))) *
q*sum(1/(1+(1-q)*lambda*tac^alpha))+22*(log(alpha)+log(lambda))+
(alpha - 1)*sum(log(tao)) - 2*sum(log(1 + lambda * tac^alpha))-sum(q *
log(1 + lambda * tac^alpha)/(1 + (1 - q) * lambda * tac^alpha))
Dif <- -2*verostF+2*verostI
final <- c(i, q1, lambda1, alpha1)
if(Dif<= 0.00001&&Dif>= -0.00001) return(final)
}
}

```

```
}  
return(-3)  
}
```

```
emglogis(tao, tac, 0.132413185, 1.098664946, 0.467857523, 50)  
[1] 33    0.5092645    0.1780919    1.1934109
```

Conclusão Final: Os valores encontrados aplicando qualquer um dos algoritmos são praticamente iguais:  $q$  ou é 0.5092644 ou é 0.5092645,  $lambda$  ou é 0.1780918 ou é 0.1780919, e  $alpha$  varia entre 1.1934108 e 1.1934110. Quanto ao número de iterações já existem diferenças consideráveis:

ECM, R=3: 14  
ECM, R=2, N-R bidimensional: 15  
ECM, R=2, N-R unidimensional: 24  
EM híbrido: 29  
EM gradiente: 33  
EM: 52

### C.3 Procedimento para a obtenção da variância das estimativas no modelo de cura baseado na distribuição log-logística

Como já referimos no anexo A, uma das desvantagens na utilização do algoritmo EM é a não obtenção simultânea da matriz de covariâncias das estimativas dos parâmetros. Vejamos pois qual a metodologia adequada para esta situação.

Alguns dos resultados aqui indicados também se encontram no capítulo 4. No entanto, optamos por repeti-los para não quebrar a sequência do raciocínio.

Começamos por notar que a matriz de informação observada,  $I(\hat{\theta}; \mathcal{O})$ , determina-se da forma que se segue (McLachlan e Krishnan, 1997, p.111),

$$I(\hat{\theta}, \mathcal{O}) = \mathcal{I}_C(\hat{\theta}; \mathcal{O}) - \mathcal{I}_m(\hat{\theta}; \mathcal{O}),$$

onde  $\theta$  representa o vector de parâmetros,  $\mathcal{I}_C(\hat{\theta}; \mathcal{O})$  a matriz de informação esperada para os dados completos, condicional aos dados observados e  $\mathcal{I}_m(\hat{\theta}; \mathcal{O})$  representa a matriz de informação dos dados omissos, condicional aos dados observados.

#### C.3.1 O modelo

O modelo de cura baseado na distribuição log-logística, escrito à custa da função de sobrevivência, é da forma

$$S(t) = \frac{1 + (1 - q)\lambda t^\alpha}{1 + \lambda t^\alpha},$$

onde  $q$ ,  $\alpha$  e  $\lambda$  são os parâmetros da distribuição. Assim, considerando a amostra  $t_1, \dots, t_n$ , o logaritmo da função de verosimilhança completa é

$$\begin{aligned} \log L_C = & (n - m) \log q + m \log(1 - q) + \sum_{i=1}^m y_i [\log q - \log(1 - q)] + \\ & (n - m)(\log \alpha + \log \lambda) + \sum_{i=m+1}^n [(\alpha - 1) \log t_i - 2 \log(1 + \lambda t_i^\alpha)] - \\ & \sum_{i=1}^m y_i \log(1 + \lambda t_i^\alpha), \end{aligned}$$

tendo em conta que  $y_i = 1$  sempre que  $t_i$  é observado, ou seja, quando  $1 = m + 1, \dots, n$ . Então, trabalhando com os dados completos e definindo  $\boldsymbol{\theta} = (q, \alpha, \lambda)$ , o vector *score* é

$$S_C(Y; \boldsymbol{\theta}) = \begin{bmatrix} \frac{n-m}{q} - \frac{m}{1-q} + \left(\frac{1}{q} + \frac{1}{1-q}\right) \sum_{i=1}^m y_i \\ \frac{n-m}{\alpha} + \sum_{i=m+1}^n [\log t_i - 2\lambda \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha}] - \lambda \sum_{i=1}^m y_i \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \\ \frac{n-m}{\lambda} - 2 \sum_{i=m+1}^n \frac{t_i^\alpha}{1+\lambda t_i^\alpha} - \sum_{i=1}^m y_i \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \end{bmatrix}.$$

Relativamente às derivadas de segunda ordem, temos

$$\begin{aligned} \frac{\partial^2 \log L_C}{\partial q^2} &= - \left[ \frac{n-m}{q^2} + \frac{m}{(1-q)^2} + \left( \frac{1}{q^2} - \frac{1}{(1-q)^2} \right) \sum_{i=1}^m y_i \right], \\ \frac{\partial^2 \log L_C}{\partial q \partial \alpha} &= \frac{\partial^2 \log L_C}{\partial \alpha \partial q} = \frac{\partial^2 \log L_C}{\partial q \partial \lambda} = \frac{\partial^2 \log L_C}{\partial \lambda \partial q} = 0, \\ \frac{\partial^2 \log L_C}{\partial \alpha^2} &= - \left[ \frac{n-m}{\alpha^2} + 2\lambda \sum_{i=m+1}^n \frac{t_i^\alpha (\log t_i)^2}{(1+\lambda t_i^\alpha)^2} + \lambda \sum_{i=1}^m y_i \frac{t_i^\alpha (\log t_i)^2}{(1+\lambda t_i^\alpha)^2} \right], \\ \frac{\partial^2 \log L_C}{\partial \lambda^2} &= - \left[ \frac{n-m}{\lambda^2} - 2 \sum_{i=m+1}^n \frac{t_i^{2\alpha}}{(1+\lambda t_i^\alpha)^2} - \sum_{i=1}^m y_i \frac{t_i^{2\alpha}}{(1+\lambda t_i^\alpha)^2} \right], \end{aligned}$$

$$\frac{\partial^2 \log L_C}{\partial \lambda \partial \alpha} = \frac{\partial^2 \log L_C}{\partial \alpha \partial \lambda} = - \left[ 2 \sum_{i=m+1}^n \frac{t_i^\alpha \log t_i}{(1 + \lambda t_i^\alpha)^2} - \sum_{i=1}^m y_i \frac{t_i^\alpha \log t_i}{(1 + \lambda t_i^\alpha)^2} \right].$$

Obtivemos assim a matriz

$$I_C(\boldsymbol{\theta}; \mathcal{O}) = \begin{bmatrix} -\frac{\partial^2 \log L_C}{\partial q^2} & 0 & 0 \\ 0 & -\frac{\partial^2 \log L_C}{\partial \alpha^2} & -\frac{\partial^2 \log L_C}{\partial \alpha \partial \lambda} \\ 0 & -\frac{\partial^2 \log L_C}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \log L_C}{\partial \lambda^2} \end{bmatrix}.$$

No entanto, para as observações censuradas, não conhecemos  $y_i$ , de modo que substituímos por

$$\tau_i = E(Y|\mathcal{O}) = \frac{q}{1 + (1 - q)\lambda t_i^\alpha}.$$

Por conseguinte, como a matriz de informação esperada para os dados completos, condicional aos dados observados, é dada por

$$\mathcal{I}_C(\boldsymbol{\theta}; \mathcal{O}) = E_{\boldsymbol{\theta}}[I_C(\boldsymbol{\theta}; \mathcal{O})|\mathcal{O}],$$

temos

$$\mathcal{I}_C(\boldsymbol{\theta}; \mathcal{O}) = \begin{bmatrix} -\frac{\partial^2 \log L_E}{\partial q^2} & 0 & 0 \\ 0 & -\frac{\partial^2 \log L_E}{\partial \alpha^2} & -\frac{\partial^2 \log L_C}{\partial \alpha \partial \lambda} \\ 0 & -\frac{\partial^2 \log L_E}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \log L_C}{\partial \lambda^2} \end{bmatrix},$$

onde o índice E em  $L_E$  indica que se substituiu em  $L_C$   $y_i$  por  $\tau_i$ , no caso das observações censuradas.

Vejamos agora o cálculo de  $\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O})$  através da fórmula

$$\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O}) = \text{cov}_{\boldsymbol{\theta}}\{S_C(Y; \boldsymbol{\theta})|\mathcal{O}\}.$$

O primeiro elemento da diagonal de  $\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O})$  é dado por

$$\begin{aligned}
 \mathcal{I}_{m;11}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i \right] \middle| \mathcal{O} \right\} \\
 &= \left( \frac{1}{q} + \frac{1}{1-q} \right)^2 \text{var}_{\boldsymbol{\theta}} \left\{ \left( \sum_{i=1}^m Y_i \right) \middle| \mathcal{O} \right\} \\
 &= \left( \frac{1}{q} + \frac{1}{1-q} \right)^2 \sum_{i=1}^m \tau_i (1 - \tau_i) \\
 &= \left( \frac{1}{q} + \frac{1}{1-q} \right)^2 \sum_{i=1}^m \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2}.
 \end{aligned}$$

Quanto ao segundo elemento, temos

$$\begin{aligned}
 \mathcal{I}_{m;22}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\alpha} + \sum_{i=m+1}^n [\log t_i - 2\lambda \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha}] - \lambda \sum_{i=1}^m Y_i \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \right] \middle| \mathcal{O} \right\} \\
 &= \lambda^2 \sum_{i=1}^m \left( \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \right)^2 \text{var}_{\boldsymbol{\theta}} \{ Y_i \middle| \mathcal{O} \} \\
 &= \lambda^2 \sum_{i=1}^m \left( \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \right)^2 \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2}.
 \end{aligned}$$

O terceiro elemento é

$$\begin{aligned}
 \mathcal{I}_{m;33}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\lambda} - 2 \sum_{i=m+1}^n \frac{t_i^\alpha}{1+\lambda t_i^\alpha} - \sum_{i=1}^m Y_i \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right] \middle| \mathcal{O} \right\} \\
 &= \sum_{i=1}^m \left( \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right)^2 \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2}.
 \end{aligned}$$

Relativamente aos restantes elementos da matriz, começamos por notar que, sendo  $X_i$  variáveis aleatórias i.i.d.,  $a$  e  $b_i$  constantes, então

$$\text{cov} \left( a \sum X_i, \sum b_i X_i \right) = a \sum b_i \text{var}(X_i).$$

Assim, temos

$$\begin{aligned}
 \mathcal{I}_{m;12}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, \right. \right. \\
 &\quad \left. \left. \frac{n-m}{\alpha} + \sum_{i=m+1}^n [\log t_i - 2\lambda \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha}] - \lambda \sum_{i=1}^m Y_i \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \right] \middle| \mathcal{O} \right\} \\
 &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, -\lambda \sum_{i=1}^m Y_i \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \right] \middle| \mathcal{O} \right\} \\
 &= -\lambda \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \text{var}_{\boldsymbol{\theta}}(Y_i) \\
 &= -\lambda \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha} \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2} \\
 &= \mathcal{I}_{m;21}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{I}_{m;13}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, \right. \right. \\
 &\quad \left. \left. \frac{n-m}{\lambda} - 2 \sum_{i=m+1}^n \frac{t_i^\alpha}{1+\lambda t_i^\alpha} - \sum_{i=1}^m Y_i \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right] \middle| \mathcal{O} \right\} \\
 &= -\left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \text{var}_{\boldsymbol{\theta}}(Y_i) \\
 &= -\left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2} \\
 &= \mathcal{I}_{m;31}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{I}_{m;23}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\alpha} + \sum_{i=m+1}^n [\log t_i - 2\lambda \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha}] - \lambda \sum_{i=1}^m Y_i \frac{t_i^\alpha \log t_i}{1+\lambda t_i^\alpha}, \right. \right. \\
 &\quad \left. \left. \frac{n-m}{\lambda} - 2 \sum_{i=m+1}^n \frac{t_i^\alpha}{1+\lambda t_i^\alpha} - \sum_{i=1}^m Y_i \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right] | \mathcal{O} \right\} \\
 &= -\lambda \sum_{i=1}^m \left( \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right)^2 (\log t_i) \text{var}_{\boldsymbol{\theta}}(Y_i) \\
 &= -\lambda \sum_{i=1}^m \left( \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right)^2 (\log t_i) \frac{q(1-q+(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2} \\
 &= \mathcal{I}_{m;32}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

Estamos agora em condições de passar ao cálculo de  $I(\hat{\boldsymbol{\theta}}, \mathcal{O})$ , bastando para tal subtrair membro a membro os elementos das matrizes  $\mathcal{I}_C(\hat{\boldsymbol{\theta}}; \mathcal{O})$  e  $\mathcal{I}_m(\hat{\boldsymbol{\theta}}; \mathcal{O})$ . Após algumas simplificações, os elementos da matriz  $I(\hat{\boldsymbol{\theta}}, \mathcal{O})$  são

$$\begin{aligned}
 I_{11}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{n-m}{q^2} + \frac{m}{(1-q)^2} - \sum_{i=1}^m \frac{1+2\lambda t_i^\alpha}{(1+(1-q)\lambda t_i^\alpha)^2} - \frac{1}{(1-q)^2} \sum_{i=1}^m \frac{q(2-q+2(1-q)\lambda t_i^\alpha)}{(1+(1-q)\lambda t_i^\alpha)^2}, \\
 I_{22}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{n-m}{\alpha^2} + 2\lambda \sum_{i=m+1}^n \left( \frac{\log t_i}{1+\lambda t_i^\alpha} \right)^2 t_i^\alpha + q\lambda \sum_{i=1}^m \left( \frac{\log t_i}{1+\lambda t_i^\alpha} \right)^2 \frac{(1-(1-q)\lambda^2 t_i^{2\alpha}) t_i^\alpha}{(1+(1-q)\lambda t_i^\alpha)^2}, \\
 I_{33}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{n-m}{\lambda^2} - 2 \sum_{i=m+1}^n \left( \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right)^2 - q \sum_{i=1}^m \left( \frac{t_i^\alpha}{1+\lambda t_i^\alpha} \right)^2 \frac{2-q+2(1-q)\lambda t_i^\alpha}{(1+(1-q)\lambda t_i^\alpha)^2}, \\
 I_{12}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{\lambda}{1-q} \sum_{i=1}^m \frac{(t_i^\alpha \log t_i)(1-q+(1-q)\lambda t_i^\alpha)}{(1+\lambda t_i^\alpha)(1+(1-q)\lambda t_i^\alpha)^2} = I_{21}(\hat{\boldsymbol{\theta}}, \mathcal{O}), \\
 I_{13}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{1}{1-q} \sum_{i=1}^m \frac{t_i^\alpha (1-q+(1-q)\lambda t_i^\alpha)}{(1+\lambda t_i^\alpha)(1+(1-q)\lambda t_i^\alpha)^2} = I_{31}(\hat{\boldsymbol{\theta}}, \mathcal{O}),
 \end{aligned}$$

$$I_{23}(\hat{\boldsymbol{\theta}}, \mathcal{O}) = -q \sum_{i=1}^m \frac{t_i^\alpha}{(1+\lambda t_i^\alpha)^2} \frac{1}{1+(1-q)\lambda t_i^\alpha} \left[ 1 + \frac{\lambda t_i^\alpha (1-q+(1-q)\lambda t_i^\alpha)}{1+(1-q)\lambda t_i^\alpha} \right] +$$

$$2 \sum_{i=m+1}^n \frac{t_i^\alpha}{(1+\lambda t_i^\alpha)^2} = I_{32}(\hat{\boldsymbol{\theta}}, \mathcal{O}).$$

Obtivemos, deste modo, a matriz de informação observada. O passo seguinte consiste em substituir cada um dos  $t_i$ ,  $q$ ,  $\alpha$  e  $\lambda$  pelo respectivo valor. Assim, temos

$$I(\hat{\boldsymbol{\theta}}; \mathcal{O}) = \begin{bmatrix} 154.582 & 25.749 & 50.359 \\ 25.749 & 21.556 & 23.174 \\ 50.359 & 23.174 & 213.641 \end{bmatrix}$$

e a matriz inversa é

$$I^{-1}(\hat{\boldsymbol{\theta}}; \mathcal{O}) = \begin{bmatrix} 0.00826 & -0.00880 & -0.00099 \\ -0.00880 & 0.06189 & -0.00464 \\ -0.00099 & -0.00464 & 0.00542 \end{bmatrix}.$$

Por fim, concluímos que o desvio padrão de  $\hat{q}$  é 0.09087, o de  $\hat{\alpha}$  é 0.24877 e o de  $\hat{\lambda}$  é 0.07361.

## C.4 Funções hazard do modelo de cura baseado na distribuição de Chen

Na Análise de Sobrevivência, a função hazard, a par da função de sobrevivência, desempenha um papel fundamental. O conhecimento das características desta função nas mais variadas distribuições permite, numa primeira abordagem, optar por um determinado modelo para ajustar aos dados em estudo. Por esta razão, apresentam-se aqui algumas funções hazard deste novo modelo.

**Figura C.6:** Funções hazard do modelo de cura baseado na distribuição de Chen, com  $q = 0.7$ ,  $\lambda = 2$  e  $\beta$  a variar entre 0.4 e 1.4.

**Figura C.7:** Funções hazard do modelo de cura baseado na distribuição de Chen, com  $q = 0.7$ ,  $\beta = 1.2$  e  $\lambda$  a variar entre 0.5 e 2.

**Figura C.8:** Funções hazard do modelo de cura baseado na distribuição de Chen, com  $\lambda = 2$ ,  $\beta = 1.2$  e  $q$  a variar entre 0.2 e 0.8.

## C.5 Estimação no modelo de cura baseado na distribuição de Chen

Os dados que usámos, são os mesmos que os usados no ajustamento com o modelo de cura baseado na distribuição log-logística. Desta vez optámos apenas pelo algoritmo EM, uma vez que já constatámos que as variações no algoritmo alteram o número de iterações mas não os valores das estimativas. A estimativa inicial do parâmetro  $q$  é a mesma do caso anterior, ou seja,  $q^{(0)} = 0.46786$ . Relativamente às estimativas iniciais dos restantes 2 parâmetros, vejamos qual foi o procedimento.

Temos

$$\begin{cases} S(0.03) = 0.53214 + 0.46786 \exp[\lambda(1 - \exp(0.03^\beta))] \\ S(4.178) = 0.53214 + 0.46786 \exp[\lambda(1 - \exp(4.178^\beta))] \end{cases} \Rightarrow$$

$$\begin{cases} \exp[\lambda(1 - \exp(0.03^\beta))] = (0.98 - 0.53214)/(0.46786) \\ \exp[\lambda(1 - \exp(4.178^\beta))] = (0.72 - 0.53214)/(0.46786) \end{cases}$$

Se  $\beta = 1$ ,

$$\begin{cases} \exp[\lambda(1 - \exp(0.03))] = -0.04369 \\ \exp[\lambda(1 - \exp(4.178))] = -0.91247 \end{cases} \Rightarrow \begin{cases} \lambda = 1.4346 \\ \lambda = 0.01421 \end{cases}$$

Se  $\beta = 0.5$ ,

$$\begin{cases} \exp[\lambda(1 - \exp(0.03^{0.5}))] = -0.04369 \\ \exp[\lambda(1 - \exp(4.178^{0.5}))] = -0.91247 \end{cases} \Rightarrow \begin{cases} \lambda = 0.23103 \\ \lambda = 0.13576 \end{cases}$$

Então,  $\beta^{(0)} = 0.5$  e  $\lambda^{(0)} = (0.23103 + 0.13576)/2 = 0.183395$ . Também experimentámos iniciar o algoritmo com outros valores (próximos) de  $\beta$  e o resultado foi de que não houve alterações significativas nos valores das estimativas obtidas.

### C.5.1 Algoritmo EM

Cada iteração é actualizada à custa dos valores da iteração anterior.

Para estimar  $q$ :

```
emqc <- function(tac, lambda, beta, q)
{
  q <- (1/50) * (22 + q * sum(exp(lambda * (1 - exp(tac ^ beta)))) / ((1 - q) +
q * exp(lambda * (1 - exp(tac ^ beta)))))
  return(q)
}

emqc(tac, 0.183395, 0.5, 0.467857523)
[1] 0.4491461
```

Para estimar  $lambda$ :

```
emlc <- function(tao, tac, lambda, beta, q)
{
  lambda <- 22 / (q * sum(((exp(lambda * (1 - exp(tac ^ beta)))) / (1 - q + q *
exp(lambda * (1 - exp(tac ^ beta))))) * (exp(tac ^ beta) - 1)) + sum(exp(tao ^ beta) -
1))
  return(lambda)
}

emlc(tao, tac, 0.183395, 0.5, 0.467857523)
[1] 0.07597151
```

Para estimar  $beta$ :

```
emb <- function(tao, tac, lambda, beta, q, n)
{
  for (i in 1 : n)
  {
    verosaI <- lambda * q * sum(((exp(lambda * (1 - exp(tac ^ beta)))) / (1 -
q + q * exp(lambda * (1 - exp(tac ^ beta))))) * (1 - exp(tac ^ beta))) +
22 * log(lambda * beta) + (beta - 1) * sum(log(tao)) + sum(tao ^ beta) -
lambda * sum(1 - exp(tao ^ beta))
```

```
d1beta <- -lambda*q*sum(((exp(lambda*(1-exp(tac^beta))))/(1-q+
q*exp(lambda*(1-exp(tac^beta))))*(log(tac))*(tac^beta)*exp(tac^beta))+
22/beta+sum((log(tao)) * (1 + tao ^ beta)) - lambda*sum((log(tao)) *
(tao^beta) * exp(tao^beta))
```

```
d2beta <- -(lambda*q*sum(((exp(lambda*(1-exp(tac^beta))))/(1-
q + q * exp(lambda * (1 - exp(tac^beta)))) * ((log(tac))^2) * (tac^beta) *
exp(tac^beta) * (1+tac^beta))+22/(beta^2)-sum(((log(tao))^2)*tao^beta)+
lambda*sum(((log(tao))^2) * (tao^beta) * exp(tao^beta) * (1 + tao^beta)))
```

```
beta <- beta-d1beta/d2beta
```

```
verosaF <- lambda * q*sum(((exp(lambda * (1 - exp(tac^beta))))/(1 -
q + q * exp(lambda * (1 - exp(tac^beta)))) * (1 - exp(tac^beta))) +
22 * log(lambda * beta) + (beta - 1)*sum(log(tao))+sum(tao^beta) -
lambda*sum(1 - exp(tao^beta))
```

```
Difa <- - 2*verosaF+2*verosaI
```

```
# Critério de paragem
```

```
if(Difa<= 0.00001&Difa>= -0.00001) return(beta)
```

```
}
```

```
# Para saber quando ainda não há convergência
```

```
return(-2)
```

```
}
```

```
emb(tao, tac, 0.183395, 0.5, 0.467857523, 5)
```

```
[1] -2
```

```
ema(tao, tac, 0.183395, 0.5, 0.467857523, 20)
```

```
[1] 1.335504
```

As 3 funções anteriores são usadas no algoritmo final, que envolve todas as iterações até à convergência.

```
emchen <- function(tao, tac, lambda, beta, q, n)
```

```
{
```

```
for (i in 1 : n)
```

```
{
```

```
verostI <- 28 * log(1 - q) + 22 * log(q) + (log(q) - log(1 - q)) *
q*sum((exp(lambda * (1 - exp(tac^beta))))/(1 - q + q * exp(lambda * (1 -
exp(tac^beta)))) + lambda * q*sum(((exp(lambda * (1 - exp(tac^beta))))/(1 -
q+q*exp(lambda*(1-exp(tac^beta))))*(1-exp(tac^beta))+22*log(lambda*
beta) + (beta - 1)*sum(log(tao))+sum(tao^beta) + lambda*sum(exp(tao^beta))
```

```

q1 <- emqc(tac, lambda, beta, q)
lambda1 <- emlc(tao, tac, lambda, beta, q, n)
beta1 <- emb(tao, tac, lambda, beta, q, n)
q <- q1
lambda <- lambda1
beta <- beta1
final <- c(i, q1, lambda1, beta1)
verostF <- - 28 * log(1 - q) + 22 * log(q) + (log(q) - log(1 - q)) *
q*sum((exp(lambda * (1 - exp(tac*beta))))/(1 - q + q * exp(lambda * (1 -
exp(tac*beta))))) + lambda * q*sum(((exp(lambda * (1 - exp(tac*beta))))/(1 -
q + q*exp(lambda*(1-exp(tac*beta))))) * (1 - exp(tac*beta))) + 22*log(lambda*
beta) + (beta - 1)*sum(log(tao)) + sum(tao*beta) + lambda*sum(exp(tao*beta))
Dif <- - 2*verostF + 2*verostI
if(Dif <= 0.00001 & Dif >= -0.00001) return(final)
}
return(-3)
}

emchen(tao, tac, 0.183395, 0.5, 0.467857523, 60)
[1] 51    0.4706010    0.1250839    0.4210026

```

## C.6 Procedimento para a obtenção da variância das estimativas no modelo de cura baseado na distribuição de Chen

### C.6.1 O modelo

O modelo de cura baseado na distribuição de Chen, escrito à custa da função de sobrevivência, é da forma

$$S(t) = 1 - q + q \exp[\lambda(1 - \exp(t^\beta))],$$

onde  $q$ ,  $\lambda$  e  $\beta$  são os parâmetros da distribuição. Assim, considerando a amostra  $t_1, \dots, t_n$ , o logaritmo da função de verosimilhança completa é

$$\begin{aligned} \log L_C = & (n - m) \log q + m \log(1 - q) + \sum_{i=1}^m y_i [\log q - \log(1 - q)] + \\ & (n - m)(\log \lambda + \log \beta) + \sum_{i=m+1}^n [(\beta - 1) \log t_i + t_i^\beta] + \\ & \lambda \sum_{i=m+1}^n (1 - \exp(t_i^\beta)) + \lambda \sum_{i=1}^m y_i (1 - \exp(t_i^\beta)), \end{aligned}$$

tendo em conta que  $y_i = 1$  sempre que  $t_i$  é observado, ou seja, quando  $1 = m + 1, \dots, n$ . Então, trabalhando com os dados completos e definindo  $\boldsymbol{\theta} = (q, \alpha, \beta)$ , o vector *score*,  $S_C(Y; \boldsymbol{\theta})$ , é

$$\begin{bmatrix} \frac{n-m}{q} - \frac{m}{1-q} + \left(\frac{1}{q} + \frac{1}{1-q}\right) \sum_{i=1}^m y_i \\ \frac{n-m}{\lambda} + \sum_{i=1}^m y_i (1 - \exp(t_i^\beta)) + \sum_{i=m+1}^n (1 - \exp(t_i^\beta)) \\ \frac{n-m}{\beta} - \lambda \sum_{i=1}^m y_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \sum_{i=m+1}^n (\log t_i) [1 + t_i^\beta - \lambda t_i^\beta \exp(t_i^\beta)] \end{bmatrix}.$$

Quanto às derivadas parciais de segunda ordem, temos

$$\frac{\partial^2 \log L_C}{\partial q^2} = -\frac{n-m}{q^2} - \frac{m}{(1-q)^2} + \left(-\frac{1}{q^2} + \frac{1}{(1-q)^2}\right) \sum_{i=1}^m y_i,$$

$$\frac{\partial^2 \log L_C}{\partial q \partial \beta} = \frac{\partial^2 \log L_C}{\partial \beta \partial q} = \frac{\partial^2 \log L_C}{\partial q \partial \lambda} = \frac{\partial^2 \log L_C}{\partial \lambda \partial q} = 0,$$

$$\begin{aligned} \frac{\partial^2 \log L_C}{\partial \beta^2} = & -\left[ \lambda \sum_{i=1}^m y_i (\log t_i)^2 t_i^\beta \exp(t_i^\beta) (1+t_i^\beta) + \frac{n-m}{\beta^2} - \right. \\ & \left. \sum_{i=m+1}^n (\log t_i)^2 t_i^\beta \left(1 - \lambda \exp(t_i^\beta) (1+t_i^\beta)\right) \right], \end{aligned}$$

$$\frac{\partial^2 \log L_C}{\partial \lambda^2} = -\frac{n-m}{\lambda^2},$$

$$\frac{\partial^2 \log L_C}{\partial \lambda \partial \beta} = \frac{\partial^2 \log L_C}{\partial \beta \partial \lambda} = -\left[ \sum_{i=1}^m y_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \sum_{i=m+1}^n (\log t_i) t_i^\beta \exp(t_i^\beta) \right].$$

Obtivemos assim a matriz

$$I_C(\boldsymbol{\theta}; \mathcal{O}) = \begin{bmatrix} -\frac{\partial^2 \log L_C}{\partial q^2} & 0 & 0 \\ 0 & -\frac{\partial^2 \log L_C}{\partial \lambda^2} & -\frac{\partial^2 \log L_C}{\partial \beta \partial \lambda} \\ 0 & -\frac{\partial^2 \log L_C}{\partial \lambda \partial \beta} & -\frac{\partial^2 \log L_C}{\partial \beta^2} \end{bmatrix}.$$

No entanto, para as observações censuradas, não conhecemos  $y_i$ , de modo que substituímos por

$$\tau_i = E(Y|\mathcal{O}) = \frac{q \exp[\lambda(1 - \exp(t_i^\beta))]}{1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]}.$$

Por conseguinte, como a matriz de informação esperada para os dados completos, condicional aos dados observados, é dada por

$$\mathcal{I}_C(\boldsymbol{\theta}; \mathcal{O}) = E_{\boldsymbol{\theta}}[I_C(\boldsymbol{\theta}; \mathcal{O})|\mathcal{O}],$$

temos

$$\mathcal{I}_C(\boldsymbol{\theta}; \mathcal{O}) = \begin{bmatrix} -\frac{\partial^2 \log L_E}{\partial q^2} & 0 & 0 \\ 0 & -\frac{\partial^2 \log L_E}{\partial \lambda^2} & -\frac{\partial^2 \log L_C}{\partial \beta \partial \lambda} \\ 0 & -\frac{\partial^2 \log L_E}{\partial \lambda \partial \beta} & -\frac{\partial^2 \log L_C}{\partial \beta^2} \end{bmatrix},$$

onde o índice E em  $L_E$  indica que se substituíu em  $L_C$   $y_i$  por  $\tau_i$ , no caso das observações censuradas.

Vejamos agora o cálculo de  $\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O})$  através da fórmula

$$\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O}) = \text{cov}_{\boldsymbol{\theta}}\{S_C(Y; \boldsymbol{\theta})|\mathcal{O}\}.$$

O primeiro elemento da diagonal de  $\mathcal{I}_m(\boldsymbol{\theta}; \mathcal{O})$  é dado por

$$\begin{aligned} \mathcal{I}_{m;11}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}}\left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i \right] | \mathcal{O} \right\} \\ &= \left( \frac{1}{q} + \frac{1}{1-q} \right)^2 \text{var}_{\boldsymbol{\theta}}\left\{ \left( \sum_{i=1}^m Y_i \right) | \mathcal{O} \right\} \\ &= \left( \frac{1}{q} + \frac{1}{1-q} \right)^2 \sum_{i=1}^m \tau_i (1 - \tau_i) \\ &= \frac{1}{q(1-q)} \sum_{i=1}^m \frac{\exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2}. \end{aligned}$$

Quanto ao segundo elemento, temos

$$\begin{aligned}
 \mathcal{I}_{m;22}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\lambda} + \sum_{i=1}^m Y_i (1 - \exp(t_i^\beta)) + \sum_{i=m+1}^n (1 - \exp(t_i^\beta)) \right] \middle| \mathcal{O} \right\} \\
 &= \sum_{i=1}^m (1 - \exp(t_i^\beta))^2 \text{var}_{\boldsymbol{\theta}} \{ Y_i \middle| \mathcal{O} \} \\
 &= q(1-q) \sum_{i=1}^m \frac{(1 - \exp(t_i^\beta))^2 \exp[\lambda(1 - \exp(t_i^\beta))]}{[1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]]^2}.
 \end{aligned}$$

O terceiro elemento é

$$\begin{aligned}
 \mathcal{I}_{m;33}(\boldsymbol{\theta}; \mathcal{O}) &= \text{var}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\beta} - \lambda \sum_{i=1}^m y_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \right. \right. \\
 &\quad \left. \left. \sum_{i=m+1}^n (\log t_i) [1 + t_i^\beta - \lambda t_i^\beta \exp(t_i^\beta)] \right] \middle| \mathcal{O} \right\} \\
 &= q(1-q) \lambda^2 \sum_{i=1}^m \left[ \frac{(\log t_i) t_i^\beta \exp(t_i^\beta)}{1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]} \right]^2 \exp[\lambda(1 - \exp(t_i^\beta))].
 \end{aligned}$$

Relativamente aos restantes elementos da matriz, obtemos

$$\begin{aligned}
 \mathcal{I}_{m;12}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, \right. \right. \\
 &\quad \left. \left. \frac{n-m}{\lambda} + \sum_{i=1}^m Y_i (1 - \exp(t_i^\beta)) + \sum_{i=m+1}^n (1 - \exp(t_i^\beta)) \right] \middle| \mathcal{O} \right\} \\
 &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, \sum_{i=1}^m Y_i (1 - \exp(t_i^\beta)) \right] \middle| \mathcal{O} \right\} \\
 &= \frac{1}{q(1-q)} \sum_{i=1}^m \text{var}_{\boldsymbol{\theta}}(Y_i) (1 - \exp(t_i^\beta)) \\
 &= q \sum_{i=1}^m \frac{\exp[\lambda(1 - \exp(t_i^\beta))] (1 - \exp(t_i^\beta))}{[1 - q + q \exp[\lambda(1 - \exp(t_i^\beta))]]^2} \\
 &= \mathcal{I}_{m;21}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{I}_{m;13}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{q} - \frac{m}{1-q} + \left( \frac{1}{q} + \frac{1}{1-q} \right) \sum_{i=1}^m Y_i, -\lambda \sum_{i=1}^m y_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \right. \right. \\
 &\quad \left. \left. \frac{n-m}{\beta} + \sum_{i=m+1}^n (\log t_i) [1 + t_i^\beta - \lambda t_i^\beta \exp(t_i^\beta)] \right] \middle| \mathcal{O} \right\} \\
 &= -\frac{\lambda}{q(1-q)} \sum_{i=1}^m (\log t_i) t_i^\beta \exp(t_i^\beta) \text{var}_{\boldsymbol{\theta}}(Y_i) \\
 &= -\lambda \sum_{i=1}^m \frac{(\log t_i) t_i^\beta \exp(t_i^\beta) \exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2} \\
 &= \mathcal{I}_{m;31}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{I}_{m;23}(\boldsymbol{\theta}; \mathcal{O}) &= \text{cov}_{\boldsymbol{\theta}} \left\{ \left[ \frac{n-m}{\lambda} + \sum_{i=1}^m Y_i (1 - \exp(t_i^\beta)) + \sum_{i=m+1}^n (1 - \exp(t_i^\beta)), \frac{n-m}{\beta} - \right. \right. \\
 &\quad \left. \left. \lambda \sum_{i=1}^m y_i (\log t_i) t_i^\beta \exp(t_i^\beta) + \sum_{i=m+1}^n (\log t_i) [1 + t_i^\beta - \lambda t_i^\beta \exp(t_i^\beta)] \right] \middle| \mathcal{O} \right\} \\
 &= -\lambda \sum_{i=1}^m (1 - \exp(t_i^\beta)) (\log t_i) t_i^\beta \exp(t_i^\beta) \text{var}_{\boldsymbol{\theta}}(Y_i) \\
 &= -\lambda q(1-q) \sum_{i=1}^m (1 - \exp(t_i^\beta)) (\log t_i) t_i^\beta \exp(t_i^\beta) \frac{\exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2} \\
 &= \mathcal{I}_{m;32}(\boldsymbol{\theta}; \mathcal{O}),
 \end{aligned}$$

Estamos agora em condições de passar ao cálculo de  $I(\hat{\boldsymbol{\theta}}, \mathcal{O})$ . Após algumas simplificações, os elementos da matriz  $I(\hat{\boldsymbol{\theta}}, \mathcal{O})$  são

$$\begin{aligned}
 I_{11}(\hat{\boldsymbol{\theta}}, \mathcal{O}) &= \frac{n-m}{q^2} + \frac{m}{(1-q)^2} + \frac{1-2q}{q(1-q)^2} \sum_{i=1}^m \frac{\exp[\lambda(1-\exp(t_i^\beta))]}{1-q+q \exp[\lambda(1-\exp(t_i^\beta))]} - \\
 &\quad \frac{1}{q(1-q)} \sum_{i=1}^m \frac{\exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2},
 \end{aligned}$$

$$I_{22}(\hat{\theta}, \mathcal{O}) = \frac{n-m}{\lambda^2} - q(1-q) \sum_{i=1}^m \frac{(1-\exp(t_i^\beta))^2 \exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2},$$

$$I_{33}(\hat{\theta}, \mathcal{O}) = \frac{n-m}{\beta^2} + \sum_{i=m+1}^n (\log t_i)^2 t_i^\beta [\lambda(1+t_i^\beta) - 1] +$$

$$q\lambda \sum_{i=1}^m \frac{(\log t_i)^2 t_i^\beta \exp(t_i^\beta) \exp[\lambda(1-\exp(t_i^\beta))]}{1-q+q \exp[\lambda(1-\exp(t_i^\beta))]} \left[ 1 + t_i^\beta - \frac{\lambda(1-q)t_i^\beta \exp(t_i^\beta)}{1-q+q \exp[\lambda(1-\exp(t_i^\beta))]} \right],$$

$$I_{12}(\hat{\theta}, \mathcal{O}) = -q \sum_{i=1}^m \frac{\exp[\lambda(1-\exp(t_i^\beta))](1-\exp(t_i^\beta))}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2} = I_{21}(\hat{\theta}, \mathcal{O}),$$

$$I_{13}(\hat{\theta}, \mathcal{O}) = \lambda \sum_{i=1}^m \frac{(\log t_i) t_i^\beta \exp(t_i^\beta) \exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2} = I_{31}(\hat{\theta}, \mathcal{O}),$$

$$I_{23}(\hat{\theta}, \mathcal{O}) = \lambda q(1-q) \sum_{i=1}^m (1-\exp(t_i^\beta)) (\log t_i) t_i^\beta \exp(t_i^\beta) \frac{\exp[\lambda(1-\exp(t_i^\beta))]}{[1-q+q \exp[\lambda(1-\exp(t_i^\beta))]]^2}$$

$$= I_{32}(\hat{\theta}, \mathcal{O}).$$

Obtivemos, assim, a matriz de informação observada. Substituindo cada um dos  $t_i$ ,  $q$ ,  $\beta$  e  $\lambda$  pelo respectivo valor, temos

$$I(\hat{\theta}; \mathcal{O}) = \begin{bmatrix} 132.01820 & 26.65537 & 48.01669 \\ 26.65537 & 1189.57600 & -217.81930 \\ 48.01669 & -217.81930 & 1130.53100 \end{bmatrix}$$

e a matriz inversa é

$$I^{-1}(\hat{\theta}; \mathcal{O}) = \begin{bmatrix} 0.00776 & -0.00024 & -0.00038 \\ -0.00024 & 0.00088 & 0.00018 \\ -0.00038 & 0.00018 & 0.00094 \end{bmatrix}.$$

Por fim, concluímos que o desvio padrão de  $\hat{q}$  é 0.08809, o de  $\hat{\lambda}$  é 0.02965 e o de  $\hat{\beta}$  é 0.03058.

## Anexo D

# Estimação no modelo cura com censura parcialmente informativa

Os dados que usámos, podem ser encontrados em Andersen *et al.* (1993, p. 709-714).

Tempos de vida:

$tempo < -c(10,30,35,99,185,204,210,232,232,279,295,355,386,426,469,493,529,621,629,659,667,718,752,779,793,817,826,833,858,869,872,967,977,982,1041,1055,1062,1075,1156,1228,1252,1271,1312,1427,1435,1499,1506,1508,1510,1512,1516,1525,1542,1548,1557,1560,1563,1584,1605,1621,1627,1634,1641,1641,1648,1652,1654,1654,1667,1678,1685,1690,1710,1710,1726,1745,1762,1779,1787,1787,1793,1804,1812,1836,1839,1839,1854,1856,1860,1864,1899,1914,1919,1920,1927,1933,1942,1955,1956,1958,1963,1970,2005,2007,2011,2024,2028,2038,2056,2059,2061,2062,2075,2085,2102,2103,2104,2108,2112,2150,2156,2165,2209,2227,2227,2256,2264,2339,2361,2387,2388,2403,2426,2426,2431,2460,2467,2492,2493,2521,2542,2559,2565,2570,2660,2666,2676,2738,2782,2787,2984,3032,3040,3042,3067,3079,3101,3144,3152,3154,3180,3182,3185,3199,3228,3229,3278,3297,3328,3330,3338,3383,3384,3385,3388,3402,3441,3458,3459,3459,3476,3523,3667,3695,3695,3776,3776,3830,3856,3872,3909,3968,4001,4103,4119,4124,4207,4310,4390,4479,4492,4668,4688,4926,5565)$



## Anexo D. Estimação no modelo cura com censura parcialmente informativa

```
verosaF <- -14*log(beta) + 71*log(alpha) + lambda*sum(taoi) + (beta +
1) * (alpha/lambda)*sum(1 - exp(lambda * tempo))
Difa <- -2*verosaF + 2*verosaI
vector <- c(lambda, alpha, beta, Difa)
if(Difa <= 0.00001 & Difa >= -0.00001) return(vector)
}
return(-1) # Critério de paragem
}
```

Para iniciar o algoritmo, precisamos de valores iniciais para os parâmetros. Relativamente a  $\beta$ , temos  $\beta^{(0)} = 14/57 = 0.2456140$ , uma vez que existem 14 observações censuradas informativamente e 57 tempos de vida observados.

Quanto aos valores iniciais para os parâmetros da distribuição de Gompertz modificada, começamos por notar que, recorrendo à aproximação  $1 + x = \exp(x)$ , "obtemos" uma distribuição exponencial de parâmetro  $\alpha$ . Então,  $\alpha^{(0)} = \text{sum}(\text{delta})/\text{sum}(\text{tempo}) = 0.0001292154$ .

Através da estimativa de Kaplan-Meier da função de sobrevivência, estimamos  $\hat{S}(\infty) = 0.644875856126596$ . Por outro lado, como estamos a admitir a distribuição de Gompertz modificada para o tempo de vida, então  $S(\infty) = \exp(\hat{\alpha}/\hat{\lambda})$ . Assim,  $\lambda^{(0)} = 0.0001292154/\log(0.644875856126596) = -0.0002945433$ .

Com os valores iniciais anteriores, temos o resultado ao fim de 15 iterações.

```
NR12(tempo, taoi, -0.0002945433, 0.0001292154, 0.2456140, 15)
[1] -1.026668e-04  1.479660e-04  2.455418e-01  -6.220556e-06
```

Agora, falta-nos actualizar a estimativa do parâmetro  $\alpha$ ,

```
alphaF <- -(-71 * (-0.0001026668))/((2.455418e - 01 + 1)*sum(1 -
exp((-0.0001026668) * tempo)))
```

```
alphaF
[1] 0.0001479468
```

bem como a estimativa do parâmetro  $\beta$ ,

Anexo D. Estimação no modelo cura com censura parcialmente informativa

```
betaF < - (-14 * (-0.0001026668))/(0.0001479468*sum(1 -  
exp((-0.0001026668) * tempo)))
```

```
betaF  
[1] 0.2455999
```



Quanto ao valor inicial do parmetro da distribuico exponencial, precismos de mais alguns cculos. Considermos o modelo sem fragilidade,  semelhana do que j tnhamos feito no caso do modelo multiplicativo com fragilidade. Assim, obtivemos:

```
survreg(Sallo1~1,dist="exponential")
```

```
Call: survreg(formula = Sallo1~1, dist = "exponential")
```

```
Coefficients:
```

```
(Intercept)
```

```
0.7490045
```

```
Scale fixed at 1
```

Note-se que  $\lambda = \exp(-\text{Intercept})$ , donde  $\lambda^{(0)} = 0.472837$ .

No que se segue, optmos por dividir os dados iniciais em dois grupos, visto que as expresses dos estimadores usadas no algoritmo a desenvolver incluem somas que envolvem separadamente os dados censurados ( $i = 1, \dots, m$ ) e os dados observados ( $i = m + 1, \dots, n$ ). Assim, os dados observados so

```
t1o <- c(0.0301,0.0384,0.0630,0.0849,0.0877,0.0959,0.1397,0.1616,0.1699,
0.2137,0.2137,0.2164,0.2384,0.2712,0.2740,0.3863,0.4384,0.4548,0.5918,0.6000,
0.6438,0.6849,0.7397,0.8575,0.9096,0.9644,1.0082,1.2822,1.3452,1.4000,1.5260,
2.2438,3.4411)
```

e os dados censurados so

```
t1c <- c(1.7205,1.9890,2.5068,2.6466,3.0384,3.1726,4.4219,4.4356,4.5863,
4.6904,4.7808,4.9863,5.000)
```

## E.1 Algoritmo EM

Cada iterao  actualizada  custa dos valores da iterao anterior.

O algoritmo para  $q$ :

```
emfq <- function(t1c, lambda, theta, q, n)
{
for (i in 1 : n)
```

```

{
  c <- theta + lambda * t1c
  tau_i <- - (q * exp(-2 * (theta * c)^(1/2) + 2 * theta)) / (1 - q + q * exp(-2 *
(theta * c)^(1/2) + 2 * theta))
  verosI <- - sum(tau_i * log(q/(1 - q))) - 13 * log(q/(1 - q)) + 46 * log(q)
  q <- - (1/46) * (sum(tau_i) + 33)
  verosF <- - sum(tau_i * log(q/(1 - q))) - 13 * log(q/(1 - q)) + 46 * log(q)
  Dif <- - 2 * verosF + 2 * verosI
  print(q)
  if(Dif <= 0.00001 & Dif >= -0.00001) return(q)
}
return(-1)
}

```

```
emfqn(t1c, 0.472837, 1, 0.736622, 50)
```

```
[1] 0.8403248
```

```
...
```

```
[1] 0.9708967
```

Ao fim de 41 iterações temos convergência.

#### O algoritmo para $\lambda$ e $\theta$ :

```

emfltn <- function(t1o, t1c, lambda, theta, q, n)
{
  for (i in 1 : n)
  {
    c <- theta + lambda * t1c
    o <- theta + lambda * t1o
    tau_i <- - (q * exp(-2 * (theta * c)^(1/2) + 2 * theta)) / (1 - q + q * exp(-2 *
(theta * c)^(1/2) + 2 * theta))
    verosI <- - sum(tau_i * ((-1/4) * log(theta) - ((1/2) * log(pi)) + (2 * theta) -
1 + ((3/4) * log(c)) + ((3/4) * (theta * c)^(-1/2)) - (theta * (theta/c)^(1/2)) -
((theta * c)^(1/2)) - ((1 + (theta * c)^(1/2)) / (c)))) + 33 * ((1/4) * log(theta) -
(1/2) * log(pi) + 2 * theta + log(lambda) - 1) + sum(((1/4) * log(o)) - ((1/4) *
(theta * o)^(-1/2)) - (theta * (o/theta)^(1/2)) - (theta * o)^(1/2))
    lambda <- - 33 / (sum(tau_i * ((theta/c)^(1/2)) * t1c) + sum(((1 + (theta *
o)^(1/2)) / (o)) * t1o))
  }
}

```

$$theta < - (\text{sum}(taui) + 33) / (2 * (\text{sum}(taui * ((theta/c) ^ (1/2) + ((1 + (theta * c) ^ (1/2)) / theta) - 2))) + \text{sum}(((1 + (theta * o) ^ (1/2)) / (o)) + (1 + (lambda/theta) * t1o) ^ (1/2)) - 2 * 33))$$

$$\text{verosF} < - \text{sum}(taui * \log(q/(1 - q))) - 13 * \log(q/(1 - q)) + 46 * \log(q) + \text{sum}(taui * ((-1/4) * \log(theta) - ((1/2) * \log(pi)) + (2 * theta) - 1 + ((3/4) * \log(c)) + ((3/4) * (theta * c) ^ (-1/2)) - (theta * (theta/c) ^ (1/2)) - ((theta * c) ^ (1/2)) - ((1 + (theta * c) ^ (1/2)) / (c)))) + 33 * ((1/4) * \log(theta) - (1/2) * \log(pi) + 2 * theta + \log(lambda) - 1) + \text{sum}(((1/4) * \log(o)) - ((1/4) * (theta * o) ^ (-1/2)) - (theta * (o/theta) ^ (1/2)) - (theta * o) ^ (1/2))$$

$$\text{Dif} < - - 2 * \text{verosF} + 2 * \text{verosI}$$

$$\text{final} < - c(i, lambda1, theta1, \text{Dif})$$

$$\text{if}(\text{Dif} <= 0.00001 \& \text{Dif} >= -0.00001) \text{return}(\text{final})$$

}

$$\text{return}(-2)$$

}

$$\text{emftn}(t1o, t1c, 0.4728370, 1, 0.9708967, 55)$$

$$[1] \ 78 \quad 0.5606802 \quad 0.03277456 \quad 9.110584e-06$$

Ao fim de 78 iterações temos convergência.

Assim,  $lambda = 0.5606802$ ,  $theta = 0.03277456$  e  $q = 0.9708967$ .

# Bibliografia

- [1] Aalen, O.O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, Vol.7, p. 1121-1137.
- [2] Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability*, Vol.2, p. 951-972.
- [3] Abdushukurov, A.A. (1984). On some estimates of the distribution function under random censorship. Em *Conference of Young Scientists*, Math. Inst. Acad. Sci., Uzbek SSR, Tashkent. VINITI Nº 8756-V.
- [4] Abreu, A.M. (1997). *Modelos de Sobrevivência para Populações Heterogêneas*. Tese de Mestrado. Faculdade de Ciências da Universidade de Lisboa.
- [5] Aitkin, M. e Aitkin, I. (1996). A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions. *Statistics and Computing*, Vol.6, p. 127-130.
- [6] Andersen, P.K., Borgan, O., Gill, R.D. e Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- [7] J.E. Anderson, e T.A. Louis (1995). Survival Analysis Using a Scale Change Random Effects Model. *Journal of the American Statistical Association*, Vol.90, p. 669-679.
- [8] Berkson, J. e Gage, R.P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, Vol.47, p. 501-515.

- [9] Bernardo, M.V.P. e Ibrahim, J.G. (2000). Group sequential designs for cure rate models with early stopping in favour of the null hypothesis. *Statistics in Medicine*, Vol.19, p. 3023-3035.
- [10] Betensky, R.A. e Schoenfeld, D.A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics*, Vol.57, p. 282-286.
- [11] Block, H. e Joe, H. (1997). Tail behavior of the failure rate functions. *Lifetime Data Analysis*, Vol.3, p. 269-288.
- [12] Boag, J.Q. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of Royal Statistical Society*, Vol.11, p. 15-53.
- [13] Braekers, R. e Veraverbeke, N. (2001). The partial Koziol-Green model with covariates. *Journal of Statistical Planning and Inference*, Vol.92, p. 55-71.
- [14] Cabral, M.S.E. (1994). Teste de Razão de Verossimilhanças para uma Mistura de Normais. Um Estudo de Simulação. Em *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística* p. 463-477.
- [15] Cantor, A.B. e Shuster, J.J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, Vol.11, p. 931-937.
- [16] Chao, E.C. (1998). Gibbs sampling for long-term survival data with competing risks. *Biometrics*, Vol.54, p. 350-366.
- [17] Chen, M.-H., Ibrahim, G. e Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, Vol.94, 447, p. 909-919.
- [18] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, Vol.49, p. 155-161.
- [19] Cheng, P.E. e Lin, G.D. (1984). Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazard model. *Technical Report*, Institute of Statistics, Academia Sinica.

- 
- [20] Cheng, P.E. e Lin, G.D. (1987). Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazard model. *Statistics and Probability Letters*, Vol.5, p. 75-80.
- [21] Choi, K.C. e Zhou, X. (2002). Large sample properties of mixture models with covariates for competing risks. *Journal of Multivariate Analysis*, Vol.82, p. 331-366.
- [22] Collet, D. (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall, London.
- [23] Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, B*, Vol.34, p. 187-220.
- [24] Cox, D.R. e Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London.
- [25] Crowder, M.J. (2001). *Classical Competing Risks*. Chapman & Hall, London.
- [26] Csörgö, S. (1998). Testing for the partial proportional hazards model of random censorship. Em *Prague Stochastics' 98 Proceedings*. Lachout, H.M. e Visek, J.A., editores. Union of Czech Mathematicians and Physicists, Prague, p. 87-92.
- [27] De Angelis, R., Capocaccia, R., Hakulinen, T., Soderman, B. e Verdecchia, A. (1999). Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, Vol.18, p. 441-454.
- [28] Dempster, A.P., Laird, N.M. e Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society, B*, Vol.39, p. 1-38.
- [29] Farewell, V.T. (1977). A model for a binary variable with time-censored observations. *Biometrika*, Vol.64, p. 43-46.
- [30] Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, Vol.38, p. 1041-1046.

- [31] Farewell, V.T. (1986). Mixture models in survival analysis: Are they worth the risk?. *The Canadian Journal of Statistics*, Vol.14, 3, p. 257-262.
- [32] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol.2. Wiley, New York.
- [33] Gamel, J.W., McLean, I.W. e Rosenberg, S.H. (1990). Proportion cured and mean log survival time as functions of tumour size. *Statistics in Medicine*, Vol.9, p. 999-1006.
- [34] Gather, U. e Pawlitschko, J. (1998). Estimating the survival function under a generalized Koziol-Green model with partially informative censoring. *Metrika*, Vol.48, p. 189-207.
- [35] Gaynor, J.J., Feuer, E.J., Tan, C.C., Wu, D.H., Little, C.R., Straus, B.C. e Brennan, M.F. (1993). On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association*, Vol.88, p. 400-409.
- [36] Ghahramani, M., Dean, C.B. e Spinelli, J.J. (2001). Simultaneous modelling of operative mortality and long-term survival after coronary artery bypass surgery. *Statistics in Medicine*, Vol.20, p. 1931-1945.
- [37] Ghitany, M.E. e Maller, R., Zhou, S. (1995). Estimating the proportion of immunes in censored samples: a simulation study. *Statistics in Medicine*, Vol.14, p. 39-49.
- [38] Gieser, P.W., Chang, M.N., Rao, P.V., Shuster, J.J. e Pullen, J. (1998). Modelling cure rates using the Gompertz model with covariate information. *Statistics in Medicine*, Vol.17, p. 831-839.
- [39] Goldman, A.I. (1984). Survivorship analysis when cure is a possibility: a Monte Carlo study. *Statistics in Medicine*, Vol.3, p. 153-163.
- [40] Gooley, T.A., Leisenring, W., Crowley, J. e Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, Vol.18, p. 695-706.

- 
- [41] Gordon, N.H. (1990). Application of the theory of finite mixtures for the estimation of “cure” rates of treated cancer patients. *Statistics in Medicine*, Vol.9, p. 397-407.
- [42] Gray, R.J. e Tsiatis, A.A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics*, Vol.45, p. 899-904.
- [43] Greenwood, M. (1926). A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects*, Vol.33, p. 1-26.
- [44] Halpern, J. e Brown, B.W. (1987). Cure rate models: Power of the log-rank and generalized Wilcoxon tests. *Statistics in Medicine*, Vol.6, p. 483-489.
- [45] Harrel Jr, F.E. (2001). *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression and Survival Analysis*. Springer-Verlag, New York.
- [46] Haybittle, J.L. (1965). A two-parameter model for the survival curve of treated cancer patients. *Journal of the American Statistical Association*, Vol.53, p. 16-26.
- [47] Haybittle, J.L. (1983). Is breast cancer ever cured?. *Reviews on Endocrine-Related Cancer*, Vol.14, p. 13-18.
- [48] Hosmer, D. W., Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley, New York.
- [49] Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, Vol.71, p. 75-83.
- [50] Hougaard, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability*, Vol.28, p. 695-701.
- [51] Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, Vol.1, p. 255-273.
- [52] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.

- [53] Hougaard, P., Mylegaard, P. e Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*, Vol.50, p. 1178-1188.
- [54] Ihaka, R. e Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, Vol.5, p. 299-314.
- [55] Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*. Lecture Notes in Statistics, **9**. Springer-Verlag, New York.
- [56] Klein, J.P., Moeschberger, M.L. (1998). *Survival Analysis. Techniques for Censored and Truncated Data*, 2ª impressão corrigida. Springer-Verlag, New York.
- [57] Koziol, J.A. e Green, S.B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika*, Vol.63, p. 465-474.
- [58] Kuk, A.Y.C. e Chen, C. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, Vol.79, p. 531-541.
- [59] Lang, K. (1995). A gradient algorithm locally equivalent to the EM algorithm. *Journal of Royal Statistical Society, B*, Vol.57, p. 425-437.
- [60] Larson, M. e Dinse, G.E. (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics*, Vol.45, p. 165-173.
- [61] Laska, E.M. e Meisner, M.J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics*, Vol.48, p. 1223-1234.
- [62] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley, New York.
- [63] Liu, C. e Rubin, D.B. (1995). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*, Vol.81, p. 633-648.
- [64] Longini, I.M. e Halloran, M.E. (1996). A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*, Vol.81, p. 633-648.

- 
- [65] Maller, R.A. e Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, Vol.79, p. 731-739.
- [66] Maller, R.A. e Zhou, S. (1994). Testing for sufficient follow-up and outliers in survival data. *Journal of the American Statistical Association*, Vol.89, 428, p. 1499-1506.
- [67] Maller, R.A. e Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, Vol.51, p. 1197-1205.
- [68] Maller, R.A. e Zhou, S. (1996). *Survival Analysis with Long-Term Survivors*. John Wiley, New York.
- [69] Maller, R.A. e Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica*, Vol.12, p. 725-750.
- [70] Manton, G.M. e Stallard, E., Vaupel, J.W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, Vol.81, p. 635-644.
- [71] Marubini, E. e Valsecchi, M.G. (1996). *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley, New York.
- [72] McLachlan, G.J. e Krishnan, T. (1997). *The EM Algorithm and Extensions*. John Wiley, New York.
- [73] McLachlan, G. e Peel, D. (2000). *Finite Mixture Models*. John Wiley, New York.
- [74] Meng, X.-L. e Pedlow, S. (1992). EM: a bibliographic review with missing articles. Em *American Statistical Association 1992 Proceedings of the Section on Statistical Computing*. American Statistical Association, Alexandria, p. 24-27.
- [75] Meng, X.-L. e Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, Vol.80, p. 267-278.
- [76] Moeschberger, M.L. e David, H.A. (1971). Life tests under competitive causes of failure and the theory of competing risks. *Biometrics*, Vol.27, p. 909-933.

- [77] Nagelkerke, N.J.D., Borgdorff, M.W. e Kim, S.J. (2001). Logistic discrimination of mixtures of *M. tuberculosis* and non-specific tuberculin reactions. *Statistics in Medicine*, Vol.20, p. 1113-1124.
- [78] Ng, S.K. e McLachlan, G.J. (1998). On modifications to the long-term survival mixture model in the presence of competing risks. *Journal of Statistics Computing Simulation*, Vol.61, p. 77-96.
- [79] Ng, S.K. e McLachlan, G.J. (2003). An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Statistics in Medicine*, Vol.22, p. 1097-1111.
- [80] Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, Vol.84, p. 487-493.
- [81] Oakes, D. (1993). A note on the Kaplan-Meier estimator. *The American Statistician*, Vol.47, p. 39-40.
- [82] Peng, Y. e Dear, K.B.G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, Vol.56, p. 237-243.
- [83] Peng, Y., Dear, K.B.G. e Carriere, K.C. (2001). Testing for the presence of cured patients: a simulation study. *Statistics in Medicine*, Vol.20, p. 1783-1796.
- [84] Peng, Y., Dear, K.B.G. e Denham, J.W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, Vol.17, p. 813-830.
- [85] Peterson, A.V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: application to competing risks. *Proceedings of the National Academy of Science, U.S.A.*, Vol.12, p. 11-13.
- [86] Phillips, N., Coldman, A. e McBride, M.L. (2002). Estimating cancer prevalence using mixture models for cancer survival *Statistics in Medicine*, Vol.21, p. 1257-1270.
- [87] Pierce, D.A., Stewart, W.H. e Kopecky, K.J. (1979). Distribution-free regression analysis of grouped survival data. *Biometrics*, Vol.35, p. 785-793.

- 
- [88] Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T. e Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, Vol.34, p. 541-554.
- [89] Price, D.L. e Manatunga, A.K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, Vol.20, p. 1515-1527.
- [90] Redner, R.A. e Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, Vol.26, p. 195-239.
- [91] Rocha, C.S. (1995). *Modelos com Fragilidade em Análise de Sobrevivência*. Tese de Doutorado. Faculdade de Ciências da Universidade de Lisboa.
- [92] Rohatgi, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley, New York.
- [93] Shao, Q. e Zhou, X. (2004). A new parametric model for survival data with long-term survivors. *Statistics in Medicine*, Vol.23, p. 3525-3543.
- [94] Sposto, R., Sather, H.N. e Baker, S.A. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, Vol.48, p. 87-99.
- [95] Sposto, R. (2002). Cure model analysis in cancer: an application to data from Children's Cancer Group. *Statistics in Medicine*, Vol.21, p. 293-312.
- [96] Steele, F. (2003). A discrete-time multilevel mixture model for event history data with long-term survivors, with an application to an analysis of contraceptive sterilisation in Bangladesh. *Lifetime Data Analysis*, Vol.9, p. 155-174.
- [97] Sy, J.P. e Taylor, J.M.G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, Vol.56, p. 227-236.
- [98] Tamura, R.N., Faries, D.E. e Feng, J. (2000). Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer-von Mises test. *Statistics in Medicine*, Vol.19, p. 2169-2184.

- [99] Taylor, J.M.G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, Vol.51, p. 899-907.
- [100] Tsiatis, A.A. (1975). A non-identifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Science, U.S.A.*, Vol.72, p. 20-22.
- [101] Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, Vol.54, p. 1508-1516.
- [102] Tsodikov, A. (2002). Semi-parametric models of long-and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Statistics in Medicine*, Vol.21, p. 895-920.
- [103] Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of Royal Statistical Society, B*, Vol.65, p. 759-774.
- [104] Tsodikov, A., Ibrahim, J.G. e Yakovlev, A.Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, Vol.98, p. 1063-1078.
- [105] Tsodikov, A., Loeffler, M. e Yakovlev, A. (1998). A cure model with time-changing risk factor: an application to the analysis of secondary leukaemia. *Statistics in Medicine*, Vol.17, p. 27-40.
- [106] Vaupel, J.W., Manton, K.G. e Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, Vol.16, p. 439-454.
- [107] Venables, W.N., Smith, D.M e R Development Core Team (2002). *An Introduction to R*. <http://cran.r-project.org/>
- [108] Yakovlev, A.Y., Asselain, B., Bardou, V.J. Fourquet, A., Hoang, R., Rochefediere, A., e Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer. Em *Biometrie et Analyse de Données Spatio-Temporelles*. Editores: Asselain, B., Boniface, M., Duby, C., Lopez, C., Masson, J.P. e Tranchefort, J., Vol.12, p. 66-82. Societ e Franaise de Biom etrie, ENSA Rennes, France.

- [109] Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of “permanent employment” in Japan. *Journal of the American Statistical Association*, Vol.87, p. 284-292.
- [110] Yau, K.K.W. e Ng, A.S.K. (2001). Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in Medicine*, Vol.20, p. 1591-1607.
- [111] Yu, B., Tiwari, R.C., Cronin, K.A. e Feuer, E.J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, Vol.23, p. 1733-1747.
- [112] Zhang, H. e Rao, M.B. (2004). On generalized maximum likelihood estimation in the proportional hazards model with partially informative censoring. *Metrika*, Vol.59, p. 125-136.