**Extensões em modelos de sobrevivência
com fração de cura e efeitos aleatórios**

Diego Ignacio Gallardo Mateluna

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Heleno Bolfarine

Coorientador: Prof. Dr. Antonio Carlos Pedroso de Lima

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro de Conicyt
(Comisión Nacional de Investigación, Ciencia y Tecnología - Chile)

São Paulo, Fevereiro de 2014

Extensões em modelos de sobrevivência com fração de cura e efeitos aleatórios

Esta tese contém as correções e alterações
sugeridas pela Comissão Julgadora durante a defesa
realizada por Diego Ignacio Gallardo Mateluna em 03/02/2014.

O original encontra-se disponível no Instituto de
Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Heleno Bolfarine (orientador) - IME-USP
- Prof. Dr. Antonio Carlos Pedroso de Lima - IME-USP
- Prof. Dr. Mário de Castro Andrade Filho - ICMC-USP
- Prof. Dr. Francisco Louzada Neto - ICMC-USP
- Prof. Dr. Filidor Edilson Vilca Labra - IMECC-UNICAMP

Agradecimentos

A Deus, por fazer meu caminho tão expedito em todo momento. Por dar-me saúde e a força necessária em todos estes anos de Doutorado.

Agradeço aos meus pais, que mesmo não entendendo muito bem o que é Estatística nem o que é um Doutorado, sentem orgulho do que faço. Agradeço esse amor e apoio incondicional, embora tivermos um par de quilômetros de distância durante meu Doutorado. (Digamos, uns 2.600 quilômetros). *También agradezco a la Wely, por su constante preocupación por mí. Los amo a todos.*

Ao professor Heleno, por ter aceitado orientar-me e pelas valiosas sugestões ao longo deste trabalho. Sem dúvida, este trabalho teve uma abordagem diferente com sua contribuição.

Agradeço ao Antonio Carlos pela paciência nos momentos mais difíceis durante o desenvolvimento deste trabalho e pela motivação a seguir trabalhando constantemente. Agradeço também as contribuições metodológicas e por sempre ser muito crítico comigo.

À banca julgadora, por ler e corrigir esta versão da tese. Certamente suas sugestões melhoraram a apresentação deste trabalho.

Gisela também merece uma menção nesta página, por ter-me ajudado em questões burocráticas na minha chegada em São Paulo e pela orientação em relação à mudança de orientador. Certamente a sugestão do Antonio Carlos foi acertada.

A mi polola y amiga Yolanda. Por mantener mi cordura en aquellos momentos más difíciles. Por la paciencia. Por tu amor y apoyo. Sin duda, la finalización de este trabajo hubiese sido mucho más aburrida y difícil sin tu presencia.

À turma de amigos chilenos aqui em São Paulo: Alvaro Cerda, Mónica Pavez, Carlos Montenegro, Manuel González, Natalia Milla, Andrés Donoso, Alejandra Tapia, Pedro Cortés, Marco Riquelme, Cristian Villegas, Nilton Lincopan, Maritza Fariás e Miguel Murúa. Ter uma parte do meu querido **CHILE** aqui em São Paulo ajudou naqueles momentos de saudade máxima.

Agradeço também à turma de amigos do IME: Tiago Magalhães, Aldo Medina, Francielle Lima, Elisa González, Lina Thomas, Jorge Torrejón, Eliardo Guimarães, pelos gratos momentos compartilhados e pelas discussões nas diferentes disciplinas cursadas em conjunto.

Não posso deixar de agradecer também aos meus amigos que ficaram no Chile: Bastian Pala-

cios, Maria Angélica Fuentes, Irma Coronado, Nataly Olivares, Carolina Aguilar, Mirko Aranda, Jaime Gómez, Hugo Peralta e Rony Palacios. Suas mensagens de apoio e longas conversas sempre alegraram meus dias.

Ao Pepe, por dar-me essa motivação inicial necessária para cursar meus estudos de Doutorado. Estarei agradecido por sempre.

Finalmente, agradeço o apoio financeiro por parte de Conicyt (*Comisión Nacional de Investigación, Ciencia y Tecnología*) através de seu programa *BECAS CHILE para Doctorado en el extranjero*, por ter financiado a totalidade deste Doutorado.

Resumo

Neste trabalho são apresentadas algumas extensões de modelos de sobrevivência com fração de cura, assumindo o contexto em que as observações estão agrupadas. Dois efeitos aleatórios são incorporados para cada grupo: um para explicar o efeito no tempo de sobrevida das observações suscetíveis e outro para explicar a probabilidade de cura. Apresenta-se uma abordagem clássica através dos estimadores REML e uma abordagem bayesiana através do uso de processos de Dirichlet. Discute-se alguns estudos de simulação em que avalia-se o desempenho dos estimadores propostos, além de comparar as duas abordagens. Finalmente, ilustram-se os resultados com dados reais.

Palavras-chave: modelo de tempos de promoção, modelos destrutivos, modelos mistos, processos Dirichlet.

Abstract

In this work some extensions in survival models with cure fraction are presented, assuming the context in which the observations are grouped into clusters. Two random effects are incorporated for each group: one to explain the effect on survival time of susceptible observations and another to explain the probability of cure. A classical approach through the REML estimators is presented as well as a bayesian approach through Dirichlet Process. Besides comparing both approaches, some simulation studies which evaluates the performance of the proposed estimators are discussed. Finally, the results are illustrated with a real database.

Keywords: promotion time cure rate model, destructive models, random effect models, Dirichlet Process.

Sumário

Lista de Abreviaturas	ix
Lista de Figuras	xi
Lista de Tabelas	xv
1 Motivação	1
1.1 Introdução	1
1.2 Descrição dos objetivos	3
1.3 Organização do trabalho	3
2 Preliminares	5
2.1 Algumas distribuições de interesse	5
2.1.1 Distribuição Weibull	5
2.1.2 Distribuição Exponencial Segmentada	8
2.1.3 Distribuição Wishart e Wishart-Invertida	10
2.2 Modelos de sobrevivência com fração de cura	11
2.2.1 Modelo de mistura padrão	11
2.2.2 Modelo de tempos de promoção	13
2.2.3 Extensões do modelo de Chen	16
2.2.4 Modelos destrutivos de Poisson ponderado com fração de cura	18
2.2.5 Outros modelos de cura	26
2.2.6 Comentários adicionais sobre os modelos de cura	28
2.3 Estimadores BLUP e REML	29
2.4 Processos de Dirichlet	31
2.4.1 Métodos de amostragem para processos de Dirichlet	35
2.4.2 Processos de Dirichlet centrados	39
3 Modelos de sobrevivência com fração de cura e efeitos aleatórios	41
3.1 Modelo de mistura com efeitos aleatórios	41
3.2 Modelo de Chen com efeitos aleatórios	46
3.2.1 Estudo de simulação: Recuperação de parâmetros	49
3.2.2 Comparação do modelo de Chen com efeitos aleatórios normais dependentes versus independentes	66
3.3 Modelo de mistura com outras ligações na cura	78
3.3.1 Estudo de simulação: Recuperação de parâmetros	79
3.3.2 Estimação da fração de cura usando as diferentes ligações	83

4	Modelos destrutivos de Poisson ponderado com fração de cura e efeitos aleatórios	85
4.1	O modelo	85
4.2	Abordagem clássica	86
4.3	Abordagem bayesiana	88
4.4	Estudo de simulação: comparação das abordagens clássica e bayesiana nos MDPPFC com efeitos aleatórios	91
5	Aplicação	105
5.1	Apresentação do banco de dados	105
5.2	Ajuste clássico dos modelos	108
5.3	Ajuste bayesiano dos modelos	122
6	Conclusões	129
6.1	Considerações finais	129
6.2	Sugestões para pesquisas futuras	129
A	Apêndice	131
A.1	Alguns aspectos dos modelos destrutivos	131
A.1.1	Cálculo de $P(D = d \mid \theta, p, \phi)$	131
A.1.2	Calculando a distribuição de D, M e $D \mid M = m$ dada a informação observada	132
A.1.3	Provando que a fração de cura do modelo DBLP não é identificável quando $\mathbf{z}_1$ e $\mathbf{z}_2$ incluem intercepto	135
A.2	Derivadas necessárias na implementação do algoritmo de Newton-Raphson	136
A.2.1	Modelo de Chen com efeitos aleatórios	136
A.2.2	Modelo de mistura com efeitos aleatórios	138
A.2.3	Modelos Destrutivos com efeitos aleatórios	140
A.3	Programas em R	141
A.3.1	Estimação clássica dos modelos de Chen e dos modelos destrutivos com efeitos aleatórios bivariados Normais	141
	Referências Bibliográficas	155

Lista de Abreviaturas

BLUP	melhor estimador linear não viciado (Best Linear Unbiased Predictor)
DLBP	modelo destrutivo de Poisson com viés de duração (Destructive length-biased Poisson)
e.a.(’s)	efeito(s) aleatório(s)
E.Q.M.	erro quadrático médio
f.d.a.	função de distribuição acumulada
f.d.p.	função de densidade de probabilidade
i.i.d.	independentes e identicamente distribuídos(as)
K-M	(estimador de) Kaplan-Meier
MDPPFC	modelo destrutivo de Poisson Ponderado com fração de cura
M.P.	medida de probabilidade
P.D.(’s)	processo(s) Dirichlet
P.D.C.(’s)	processo(s) Dirichlet Centrado(s)
q.c.	quase certamente
REML	máxima verossimilhança residual (Residual Maximum Likelihood)
v.a.(’s)	variável(variáveis) aleatória(s)

Lista de Figuras

1.1	Curva de K-M para pacientes com câncer de língua.	2
2.1	Função de densidade do modelo Weibull ($\lambda = 1$).	6
2.2	Função de sobrevivência do modelo Weibull ($\lambda = 1$).	7
2.3	Função de risco do modelo Weibull ($\lambda = 1$).	7
2.4	Função de densidade do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$	9
2.5	Função de sobrevivência do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$	9
2.6	Função de risco do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$	10
2.7	Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha\mathcal{G}_0)$, com $\alpha = 1$ e $\mathcal{G}_0$ a respectiva medida basal.	32
2.8	Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha\mathcal{G}_0)$, com $\alpha = 10$ e $\mathcal{G}_0$ a respectiva medida basal.	33
2.9	Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha\mathcal{G}_0)$, com $\alpha = 100$ e $\mathcal{G}_0$ a respectiva medida basal.	33
2.10	Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha\mathcal{G}_0)$, com $\alpha = 1000$ e $\mathcal{G}_0$ a respectiva medida basal.	34
3.1	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo I	54
3.2	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo II	55
3.3	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo III	56
3.4	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo IV	57
3.5	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo I	62
3.6	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo II	63
3.7	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo III	64

3.8	Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo IV	65
3.9	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo I	70
3.10	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo II	71
3.11	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo III	72
3.12	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo I	75
3.13	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo II	76
3.14	Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo III	77
3.15	Histograma dos parâmetros estimados no modelo de mistura com efeitos aleatórios usando ligação complementar log-log para a cura e sobrevivência basal Weibull	82
4.1	Funções de densidade (marginais) das distribuições seleccionadas para os e.a.'s. . . .	93
5.1	Curva de K-M para as seis clínicas que participavam no estudo.	106
5.2	Curva de K-M para a covariável recategorizada T -stage.	107
5.3	Curva de K-M para a variável Idade (categorizada).	107
5.4	Função de risco acumulada estimada	108
5.5	Fração de cura estimada para pacientes da clínica 2	110
5.6	Função de sobrevivência estimada via K-M e pelo modelo escolhido para a clínica 2 .	110
5.7	Curva de K-M para os resíduos de Cox-Snell do modelo escolhido.	111
5.8	Função de sobrevivência estimada em cada clínica	111
5.9	Crítérios cAIC e cBIC para o modelo DBN (caso 1).	112
5.10	Função de sobrevivência populacional estimada na clínica 2 para diferentes modelos de cura e diferentes perfis de pacientes	115
5.11	Função de sobrevivência estimada para os indivíduos suscetíveis da clínica 2 para diferentes modelos de cura e diferentes perfis de pacientes	116
5.12	Fração de cura estimada na clínica 2 para os diferentes modelos ajustados	117
5.13	Estimador de K-M para os resíduos de Cox-Snell para os modelos com e.a.'s ajustados	118
5.14	Diferentes componentes estimadas para os modelos de Chen, DLBP e DEWPM com e.a.'s	119

5.15	Diferentes componentes estimadas para os modelos DLBP e DEWPM com e.a.'s em todos os pacientes do estudo	120
5.16	Função de log-verossimilhança perfilada para: (a) β_{11} no modelo DLBP com e.a.'s, (b) β_{11} no modelo DEWPM com e.a.'s, (c) V_1 no modelo DLBP com e.a.'s (d) e no modelo DEWPM com e.a.'s.	121
5.17	Função de sobrevivência populacional da clínica 2 para os modelos DLBP e DEWPM com e.a.'s e diferentes perfis	125
5.18	Fração de cura estimada na clínica 2	126
5.19	Intervalos de credibilidade ao 95% para o número inicial de células, número de células ativadas e probabilidade de ativação das células para alguns pacientes.	127
5.20	Função de densidade estimada baseada na amostra <i>a posteriori</i>	128

Lista de Tabelas

1.1	Relação entre as funções de interesse na análise de sobrevivência.	2
3.1	Vício estimado dos estimadores REML no modelo de Chen com efeitos aleatórios dependentes: caso sobrevivência basal Weibull.	51
3.2	Vício estimado dos estimadores REML no modelo de Chen com efeitos aleatórios dependentes: caso sobrevivência Exponencial Segmentada ($L = 3$).	59
3.3	Vício e Desvio Padrão estimado dos estimadores REML no modelo de Chen com efeitos aleatórios independentes e dependentes: caso sobrevivência Weibull.	68
3.4	Vício e Desvio Padrão estimado dos estimadores REML no modelo de Chen com efeitos aleatórios independentes e dependentes: caso sobrevivência Exponencial Segmentada ($L = 3$).	73
3.5	Ligações para modelar a proporção de não curados	78
3.6	Vício estimado para os estimadores REML baseados em 500 réplicas (Função de sobrevivência dos não curados considerada do modelo Weibull).	81
3.7	Vício estimado e vício máximo estimado para a fração de cura usando as 3 ligações propostas.	84
4.1	Funções de peso para cada modelo Destrutivo.	85
4.2	Funções de sobrevivência e risco para os modelos DLBP, EWPM e DBN.	87
4.3	Comparação das estimativas dos efeitos fixos nos MDPPFC e efeitos aleatórios usando abordagem Clássica e bayesiana.	95
4.4	Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por caso).	101
4.5	Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por parâmetro).	101
4.6	Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por combinação de número de clínicas e pacientes em cada).	101
4.7	Análise comparativa para os efeitos aleatórios entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por caso).	103
4.8	Análise comparativa para os efeitos aleatórios entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por combinação de número de clínicas e pacientes em cada).	103
5.1	Distribuição de clínica e da variável T -stage categorizada.	106
5.2	Crerios de seleção para diferentes ligações no modelo de mistura com e.a.'s	109

5.3	Estimativas para o modelo de mistura com e.a.'s dependentes usando ligação cloglog na cura e função de sobrevivência basal Weibull	109
5.4	Combinações de covariáveis e e.a.'s na cura consideradas para os modelos Destrutivos.	112
5.5	Crîtérios cAIC e cBIC para os modelos de Chen e Destrutivos ajustados.	113
5.6	Estimativas obtidas para os modelos escolhidos usando o critério cAIC e cBIC.	114
5.7	Distribuições <i>a priori</i> consideradas para α	122
5.8	Valores de $\sqrt{\widehat{R}}$ nos modelos ajustados.	123
5.9	Fator de Bayes para os modelos ajustados. Na diagonal, os valores da função de log-verossimilhança marginais para cada modelo. Na parte abaixo da diagonal, o logaritmo do fator de Bayes do modelo da respectiva linha versus o modelo da respectiva coluna.	123
5.10	Estimativas obtidas para o modelo DLBP e DEWPM com e.a.'s normais usando a abordagem bayesiana.	124
5.11	Crîtério DIC para o modelos DLBP e DEWPM com e.a.'s.	127
A.1	F.d.a. das distribuições Logística, Normal e de Gumbel e suas derivadas de primeiro e segundo ordem.	138

Capítulo 1

Motivação

1.1 Introdução

A análise de sobrevivência é utilizada em estudos em que a variável em estudo é o tempo até a ocorrência de algum evento de interesse (que será denotado por T). Desta forma, a função de sobrevivência $S(\cdot)$ avaliada em t define-se como

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du = 1 - F(t),$$

em que $f(\cdot)$ e $F(\cdot)$ denotam a função de densidade e a f.d.a. de T respectivamente. Uma propriedade básica da função de sobrevivência (pelo menos na análise de sobrevivência usual) é o fato que ela é própria, i.e.,

$$\lim_{t \rightarrow \infty} S(t) = 0. \quad (1.1)$$

Em termos práticos, a equação (1.1) indica que se o tempo de seguimento do estudo for suficientemente longo, então espera-se observar o evento de interesse em todos os indivíduos da população.

Uma outra função de interesse no contexto de análise de sobrevivência é a função de risco, conhecida na literatura também como taxa de falha e que será denotada por $h(\cdot)$, a qual é definida por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t}.$$

A função de risco pode ser interpretada como a probabilidade aproximada do evento de interesse ocorrer no instante imediatamente posterior a t dado que o indivíduo sobreviveu até t .

Finalmente, a função de risco acumulado $H(\cdot)$ é definida por

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

A função de risco acumulado pode ser interpretada como o risco de ocorrência do evento de interesse até o instante t , dado que o evento não aconteceu até instante t .

Há uma relação um a um entre as funções $f(\cdot)$, $S(\cdot)$, $h(\cdot)$ e $H(\cdot)$ discutidas anteriormente. Desta forma, dada uma delas pode-se encontrar todas as outras. Este fato também implica que pode-se fazer a modelagem de qualquer uma delas, definindo um modelo único. A Tabela 1.1 apresenta a relação entre as funções mencionadas anteriormente.

Os modelos com fração de cura são modelos *ad hoc* em estudos em que o evento de interesse não necessariamente manifesta-se em todos os indivíduos da população. Esta situação implica que

Função obtida	Função requerida			
	$f(t)$	$S(t)$	$h(t)$	$H(t)$
$f(t)$	-	$\int_t^\infty f(u)du$	$\frac{f(t)}{\int_t^\infty f(u)du}$	$-\log \left(\int_t^\infty f(u)du \right)$
$S(t)$	$-\frac{dS(t)}{dt}$	-	$-\frac{[dS(t)/dt]}{S(t)}$	$-\log S(t)$
$h(t)$	$h(t) \exp \left\{ -\int_0^t h(u)du \right\}$	$\exp \left\{ -\int_0^t h(u)du \right\}$	-	$\int_0^t h(u)du$
$H(t)$	$\frac{dH(t)}{dt} \exp \{ -H(t) \}$	$\exp \{ -H(t) \}$	$\frac{dH(t)}{dt}$	-

Tabela 1.1: Relação entre as funções de interesse na análise de sobrevivência.

a função de sobrevivência da população não necessariamente satisfaça a equação em (1.1). Por exemplo, muitos tipos de câncer podem ser encaixados nestas circunstâncias. Kalbfleisch e Prentice [2002] apresentam um banco de dados referente a câncer de língua localizado em três zonas diferentes dela. Considerando só os pacientes com câncer de língua na parte posterior dela, construiu-se a curva de Kaplan-Meier da Figura 1.1.

Figura 1.1: Curva de K-M para pacientes com câncer de língua.

Note que o estimador de K-M estabiliza-se em um valor positivo (próximo de 0.2), sugerindo que a função de sobrevivência desta população (pelo menos olhada sem considerar eventuais covariáveis) não satisfaz a condição de ser própria como na equação (1.1).

A literatura para os modelos com fração de cura tem crescido bastante nas últimas décadas. Dentre esses modelos, o modelo de mistura (Berkson e Gage [1952]) tem sido o mais utilizado sobretudo por conta da facilidade na interpretação dos coeficientes estimados nesse modelo. As interpretações neste modelo só podem ser feitas em relação à probabilidade de ser curado ou da função de sobrevivência dos indivíduos suscetíveis para um indivíduo com determinadas características, o que é suficiente em muitos contextos de fração de cura.

Porém, existem outros modelos que podem fornecer ainda mais informação da mencionada anteriormente. Por exemplo, no contexto de estudos de câncer, o modelo de tempos de promoção, discutido em Yakolev e Tsodikov [1996] e Chen et al. [1999a], permite também fazer inferência em relação à distribuição das células que produzem o câncer. Já no caso dos modelos destrutivos (Rodrigues et al. [2011]), introduz-se no modelo a suposição de existência de células que podem ou não produzir câncer (algumas delas podem ser destruídas). Desta forma, nestes modelos pode-se fazer inferência tanto da distribuição das células que podem produzir o câncer quanto da probabilidade de ativação de cada célula.

Por outro lado, no contexto em que os pacientes foram tratados em diferentes clínicas, é natural supor que cada clínica pode ter influência sobre os pacientes que tratam. Porém, no contexto de fração de cura há dois grupos de pacientes: suscetíveis e curados. Portanto, pode-se pensar que cada clínica influi tanto nos indivíduos suscetíveis quanto nos indivíduos curados. Considerando que as clínicas só representam uma a.a. de clínicas, então é mais apropriado considerar um modelo de efeitos aleatórios, incluindo para cada clínica um efeito aleatório associado aos indivíduos suscetíveis (U_j) e um efeito aleatório associado aos indivíduos curados (V_j), em que $j = 1, \dots, J$, e J representa o número de clínicas.

Neste contexto, foram encontrados na literatura só os trabalhos de Li et al. [2001] e Lai e Yau [2008], ambos para o modelo de mistura e assumindo que os pares (U_j, V_j) tem distribuição normal em sua versão independente e dependente respectivamente, além do trabalho de Lopes e Bolfarine [2012], em que foi trabalhado o modelo de Chen com e.a.'s, mas incluindo efeitos só na parte associada à cura (V_j). Este trabalho foi feito de um ponto de vista clássico e bayesiano. Mas como mencionado anteriormente, existem outros modelos de cura que podem ser mais informativos e mais complexos, sobretudo no contexto de estudos de câncer.

1.2 Descrição dos objetivos

O objetivo geral deste trabalho é propor extensões em modelos com fração de cura, incluindo efeitos aleatórios nos modelos de Chen et al. [1999a] e os modelos destrutivos de Poisson ponderado propostos por Rodrigues et al. [2011]. Será considerada tanto uma perspectiva clássica quanto uma bayesiana.

Neste trabalho, procura-se também implementar as metodologias desenvolvidas através de simulações e aplicação a dados reais, comparando as metodologias clássicas e bayesianas. Para atingir o objetivo proposto, definimos os objetivos específicos desta tese:

1. Propor um método de estimação clássico no modelo de Chen et al. [1999a] com efeitos aleatórios (U_j, V_j) , $j = 1, \dots, J$ como explicados anteriormente.
2. Estudar o impacto nas estimativas de considerar de forma errônea o modelo de Chen et al. [1999a] com efeitos aleatórios dependentes quando deveria ser usado o modelo com efeitos aleatórios independentes e viceversa.
3. Propor métodos de estimação clássico e bayesiano nos modelos destrutivos (Rodrigues et al. [2011]) com efeitos aleatórios (U_j, V_j) , $j = 1, \dots, J$.
4. Estudar vantagens e desvantagens dos dois métodos propostos no item anterior.

1.3 Organização do trabalho

Este trabalho contém seis capítulos. O capítulo 2 é um capítulo introdutório, que apresenta algumas distribuições de grande importância no desenvolvimento deste trabalho. Também faz-se

uma revisão bibliográfica dos modelos com fração de cura e as principais características desses modelos. Apresenta-se uma breve revisão dos métodos BLUP (*Best Linear Unbiased Predictor*) e REML (*Restricted Maximum Likelihood*) com a finalidade de facilitar seu uso em futuras secções. Finalmente, faz-se uma introdução aos processos de Dirichlet, os que serão utilizados para fazer uma modelagem não paramétrica da distribuição dos efeitos aleatórios.

No capítulo 3 apresentamos brevemente o modelo de mistura com efeitos aleatórios normais independentes e dependentes propostos em Li et al. [2001] e Lai e Yau [2008], discutindo também a metodologia clássica baseada nos estimadores REML. Depois, é desenvolvida uma abordagem similar para o modelo de Chen et al. [1999a] com efeitos aleatórios normais. São apresentados dois estudos de simulação: o primeiro para avaliar se dado um conjunto de parâmetros o método proposto consegue estimá-los e outro em que avalia-se o ganho ou perda de usar a versão bivariada da normal *versus* a versão independente para os efeitos aleatórios. Também foi incluída uma discussão do modelo de mistura usando outras ligações na cura, como são a ligação probito e complementar log-log.

No capítulo 4 desenvolvemos uma metodologia clássica similar à do capítulo anterior para os modelos destrutivos com efeitos aleatórios baseada também nos estimadores REML. Também propõe-se uma metodologia bayesiana para esse modelo baseada no uso de processos de Dirichlet para os efeitos aleatórios, o que permite flexibilizar a distribuição utilizada para os efeitos aleatórios. As duas abordagens são comparadas através um estudo de simulação em que os efeitos aleatórios foram gerados de diferentes distribuições, incluindo a distribuição normal.

O capítulo 5 apresenta uma aplicação ao banco de dados real de Kalbfleisch e Prentice [2002] discutido no começo deste capítulo. Ilustra-se a vantagem interpretativa de utilizar os modelos destrutivos com efeitos aleatórios em relação aos modelos de Chen et al. [1999a] e de mistura com efeitos aleatórios.

Por último, o capítulo 6 discute as principais conclusões deste trabalho e perspectivas futuras de pesquisa.

Parte deste trabalho já foi aceito em periódicos de circulação internacional (Gallardo et al. [2013a] e Gallardo et al. [2013b]).

Capítulo 2

Preliminares

Este capítulo está centrado na apresentação das distribuições a serem utilizadas neste trabalho, algumas técnicas de estimação que serão utilizadas em futuros capítulos desenvolvidos e introduzir a classe de modelos com fração de cura no contexto de análise de sobrevivência.

Na seção 2.1 apresentam-se as distribuições Weibull e Exponencial Segmentada, distribuições muito importantes neste trabalho pois pertencem à reduzida classe de modelos paramétricos de riscos proporcionais, além de introduzir a distribuição Wishart e Wishart Invertida. Posteriormente, na seção 2.2 faz-se uma rápida revisão dos principais modelos com fração de cura na literatura. Na seção 2.3 são apresentados os estimadores BLUP (*Best Linear Unbiased Prediction*) e os estimadores REML (*Restricted Maximum Likelihood*) em um modelo de efeitos aleatórios, enunciando algumas propriedades deles. Finalmente, na seção 2.4 apresenta-se uma breve introdução aos Processos Dirichlet (P.D.'s), incluindo alguns métodos de amostragem para estes processos em modelos conjugados e não conjugados, além de introduzir a noção de Processos Dirichlet Centrados (P.D.C.'s).

2.1 Algumas distribuições de interesse

Esta seção visa ajudar ao leitor ao entendimento de duas distribuições utilizadas neste trabalho: as distribuições Weibull e Exponencial Segmentada. Serão indicadas algumas características e propriedades de cada uma delas, além de unificar a notação das duas distribuições, ajudando ao desenvolvimento deste trabalho. Também serão introduzidas as distribuições Wishart e Wishart-Invertida, esta última de grande relevância na Inferência bayesiana pois permite a conjugação da matriz de variância-covariância no modelo Normal.

2.1.1 Distribuição Weibull

A distribuição Weibull tem sido amplamente utilizada na análise de sobrevivência. Foi apresentada originalmente por Weibull [1939] e utilizada na análise de sobrevivência posteriormente por Weibull [1951]. Desde então, tem sido extensamente utilizada como modelo em aplicações biométricas, industriais e muitas outras áreas. Uma das principais características desta distribuição na modelagem é sua flexibilidade em acomodar diferentes formas na taxa de falha (função de risco), apenas variando o valor do parâmetro de forma.

A variável T tem distribuição Weibull com parâmetro de forma $\nu > 0$ e parâmetro de escala $\lambda > 0$ (denotamos $T \sim \text{Weibull}(\lambda, \nu)$) se a função de densidade é dada por

$$f(t \mid \lambda, \nu) = \lambda \nu t^{\nu-1} \exp\{-\lambda t^\nu\}, \quad t > 0. \quad (2.1)$$

As funções de sobrevivência e de risco são dadas respectivamente por

$$\begin{aligned} S(t | \lambda, \nu) &= \exp\{-\lambda t^\nu\}, \quad t > 0, \quad \text{e} \\ h(t | \lambda, \nu) &= \lambda \nu t^{\nu-1}, \quad t > 0 \end{aligned} \quad (2.2)$$

Os dois primeiros momentos desta distribuição são dados por

$$\begin{aligned} E(T) &= \frac{1}{\lambda^{\frac{1}{\nu}}} \Gamma\left(1 + \frac{1}{\nu}\right), \quad \text{e} \\ E(T^2) &= \frac{1}{\lambda^{\frac{2}{\nu}}} \Gamma\left(1 + \frac{2}{\nu}\right). \end{aligned}$$

A distribuição Exponencial de parâmetro λ é o caso particular da distribuição Weibull quando $\nu = 1$. Note que, nesse caso, a função de risco é constante através do tempo. Os comportamentos das funções de densidade, de sobrevivência e de risco estão representadas nas Figuras 2.1, 2.2 e 2.3 respectivamente.

Observa-se na Figura 2.3 as diferentes formas da taxa de falha para a distribuição Weibull: taxa de falha decrescente ($\nu < 1$), taxa de falha constante ($\nu = 1$) e taxa de falha crescente ($\nu > 1$).

Figura 2.1: Função de densidade do modelo Weibull ($\lambda = 1$).

Figura 2.2: Função de sobrevivência do modelo Weibull ($\lambda = 1$).

Figura 2.3: Função de risco do modelo Weibull ($\lambda = 1$).

2.1.2 Distribuição Exponencial Segmentada

A distribuição Exponencial Segmentada foi introduzida por [Feigl e Zelen \[1965\]](#) para estudos com uma covariável e foi estendida por [Friedman \[1982\]](#) para o caso de múltiplas covariáveis. O modelo assume que o risco é constante dentro de algum intervalo de tempo. Considera-se uma partição do eixo do tempo tal que $0 = a_0 < a_1 < \dots < a_{L-1} < a_L = \infty$ e o risco é constante em cada um desses intervalos.

Suponha o vetor $\mathbf{a}$ conhecido, em que $\mathbf{a} = (a_1, \dots, a_{L-1})$, com $0 < a_1 < \dots < a_{L-1} < \infty$ e defina $a_0 = 0$ e $a_L = \infty$. A variável T tem distribuição Exponencial Segmentada de parâmetros $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_L)$ e partição $\mathbf{a}$ (denotamos $T \sim ES_{\mathbf{a}}(\boldsymbol{\lambda})$) se a função de densidade é dada por

$$f(t | \boldsymbol{\lambda}) = \kappa_l \lambda_l \exp\{-\lambda_l(t - a_{l-1})\}, \quad t \in (a_{l-1}, a_l], \quad l = 1, \dots, L, \quad (2.3)$$

em que

$$\kappa_l = \begin{cases} 1 & , \text{ se } l = 1, \\ \exp\left[-\sum_{i=1}^{l-1} \lambda_i(a_i - a_{i-1})\right] & , \text{ se } l = 2, \dots, L. \end{cases}$$

As funções de sobrevivência e de risco são dadas respectivamente por

$$\begin{aligned} S(t | \boldsymbol{\lambda}) &= \exp\left\{-\sum_{l=1}^L \lambda_l \nabla_l(t)\right\}, \quad t > 0, \quad \text{e} \\ h(t | \boldsymbol{\lambda}) &= \lambda_l, \quad t \in (a_{l-1}, a_l], \quad l = 1, \dots, L, \end{aligned} \quad (2.4)$$

em que

$$\nabla_l(t) = \begin{cases} 0, & \text{se } t < a_{l-1}, \\ t - a_{l-1}, & \text{se } a_{l-1} \leq t < a_l, \\ a_l - a_{l-1}, & \text{se } t \geq a_l \end{cases} \quad l = 1, \dots, L. \quad (2.5)$$

Os dois primeiros momentos desta distribuição são dados por

$$\begin{aligned} E(T) &= \sum_{l=1}^L \kappa_l \left(a_l + \frac{1}{\lambda_l}\right) (1 - \exp\{-\lambda_l(a_l - a_{l-1})\}), \quad \text{e} \\ E(T^2) &= \sum_{l=1}^L \kappa_l \left[a_{l-1}^2 + \frac{2a_{l-1}}{\lambda_l} + \frac{2}{\lambda_l^2} - \exp\{-\lambda_l(a_l - a_{l-1})\} \left(a_l^2 + \frac{2a_l}{\lambda_l} + \frac{2}{\lambda_l^2}\right)\right]. \end{aligned}$$

A distribuição Exponencial de parâmetro λ é o caso particular da distribuição Exponencial Segmentada quando $\lambda_l = \lambda$, para $l = 1, \dots, L$. Os comportamentos das funções de densidade, de sobrevivência e de risco estão representadas nas Figuras 2.4, 2.5 e 2.6 respectivamente. Na Figura 2.6 pode-se verificar que a função de risco da distribuição Exponencial Segmentada não é contínua nos pontos do vetor $\mathbf{a}$. Isso implica que o conhecimento do vetor $\mathbf{a}$ é muito importante. Porém, esta distribuição permite a modelagem de eventos em que a função de risco não é monótona (crescente, decrescente ou constante) como visto na figura 2.6 no caso em que $\lambda_1 = 1.7, \lambda_2 = 0.8, \lambda_3 = 1.7$ e a principal diferença com a Weibull é que a esta última distribuição é adequada só nos casos de eventos com função de risco monótona.

Na prática, o uso desta distribuição também está associado com o uso de uma abordagem semi-paramétrica. Por exemplo, veja-se [Pan e Kozakai \[2013\]](#).

Figura 2.4: Função de densidade do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$.

Figura 2.5: Função de sobrevivência do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$.

Figura 2.6: Função de risco do modelo Exponencial Segmentado com $\mathbf{a} = (2, 5)$.

Observação 2.1.1 As distribuições Weibull e Exponencial Segmentada serão utilizadas para a modelagem de tempos de sobrevivência. Para unificar a notação, $\boldsymbol{\lambda}$ representará o conjunto de parâmetros a estimar em cada distribuição. Isto é,

- Caso Weibull: $\boldsymbol{\lambda} = (\lambda, \nu)$.
- Caso Exponencial Segmentada: $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_L)$.

2.1.3 Distribuição Wishart e Wishart-Invertida

Esta distribuição foi formulada por primeira vez por Wishart [1928]. Sejam $\mathbf{X}_i$, $i = 1, \dots, n$ vetores q -dimensionais i.i.d. tal que $\mathbf{X}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$. Diz-se que a matriz aleatória (simétrica) $\mathbf{W}$ q -dimensional tem distribuição Wishart com matriz de escala $\boldsymbol{\Psi}$ e b graus de liberdade (denotamos

$\mathbf{W} \sim \mathcal{W}_q(b, \boldsymbol{\Psi})$) se e só se $\mathbf{W} = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i$. A f.d.p. da distribuição Wishart é dada por

$$\begin{aligned} f(\mathbf{w}) &= \frac{|\mathbf{w}|^{\frac{b-q-1}{2}}}{2^{\frac{bq}{2}} |\boldsymbol{\Psi}|^{\frac{b}{2}}} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Psi}^{-1} \mathbf{w}) \right\}, \\ &\propto |\mathbf{w}|^{\frac{b-q-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Psi}^{-1} \mathbf{w}) \right\}. \end{aligned}$$

Sejam $\mathbf{W}_1$ e $\mathbf{W}_2$ matrizes aleatórias independentes com distribuição $\mathcal{W}_{q_1}(b_1, \boldsymbol{\Psi}_1)$ e $\mathcal{W}_{q_2}(b_2, \boldsymbol{\Psi}_2)$ respectivamente. Algumas propriedades da distribuição Wishart são enunciadas a seguir.

- $E(\mathbf{W}_1) = b_1 \boldsymbol{\Psi}_1$.
- $\mathbf{W}_1(b, \sigma^2) \stackrel{\mathcal{D}}{=} \sigma^2 \chi_b^2$, em que χ_b^2 representa a distribuição Qui-quadrado com b graus de liberdade.

- Se $\mathbf{A}$ uma matriz $r \times q_1$, então $\mathbf{A}\mathbf{W}_1\mathbf{A}^\top \sim \mathcal{W}_r(b_1, \mathbf{A}\mathbf{\Psi}_1\mathbf{A}^\top)$.
- Se $\mathbf{a} \in \mathbb{R}^{q_1}$ um vetor não nulo, então $\frac{\mathbf{a}^\top \mathbf{W}_1 \mathbf{a}}{\mathbf{a}^\top \mathbf{\Psi}_1 \mathbf{a}} \sim \chi_{b_1}^2$.
- Se $q_1 = q_2 = q$ e $\mathbf{\Psi}_1 = \mathbf{\Psi}_2 = \mathbf{\Psi}$, então $\mathbf{W}_1 + \mathbf{W}_2 \sim \mathcal{W}_q(b_1 + b_2, \mathbf{\Psi})$.

Desde um ponto de vista Bayesiano, ainda mais importante é a distribuição Wishart Invertida, pois ela permite a conjugação da matriz de variância-covariância $\mathbf{\Sigma}$ no modelo $N(\mathbf{0}, \mathbf{\Sigma})$.

Diz-se que a matriz aleatória (simétrica) $\mathbf{W}$ q -dimensional tem distribuição Wishart Invertida com matriz de escala $\mathbf{\Psi}$ e b graus de liberdade (denotamos $\mathbf{W} \sim \mathcal{WI}_q(b, \mathbf{\Psi})$) se e só se $\mathbf{W}^{-1} \sim \mathcal{W}_q(b, \mathbf{\Psi}^{-1})$. A f.d.p. da distribuição Wishart Invertida é dada por

$$\begin{aligned} f(\mathbf{w}) &= \frac{|\mathbf{\Psi}|^{b/2} |\mathbf{w}|^{-(b+q+1)/2}}{2^{bq/2} \Gamma_q(b/2)} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{\Psi} \mathbf{w}^{-1}) \right\} \\ &\propto |\mathbf{w}|^{-(b+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{\Psi} \mathbf{w}^{-1}) \right\}, \end{aligned}$$

em que $\Gamma_q(\cdot)$ denota a função gamma q -variada. A esperança da distribuição Wishart Invertida é $E(\mathbf{W}) = \frac{\mathbf{\Psi}}{b-q-1}$ e a variância dos elementos da matriz $\mathbf{W}$ são dados por

$$\text{Var}(W_{ij}) = \begin{cases} \frac{2\psi_{ii}^2}{(b-q-1)^2(b-q-3)} & , \text{ se } i = j, \\ \frac{(b-q-1)(\psi_{ij}^2 + \psi_{ii}\psi_{jj})}{(b-q)(b-q-1)^2(b-q-3)} & , \text{ se } i \neq j. \end{cases}$$

2.2 Modelos de sobrevivência com fração de cura

Nos modelos de sobrevivência usuais assume-se que se o tempo de seguimento do estudo é suficientemente longo, então em todos os indivíduos da população acontecerá o evento de interesse. Na prática, há situações nas quais isso nem sempre acontece. Por exemplo, em alguns tipos de câncer, os avanços tecnológicos tem permitido que pacientes possam ter uma vida relativamente normal. Essa porcentagem de pacientes é conhecida na literatura como *fração de cura*.

Os modelos com fração de cura obtiveram uma grande evolução desde o trabalho pioneiro de Berkson e Gage [1952], também conhecido na literatura como modelo de mistura. Este capítulo apresenta esse modelo e algumas extensões que surgiram a partir dele.

2.2.1 Modelo de mistura padrão

Suponha que a função de sobrevivência da população inteira e dos indivíduos não curados seja representada por $S_{pop}(t \mid \boldsymbol{\lambda}, \pi)$ e $S(t \mid \boldsymbol{\lambda})$ respectivamente. Então,

$$\begin{aligned} S_{pop}(t \mid \boldsymbol{\lambda}, \pi) &= 1 - \pi + \pi \times S(t \mid \boldsymbol{\lambda}) \quad \text{e} \\ f_{pop}(t \mid \boldsymbol{\lambda}, \pi) &= \pi \times f(t \mid \boldsymbol{\lambda}), \end{aligned} \tag{2.6}$$

em que $f(t \mid \boldsymbol{\lambda}) = -\frac{\partial S(t \mid \boldsymbol{\lambda})}{\partial t}$ e π é a proporção de indivíduos suscetíveis (não curados) da população. No entanto, para $S(t \mid \boldsymbol{\lambda})$ podem ser adotadas abordagens paramétricas como em Maller e Zhou [1996]; semi-paramétricas como a proposta por Peng [2003] e não paramétricas como pode ser visto em Lu [2010]. Neste trabalho será considerada apenas uma abordagem paramétrica.

Defina uma variável indicadora (não observável) M tal que

$$M = \begin{cases} 1 & , \text{ se o indivíduo é suscetível ao evento de interesse,} \\ 0 & , \text{ se o indivíduo é imune ao evento de interesse.} \end{cases}$$

Por outro lado, suponha T^* e C os tempos de falha e censura respectivamente e defina $\delta = I(T^* \leq C)$ e $T = \min(T^*, C)$, o indicador de falha e o tempo observado respectivamente. Assim, dado $M = m$, tem-se que

$$\begin{aligned} S_T(t \mid M = m, \boldsymbol{\lambda}) &= S(t \mid \boldsymbol{\lambda})^m \quad \text{e} \\ f_T(t \mid M = m, \boldsymbol{\lambda}) &= mf(t \mid \boldsymbol{\lambda})S(t \mid \boldsymbol{\lambda})^{m-1}, \end{aligned}$$

e portanto

$$\begin{aligned} f(t, \delta \mid M = m, \boldsymbol{\lambda}) &= \{S_T(t \mid M = m, \boldsymbol{\lambda})\}^{1-\delta} \{f_T(t \mid M = m, \boldsymbol{\lambda})\}^\delta, \\ &= \{S(t \mid \boldsymbol{\lambda})^m\}^{1-\delta} \{mf(t \mid \boldsymbol{\lambda})S(t \mid \boldsymbol{\lambda})^{m-1}\}^\delta, \\ &= S(t \mid \boldsymbol{\lambda})^{m-\delta} [mf(t \mid \boldsymbol{\lambda})]^\delta. \end{aligned}$$

Além disso, note que $\delta = 1$ implica $M = 1$ e dado que

$$\begin{aligned} f(t, \delta = 0 \mid \boldsymbol{\lambda}) &= \sum_{m=0}^1 f(t, \delta = 0 \mid M = m, \boldsymbol{\lambda}) P(M = m \mid \pi), \\ &= (1 - \pi) + \pi S(t \mid \boldsymbol{\lambda}), \end{aligned}$$

então tem-se que

$$\begin{aligned} P(M = 1 \mid t, \delta = 0, \boldsymbol{\lambda}) &= \frac{f(t, \delta = 0, m = 1 \mid \boldsymbol{\lambda})}{f(t, \delta = 0 \mid \boldsymbol{\lambda})}, \\ &= \frac{\pi S(t \mid \boldsymbol{\lambda})}{(1 - \pi) + \pi S(t \mid \boldsymbol{\lambda})}. \end{aligned}$$

Assim, conclui-se que a esperança condicional de M dada a informação observada é

$$E(M \mid \mathbf{D}_{obs}) = \delta + (1 - \delta) \frac{\pi S(t \mid \boldsymbol{\lambda})}{(1 - \pi) + \pi S(t \mid \boldsymbol{\lambda})}. \quad (2.7)$$

Adicionalmente, para populações heterogêneas é possível introduzir o vetor $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ir})$ de covariáveis observadas no i -ésimo indivíduo e associadas à parte suscetível do modelo mediante

$$S(t \mid \boldsymbol{\lambda}, \eta_i) = [S_0(t \mid \boldsymbol{\lambda})]^{\exp \eta_i}, \quad i = 1, \dots, n, \quad (2.8)$$

em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\gamma}$, $\boldsymbol{\gamma}^\top = (\gamma_1, \dots, \gamma_r)$ é vetor de parâmetros e $S_0(t \mid \boldsymbol{\lambda})$ é a função de sobrevivência basal, os dois ligados à sobrevivência para os indivíduos suscetíveis.

$S_0(t \mid \boldsymbol{\lambda})$ deve pertencer à classe de modelos de riscos proporcionais. (Entre outras, as distribuições Weibull e Exponencial Segmentadas apresentadas na seção 2.1.1 e 2.1.2 respectivamente). A propriedade de riscos proporcionais garante que $S_0(t \mid \boldsymbol{\lambda})$ e $S(t \mid \boldsymbol{\lambda}, \eta_i)$ pertençam à mesma classe de modelos. Por exemplo, se $S_0(t \mid \boldsymbol{\lambda})$ corresponde a distribuição Weibull como em (2.1), então

$$S(t \mid \eta_i, \boldsymbol{\lambda}) = [\exp\{-\lambda t^\nu\}]^{\exp \eta_i} = \exp\{-\lambda e^{\eta_i} t^\nu\}, \quad (2.9)$$

que corresponde à função de sobrevivência da distribuição Weibull com parâmetro de escala λe^{η_i} e parâmetro de forma ν .

Por outro lado, considere que o vetor de covariáveis $\mathbf{z}_i$ para o indivíduo i está associado à

probabilidade de ser suscetível. Então, a proporção π de não curados pode ser ajustada através de um modelo para dados binários em $\mathbf{z}$, de modo que

$$\pi_i = P(M_i = 1 \mid \vartheta_i) = \frac{\exp(\vartheta_i)}{1 + \exp(\vartheta_i)}, \quad i = 1, \dots, n, \quad (2.10)$$

em que $\vartheta_i = \mathbf{z}_i^\top \boldsymbol{\beta}$, $\mathbf{z}_i^\top = (z_{i1}, \dots, z_{is})$ e $\boldsymbol{\beta}$ é um vetor de parâmetros desconhecidos de dimensão s .

Dessa forma, sob a suposição de censura não informativa como explicado em Williams e Lagakos [1977] e Lagakos [1979], tem-se que a densidade conjunta de $(\mathbf{T}, \boldsymbol{\delta}, \mathbf{M})$ pode ser escrita como

$$\begin{aligned} f(\mathbf{t}, \boldsymbol{\delta}, \mathbf{m} \mid \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\vartheta}) &= \prod_{i=1}^n f(t_i, \delta_i, m_i \mid \boldsymbol{\lambda}, \eta_i, \vartheta_i), \\ &= \prod_{i=1}^n f(t_i, \delta_i \mid M_i = m_i, \boldsymbol{\lambda}, \eta_i) P(M_i = m_i \mid \vartheta_i), \\ &= \prod_{i=1}^n S(t_i \mid \boldsymbol{\gamma}, \boldsymbol{\lambda})^{m_i - \delta_i} [m_i f(t_i \mid \boldsymbol{\gamma}, \boldsymbol{\lambda})]^{\delta_i} \pi_i^{m_i} (1 - \pi_i)^{1 - m_i}. \end{aligned}$$

Portanto, a função de verossimilhança aumentada do modelo é dada por

$$L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}_c) = \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda}, \eta_i)^{M_i - \delta_i} [M_i f(t_i \mid \boldsymbol{\lambda}, \eta_i)]^{\delta_i} \pi_i^{M_i} (1 - \pi_i)^{1 - M_i}. \quad (2.11)$$

Também é possível trabalhar com a verossimilhança observada, a qual é obtida somando (2.11) para $\mathbf{M}$ no conjunto $\{0, 1\}^n$. Como resultado disso, obtém-se

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^n f_{pop}(t_i \mid \boldsymbol{\lambda}, \eta_i, \vartheta_i)^{\delta_i} [S_{pop}(t_i \mid \boldsymbol{\lambda}, \eta_i, \vartheta_i)]^{1 - \delta_i}, \\ &= \prod_{i=1}^n \left\{ \pi_i f(t_i \mid \boldsymbol{\lambda}, \eta_i) \right\}^{\delta_i} \left\{ 1 - \pi_i + \pi_i S(t_i \mid \boldsymbol{\lambda}, \eta_i) \right\}^{1 - \delta_i}. \end{aligned} \quad (2.12)$$

2.2.2 Modelo de tempos de promoção

Embora o modelo de mistura seja muito atraente e intuitivo, apresenta alguns inconvenientes comentados por Yakolev e Tsodikov [1996] e Chen et al. [1999a]. Primeiro, na presença de covariáveis, não possui uma estrutura de riscos proporcionais, que é uma propriedade desejável nos modelos de sobrevivência, especialmente quando é considerada uma perspectiva frequentista, pois muitos resultados assintóticos e computacionais requerem uma estrutura de riscos proporcionais. Segundo, quando são incluídas covariáveis para a proporção de suscetíveis através de um modelo de regressão binomial padrão, (2.6) leva a distribuições à posteriori impróprias para muitos tipos de distribuições não informativas *a priori*, que são impróprias, incluindo a priori uniforme para os coeficientes de regressão. Esse é um inconveniente crucial de (2.6) porque implica que a Inferência bayesiana nesse modelo requer uma distribuição *a priori* própria.

O modelo de Chen et al. [1999a] surge em um contexto em que ocorre recaída em pacientes com câncer. Seja M o número de células carcinogênicas ativas que o indivíduo apresentou quando começou o tratamento. Assuma que M tem distribuição de Poisson com média θ . Seja Y_c o tempo até que a c -ésima célula carcinogênica produza um câncer detectável. Suponha que as variáveis $\{Y_c, c = 1, 2, \dots\}$ são i.i.d. com f.d.a. $F(t \mid \boldsymbol{\lambda}) = 1 - S(t \mid \boldsymbol{\lambda})$, independentes de M . Assim, o tempo

para o aparecimento do câncer é definido pela variável aleatória $T = \min\{Y_c, 0 \leq c \leq M\}$, em que $P(Y_0 = \infty) = 1$. Nessas condições, tem-se que

$$\begin{aligned}
S_{pop}(t \mid \boldsymbol{\lambda}, \theta) &= P(T > t \mid \boldsymbol{\lambda}, \theta), \\
&= \sum_{m=0}^{\infty} P(T > t \mid M = m, \boldsymbol{\lambda}) P(M = m \mid \theta), \\
&= P(T > t \mid M = 0, \boldsymbol{\lambda}) P(M = 0 \mid \theta) + \sum_{m=1}^{\infty} P(T > t \mid M = m, \boldsymbol{\lambda}) P(M = m \mid \theta), \\
&= P(M = 0 \mid \theta) + \sum_{m=1}^{\infty} P(Y_1 > t, \dots, Y_m > t \mid M = m, \boldsymbol{\lambda}) P(M = m \mid \theta), \\
&= P(M = 0 \mid \theta) + \sum_{m=1}^{\infty} \{P(Y_1 > t \mid \boldsymbol{\lambda})\}^m P(M = m \mid \theta), \\
&= \sum_{m=0}^{\infty} \{S(t \mid \boldsymbol{\lambda})\}^m P(M = m \mid \theta).
\end{aligned} \tag{2.13}$$

Assim, usando o fato de que $M \sim Po(\theta)$, tem-se que

$$\begin{aligned}
S_{pop}(t \mid \boldsymbol{\lambda}, \theta) &= \sum_{m=0}^{\infty} \{S(t \mid \boldsymbol{\lambda})\}^m \frac{e^{-\theta} \theta^m}{m!}, \\
&= \sum_{m=0}^{\infty} \frac{e^{-\theta} \{\theta S(t \mid \boldsymbol{\lambda})\}^m}{m!}, \\
&= e^{-\theta} \sum_{m=0}^{\infty} \frac{\{\theta S(t \mid \boldsymbol{\lambda})\}^m}{m!}, \\
&= e^{-\theta} e^{\theta S(t \mid \boldsymbol{\lambda})}, \\
&= e^{-\theta(1-S(t \mid \boldsymbol{\lambda}))}, \\
&= e^{-\theta F(t \mid \boldsymbol{\lambda})}.
\end{aligned} \tag{2.14}$$

Derivando em relação a t , tem-se que as funções de densidade e de risco da população são dadas respectivamente por

$$\begin{aligned}
f_{pop}(t \mid \boldsymbol{\lambda}, \theta) &= \theta f(t \mid \boldsymbol{\lambda}) e^{-\theta F(t \mid \boldsymbol{\lambda})} \quad e \\
h_{pop}(t \mid \boldsymbol{\lambda}, \theta) &= \frac{f_{pop}(t \mid \boldsymbol{\lambda})}{S_{pop}(t \mid \boldsymbol{\lambda})} = \theta f(t \mid \boldsymbol{\lambda}).
\end{aligned}$$

Além disso, Mizoi [2004] mostra que

$$P(M = m \mid t, \delta, \boldsymbol{\lambda}, \theta) = \frac{e^{-\theta S(t \mid \boldsymbol{\lambda}, \theta)} [\theta S(t \mid \boldsymbol{\lambda}, \theta)]^{(m-\delta)}}{(m-\delta)!}.$$

Portanto, $M - \delta \mid t, \delta \sim P(\theta S(t \mid \boldsymbol{\lambda}))$. Assim, conclui-se que a esperança condicional de M dada a informação observada é

$$E(M \mid \boldsymbol{D}_{obs}) = \delta + \theta S(t \mid \boldsymbol{\lambda}).$$

Note que os indivíduos curados ficam caracterizados por aqueles em que $M \geq 1$. Portanto, a função de sobrevivência dos indivíduos suscetíveis (que denotaremos por $S_{sus}(t \mid \boldsymbol{\lambda}, \theta)$) pode ser calculada como segue

$$\begin{aligned}
S_{sus}(t \mid \boldsymbol{\lambda}, \theta) &= P(T > t \mid M \geq 1), \\
&= \frac{P(T > t \mid M \geq 0, \boldsymbol{\lambda}, \theta) - P(T > t \mid M = 0, \boldsymbol{\lambda}, \theta)}{P(M \geq 1 \mid \theta)}, \\
&= \frac{S_{pop}(t \mid \boldsymbol{\lambda}, \theta) - P(M = 0 \mid \theta)}{P(M \geq 1 \mid \theta)}, \\
&= \frac{e^{-\theta F(t \mid \boldsymbol{\lambda})} - e^{-\theta}}{1 - e^{-\theta}}.
\end{aligned} \tag{2.15}$$

Na proposta inicial de [Chen et al. \[1999a\]](#), é assumido que $F(t \mid \boldsymbol{\lambda})$ provém da distribuição Weibull. Portanto, nesse modelo em particular, (2.15) corresponde ao modelo Weibull Poisson em [Morais e Barreto-Souza \[2011\]](#).

Uma característica muito importante do modelo de Chen é que ele pertence à família dos modelos de riscos proporcionais de Cox. Isso é fácil de visualizar. Suponha que dois indivíduos, digamos i e j , possuam o parâmetro associado à fração de cura dado por θ_i e θ_j respectivamente. Desta forma, tem-se que

$$\frac{h_{pop}(t \mid \boldsymbol{\lambda}, \theta_i)}{h_{pop}(t \mid \boldsymbol{\lambda}, \theta_j)} = \frac{\theta_i f(t \mid \boldsymbol{\lambda})}{\theta_j f(t \mid \boldsymbol{\lambda})} = \frac{\theta_i}{\theta_j}.$$

Como a última expressão não depende do tempo, então fica provado que o modelo de Chen pertence à classe de modelos de riscos proporcionais. Por outro lado, a fração de cura do modelo é dada por

$$S_{pop}(\infty \mid \boldsymbol{\lambda}, \theta) = \lim_{t \rightarrow \infty} S_{pop}(t \mid \boldsymbol{\lambda}, \theta) = \lim_{t \rightarrow \infty} e^{-\theta F(t \mid \boldsymbol{\lambda})} = e^{-\theta}. \tag{2.16}$$

Similarmente ao modelo de Berkson e Gage, a probabilidade do i -ésimo indivíduo ser suscetível pode ser modelada através de um vetor de covariáveis $\mathbf{z}_i$, por exemplo, através de

$$\theta_i = \exp(\vartheta_i), \quad i = 1, \dots, n,$$

em que $\vartheta_i = \mathbf{z}_i^\top \boldsymbol{\beta}$. Também um conjunto $\mathbf{x}_i$ de covariáveis observadas no i -ésimo indivíduo podem ser incorporadas através da função $F(t \mid \boldsymbol{\lambda})$ (ou da função $S(t \mid \boldsymbol{\lambda})$) para modelar a sobrevivência dos indivíduos suscetíveis como em (2.8), mas devemos tomar cuidado no momento de interpretar estes coeficientes, pois neste caso estas covariáveis são incluídas através da função de sobrevivência das células carcinogênicas e não diretamente através da função de sobrevivência dos indivíduos suscetíveis.

Assumindo censura à direita, é possível trabalhar com a verossimilhança aumentada do modelo. Usando argumentos similares apresentados na seção anterior, verifica-se que sob a suposição de censura não informativa, tem-se que

$$f(t_i, \delta_i \mid M_i = m_i, \boldsymbol{\lambda}, \eta_i, \vartheta_i) = S(t_i \mid \boldsymbol{\lambda}, \eta_i)^{m_i - \delta_i} [m_i f(t_i \mid \boldsymbol{\lambda}, \eta_i)]^{\delta_i}.$$

Portanto, segue que

$$\begin{aligned}
f(\mathbf{t}, \boldsymbol{\delta}, \mathbf{m} \mid \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\vartheta}) &= \prod_{i=1}^n f(t_i, \delta_i, m_i \mid \boldsymbol{\lambda}, \eta_i, \vartheta_i), \\
&= \prod_{i=1}^n f(t_i, \delta_i \mid M_i = m_i, \boldsymbol{\lambda}, \eta_i) P(M_i = m_i \mid \vartheta_i), \\
&= \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda}, \eta_i)^{m_i - \delta_i} [m_i f(t_i \mid \boldsymbol{\lambda}, \eta_i)]^{\delta_i} \frac{e^{-\theta_i} \theta_i^{m_i}}{m_i!}.
\end{aligned}$$

Finalmente, somando sob $\mathbf{M}$ no conjunto $\{0, 1, 2, \dots\}^n$, obtém-se a verossimilhança observada do modelo, que é dada por

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^n \{f_{pop}(t_i | \boldsymbol{\lambda}, \eta_i)\}^{\delta_i} \{S_{pop}(t_i | \boldsymbol{\lambda}, \eta_i)\}^{1-\delta_i}, \\ &= \prod_{i=1}^n \left\{ \theta_i e^{-\theta_i F(t_i | \boldsymbol{\lambda}, \eta_i)} \right\}^{\delta_i} \left\{ \theta_i f(t_i | \boldsymbol{\lambda}, \eta_i) e^{-\theta_i F(t_i | \boldsymbol{\lambda}, \eta_i)} \right\}^{1-\delta_i}, \\ &= \prod_{i=1}^n \left\{ e^{-\theta_i F(t_i | \boldsymbol{\lambda}, \eta_i)} [\theta_i f(t_i | \boldsymbol{\lambda}, \eta_i)]^{\delta_i} \right\}. \end{aligned}$$

2.2.3 Extensões do modelo de Chen

Na literatura, há alguns trabalhos que unificam os dois modelos apresentados anteriormente em uma única classe de modelos. Por exemplo, [Yin and Ibrahim \[2005\]](#) propõem uma classe mais ampla de modelos com fração de cura, introduzindo um novo parâmetro $\phi \in [0, 1]$ na função de sobrevivência populacional fazendo

$$S_{pop}(t | \boldsymbol{\lambda}, \theta, \phi) = \{1 - \phi \theta F(t | \boldsymbol{\lambda})\}^{\phi^{-1}}. \quad (2.17)$$

Note que para o caso limite $\phi \rightarrow 0$, recupera-se o modelo de [Chen et al. \[1999a\]](#) e para $\phi = 1$, tem-se como caso particular o modelo de [Berkson e Gage \[1952\]](#).

Porém, a extensão proposta por [Rodrigues et al. \[2009a\]](#) parece mais natural, além de estender o espaço paramétrico para ϕ . Similarmente ao modelo de Chen, seja M o número de células carcinogênicas ativas que um indivíduo apresentou quando começou o tratamento. Agora, assuma que M tem distribuição Binomial Negativa com parâmetros θ e ϕ tal que $\theta > 0$ e $\theta\phi > -1$, i.e., $E(M) = \theta$ e $\text{Var}(M) = \theta(1 + \phi)$. Seja Y_c o tempo até que a c -ésima célula carcinogênica produza um câncer detectável. Suponha que as variáveis $\{Y_c, c = 1, 2, \dots\}$ são i.i.d. com f.d.a. $F(t | \boldsymbol{\lambda}) = 1 - S(t | \boldsymbol{\lambda})$, independentes de M . Dessa forma, o tempo para o aparecimento do câncer é definido pela variável aleatória $T = \min\{Y_c, 0 \leq c \leq M\}$, em que $P(Y_0 = \infty) = 1$. Assim, e usando (2.13), tem-se que

$$\begin{aligned} S_{pop}(t | \boldsymbol{\lambda}, \theta, \phi) &= \sum_{m=0}^{\infty} \{S(t | \boldsymbol{\lambda})\}^m P(M = m | \theta, \phi), \\ &= \sum_{m=0}^{\infty} \{S(t | \boldsymbol{\lambda})\}^m \frac{\Gamma(\phi^{-1} + m)}{\Gamma(\phi^{-1})m!} \left(\frac{\phi\theta}{1 + \phi\theta} \right)^m (1 + \phi\theta)^{-\phi^{-1}}, \\ &= (1 + \phi\theta)^{-\phi^{-1}} \sum_{m=0}^{\infty} \frac{\Gamma(\phi^{-1} + m)}{\Gamma(\phi^{-1})m!} \left(\frac{\phi\theta S(t | \boldsymbol{\lambda})}{1 + \phi\theta} \right)^m, \\ &= (1 + \phi\theta)^{-\phi^{-1}} \left\{ 1 - \frac{\phi\theta S(t | \boldsymbol{\lambda})}{1 + \phi\theta} \right\}^{-\phi^{-1}} \times \\ &\quad \times \sum_{m=0}^{\infty} \frac{\Gamma(\phi^{-1} + m)}{\Gamma(\phi^{-1})m!} \left(\frac{\phi\theta S(t | \boldsymbol{\lambda})}{1 + \phi\theta} \right)^m \left\{ 1 - \frac{\phi\theta S(t | \boldsymbol{\lambda})}{1 + \phi\theta} \right\}^{\phi^{-1}}, \\ &= \{1 + \phi\theta(1 - S(t | \boldsymbol{\lambda}))\}^{-\phi^{-1}}, \\ &= \{1 + \phi\theta F(t | \boldsymbol{\lambda})\}^{-\phi^{-1}}. \end{aligned} \quad (2.18)$$

Dessa forma, as funções de densidade e de risco do modelo são dadas por

$$f_{pop}(t | \boldsymbol{\lambda}, \theta, \phi) = \theta f(t | \boldsymbol{\lambda}) \{1 + \phi \theta F(t | \boldsymbol{\lambda})\}^{-\phi^{-1}-1}, \quad e$$

$$h_{pop}(t | \boldsymbol{\lambda}, \theta, \phi) = \frac{f_{pop}(t | \boldsymbol{\lambda})}{S_{pop}(t | \boldsymbol{\lambda})} = \theta f(t | \boldsymbol{\lambda}) (1 + \phi \theta F(t | \boldsymbol{\lambda}))^{-1}.$$

A fração de cura do modelo é dada por

$$S_{pop}(\infty | \boldsymbol{\lambda}, \theta, \phi) = \lim_{t \rightarrow \infty} S_{pop}(t | \boldsymbol{\lambda}, \theta, \phi) = \lim_{t \rightarrow \infty} \{1 + \phi \theta F(t | \boldsymbol{\lambda})\}^{-\phi^{-1}} = \{1 + \phi \theta\}^{-\phi^{-1}}.$$

Usando um argumento similar a (2.15), verifica-se que a função de sobrevivência dos indivíduos suscetíveis é dada por

$$S_{sus}(t | \boldsymbol{\lambda}, \theta, \phi) = \frac{\{1 + \phi \theta F(t | \boldsymbol{\lambda})\}^{-\phi^{-1}} - \{1 + \phi \theta\}^{-\phi^{-1}}}{1 - \{1 + \phi \theta\}^{-\phi^{-1}}}. \quad (2.19)$$

Uma forma alternativa de ver o modelo em (2.18) pode ser vista em [Azevedo, R. \[2011\]](#), em que assumindo o modelo de [Chen et al. \[1999a\]](#) e que existe uma variável de fragilidade comum com distribuição Gama de média 1 e variância ϕ para os respectivos fatores latentes de cada indivíduo.

O caso particular em que $\phi = 1$ merece a pena ser mencionado pois é muito utilizado na literatura (veja por exemplo, [Cancho et. al \[2012b\]](#)). Corresponde a assumir que a variável M tem distribuição Geométrica de média θ . Neste caso, a equação (2.18) fica na forma

$$S_{pop}(t | \boldsymbol{\lambda}, \theta) = \{1 + \theta F(t | \boldsymbol{\lambda})\}^{-1}.$$

Neste caso, a função de sobrevivência dos indivíduos suscetíveis em (2.19) assume a forma

$$S_{sus}(t | \boldsymbol{\lambda}, \theta, \phi) = \frac{\{1 + \theta F(t | \boldsymbol{\lambda})\}^{-1} - \{1 + \theta\}^{-1}}{1 - \{1 + \theta\}^{-1}}.$$

Por outro lado, note que a parametrização usada por [Yin and Ibrahim \[2005\]](#) para ϕ em (2.17) não é a mesma que a utilizada por [Rodrigues et al. \[2009a\]](#) em (2.18), mas difere só no sinal.

Em relação ao parâmetro ϕ , [Yin and Ibrahim \[2005\]](#) discutem a estimação deste parâmetro desde um ponto de vista Bayesiano, afirmando que uma priori continua para ϕ pode fazer o processo de estimação numericamente instável quando ϕ é próximo de zero. A solução proposta por eles foi utilizar uma priori discreta, digamos sob o conjunto $\{\phi_1, \dots, \phi_k\}$, em que $\phi_1 = 0$ e $\phi_k = 1$ e atribuir a mesma probabilidade para todos os valores nesse conjunto. Esta mesma ideia foi utilizada por [de Castro et al. \[2009\]](#) no modelo com o espaço estendido para ϕ , i.e., $\phi \in [-1, \infty)$. No entanto, desde um ponto de vista clássico, [Rodrigues et al. \[2009a\]](#) ajustaram o modelo para diferentes valores de ϕ e escolheram aquele que produzia uma maior verossimilhança.

Como discutido em [Rodrigues et al. \[2008\]](#), um conjunto de covariáveis $\mathbf{Z}$ podem ser incluídas no modelo para explicar a proporção de cura existente na população através de

$$\theta_i = \begin{cases} \exp\{\mathbf{z}_i^\top \boldsymbol{\beta}\} & , i = 1, \dots, n, \text{ se } \phi > 0, \\ \frac{\exp\{\mathbf{z}_i^\top \boldsymbol{\beta}\}}{1 - \phi \exp\{\mathbf{z}_i^\top \boldsymbol{\beta}\}} & , i = 1, \dots, n, \text{ se } \phi \in [-1, 0]. \end{cases}$$

Adicionalmente, um conjunto de covariáveis $\mathbf{X}$ podem ser incluídas através da função $F(t | \boldsymbol{\lambda})$ (ou através de $S(t | \boldsymbol{\lambda})$) para modelar a função de sobrevivência das células dos indivíduos não curados.

Supondo censura à direita, a verossimilhança do modelo é dada por

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^n \{f_{pop}(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})\}^{\delta_i} \{S_{pop}(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})\}^{1-\delta_i}, \\ &= \prod_{i=1}^n \left\{ \theta_i f(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}) \{1 + \phi \theta_i F(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})\}^{-\phi^{-1}-1} \right\}^{\delta_i} \left\{ \{1 + \phi \theta_i F(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})\}^{-\phi^{-1}} \right\}^{1-\delta_i}, \\ &= \prod_{i=1}^n \left\{ \{1 + \phi \theta_i F(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})\}^{-(\delta_i + \phi^{-1})} [\theta_i f(t \mid \boldsymbol{\lambda}, \boldsymbol{\gamma})]^{\delta_i} \right\}. \end{aligned}$$

2.2.4 Modelos destrutivos de Poisson ponderado com fração de cura

Rodrigues et al. [2011] apresentam os *modelos destrutivos de Poisson ponderado com fração de cura (MDPPFC)*, uma nova extensão do modelo de Chen et al. [1999a]. Nesta classe, o modelo de Rodrigues et al. [2009a] é um caso particular (e portanto, o modelo de mistura e de Chen também são).

A abordagem de Rodrigues et al. [2011] parece adaptar-se melhor à vida real em algumas circunstâncias. Lembremos que nos modelos de Chen et al. [1999a] e de Rodrigues et al. [2009a], o número de células ou causas concorrentes com o evento de interesse era representado com a variável M . Se um indivíduo possuía inicialmente um número $M > 0$ de células ou causas, então qualquer uma delas podia causar no indivíduo o evento de interesse e se o indivíduo possuía $M = 0$ células ou causas, então diz-se que ele era imune ou curado. Nesta nova abordagem, a suposição é que o indivíduo começa com M células ou causas concorrentes, mas por diversos motivos, só uma porcentagem p delas pode motivar a falha no indivíduo. O número de células que podem causar a falha é denotada por D .

Na prática, essa abordagem parece ser mais realista. Por exemplo, os oncogenes são genes que de alguma forma controlam o ciclo celular e que estão presentes em todos os seres humanos. Alterações desse genes (sejam hereditárias ou ambientais) podem fazer que em algum momento as células de um tecido específico “percam o controle” e comece a se replicar de maneira muito mais rápida que o usual, o que dá origem ao que conhecemos como câncer. Portanto, neste problema, se um indivíduo for curado é mais realista supor que ele começou com M oncogenes e que nenhum deles foi ativado, que corresponde a supor que ele inicialmente não possuía oncogenes, como é suposto nos modelos de Chen et al. [1999a] e Rodrigues et al. [2009a].

Considere agora que M é uma variável aleatória não observável com distribuição de Poisson ponderada, e que denota o número de causas concorrentes relacionadas à ocorrência do evento de interesse, com função de probabilidade

$$p(m \mid \theta, \phi) = P(M = m \mid \theta, \phi) = \frac{w(m \mid \phi) p^*(m \mid \theta)}{E_\theta[w(M \mid \phi)]}, \quad m = 0, 1, 2, \dots \quad (2.20)$$

em que $w(\cdot \mid \phi)$ é uma função de pesos não negativa com parâmetro ϕ , $p^*(\cdot \mid \theta)$ é a função de probabilidade da Poisson com média θ e $E_\theta[\cdot]$ indica que a esperança deve ser calculada considerando que a variável M tem distribuição Poisson com média θ . A expressão em (2.20) é chamada distribuição de Poisson Ponderada com parâmetro θ e função de peso $w(\cdot \mid \phi)$. Note que se $w(m \mid \phi) = c$, $c \in \mathbb{R}$ então recupera-se a distribuição de Poisson usual.

Sejam A_k , $k = 1, 2, \dots, M$, variáveis aleatórias que dado $M = m$, são independentes e identicamente distribuídas seguindo uma distribuição Bernoulli com probabilidade de sucesso p . Essencial-

mente, $A_k = 1$ indica a presença da k -ésima causa. O dano total D , que representa o número total de causas concorrentes entre as M causas iniciais que não são destruídas, é definido como

$$D = \begin{cases} A_1 + A_2 + \dots + A_M & , \text{ se } M > 0, \\ 0 & , \text{ se } M = 0. \end{cases}$$

Devido ao dano ou destruição, temos que $D \leq M$. Note que $D \mid M = m, p \sim \text{Bin}(m, p)$ se $m > 0$ e $P(D = 0 \mid M = 0, p) = 1$. Sejam Y_c , $c = 1, \dots, D$ as variáveis aleatórias associadas aos tempos de vida de cada uma das causas não destruídas, todas elas i.i.d. com f.d.a. $F(t \mid \boldsymbol{\lambda}) = 1 - S(t \mid \boldsymbol{\lambda})$. Então, analogamente aos modelos anteriores, o tempo de vida do indivíduo será definido como

$$T = \begin{cases} \min\{Y_1, Y_2, \dots, Y_D\} & , \text{ se } D > 0, \\ \infty & , \text{ se } D = 0. \end{cases}$$

Nessas condições, e usando (2.13) temos que

$$S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) = \sum_{d=0}^{\infty} \{S(t \mid \boldsymbol{\lambda})\}^d P(D = d \mid \theta, p, \phi), \quad (2.21)$$

onde

$$P(D = d \mid \theta, p, \phi) = \frac{(\theta p)^d e^{-\theta p} E_{\theta(1-p)}[w(M + d \mid \phi)]}{d! E_{\theta}[w(M \mid \phi)]}, \quad d = 0, 1, 2, \dots \quad (2.22)$$

No Apêndice A.1.1 aparecem detalhes sobre o cálculo de $P(D = d \mid \theta, p, \phi)$. Dessa forma, substituindo (2.22) em (2.21) temos que

$$\begin{aligned} S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \sum_{d=0}^{\infty} \{S(t \mid \boldsymbol{\lambda})\}^d P(D = d \mid \theta, p, \phi), \\ &= \sum_{d=0}^{\infty} \{S(t \mid \boldsymbol{\lambda})\}^d \frac{(\theta p)^d e^{-\theta p} E_{\theta(1-p)}[w(M + d \mid \phi)]}{d! E_{\theta}[w(M \mid \phi)]}, \\ &= \frac{e^{-\theta p + \theta p S(t \mid \boldsymbol{\lambda})}}{E_{\theta}[w(M \mid \phi)]} \sum_{d=0}^{\infty} \frac{E_{\theta(1-p)}[w(M + d \mid \phi)] [\theta p S(t \mid \boldsymbol{\lambda})]^d e^{-\theta p S(t \mid \boldsymbol{\lambda})}}{d!}, \\ &= \frac{e^{-\theta p F(t \mid \boldsymbol{\lambda})}}{E_{\theta}[w(M \mid \phi)]} \sum_{d=0}^{\infty} \frac{w^*(d \mid \phi) [\theta p S(t \mid \boldsymbol{\lambda})]^d e^{-\theta p S(t \mid \boldsymbol{\lambda})}}{d!}, \end{aligned}$$

em que $w^*(d \mid \phi) = E_{\theta(1-p)}[w(M + d \mid \phi)]$. Dessa forma,

$$S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) = \frac{e^{-\theta p F(t \mid \boldsymbol{\lambda})} E_{\theta p S(t \mid \boldsymbol{\lambda})}[w^*(M \mid \phi)]}{E_{\theta}[w(M \mid \phi)]}.$$

Defina $\theta_1 = \theta p S(t \mid \boldsymbol{\lambda})$ e $\theta_2 = \theta(1 - p)$. Então,

$$\begin{aligned} E_{\theta_1}[w^*(M \mid \phi)] &= \sum_{i=0}^{\infty} \frac{\theta_1^i e^{-\theta_1}}{i!} E_{\theta_2}[w(M + i \mid \phi)], \\ &= \sum_{i=0}^{\infty} \frac{\theta_1^i e^{-\theta_1}}{i!} \sum_{j=0}^{\infty} \frac{\theta_2^j e^{-\theta_2}}{j!} w(i + j \mid \phi), \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{\theta_1^i \theta_2^j e^{-(\theta_1 + \theta_2)}}{i! j!} w(i + j \mid \phi), \\ &= E_{\theta_1 + \theta_2}[w(\mathfrak{A} + \mathfrak{B} \mid \phi)], \end{aligned}$$

em que $\mathfrak{A} \sim Po(\theta_1)$ é independente de $\mathfrak{B} \sim Po(\theta_2)$. Mas, como $\mathfrak{A} + \mathfrak{B} \sim Po(\theta_1 + \theta_2)$ e $\theta_1 + \theta_2 = \theta p S(t | \boldsymbol{\lambda}) + \theta(1 - p) = \theta - \theta p(1 - S(t | \boldsymbol{\lambda})) = \theta(1 - pF(t | \boldsymbol{\lambda}))$, então conclui-se que

$$E_{\theta_1}[w^*(M | \phi)] = E_{\theta(1-pF(t|\boldsymbol{\lambda}))}[w(M | \phi)].$$

Dessa forma, verifica-se que

$$S_{pop}(t | \boldsymbol{\lambda}, \theta, p, \phi) = \frac{e^{-\theta p F(t|\boldsymbol{\lambda})} E_{\theta(1-pF(t|\boldsymbol{\lambda}))}[w(M | \phi)]}{E_{\theta}[w(M | \phi)]}. \quad (2.23)$$

Neste caso, não é possível obter uma forma fechada para as funções $f_{pop}(t | \boldsymbol{\lambda}, \theta, p, \phi)$ e $h_{pop}(t | \boldsymbol{\lambda}, \theta, p, \phi)$. Porém, elas serão proporcionadas para os casos particulares de funções $w(\cdot | \phi)$ consideradas neste trabalho.

Por outro lado, note que a fração de cura nos MDPPFC é dada por

$$\lim_{t \rightarrow \infty} S_{pop}(t | \boldsymbol{\lambda}, \theta, p, \phi) = \lim_{t \rightarrow \infty} \frac{e^{-\theta p F(t|\boldsymbol{\lambda})} E_{\theta(1-pF(t|\boldsymbol{\lambda}))}[w(M | \phi)]}{E_{\theta}[w(M | \phi)]} = \frac{e^{-\theta p} E_{\theta(1-p)}[w(M | \phi)]}{E_{\theta}[w(M | \phi)]}.$$

A seguir, particularizamos o modelo para diferentes funções de peso $w(\cdot | \phi)$ apresentadas em [Rodrigues et al. \[2011\]](#).

1. Modelo destrutivo de Poisson com viés de duração (DLBP)

Considere $w(m | \phi) = m$, i.e., a função $w(\cdot)$ não depende de ϕ . Neste caso, $E_{\theta}[w(M | \phi)] = E_{\theta}[M] = \theta$ e portanto, segundo (2.20) tem-se que

$$\begin{aligned} P(M = m | \theta) &= \frac{m \theta^m e^{-\theta}}{m! \theta}, \quad m = 0, 1, 2, \dots \\ &= \frac{\theta^{m-1} e^{-\theta}}{(m-1)!}, \quad m = 1, 2, 3, \dots \end{aligned}$$

Em outras palavras, a distribuição de M é Poisson truncada em 1. Portanto, $E(M) = \theta + 1$ e $\text{Var}(M) = \theta$. Isso é muito relevante, pois das distribuições propostas em [Rodrigues et al. \[2011\]](#), esta é a única que possui a característica de ser truncada em 1, adaptando-se à abordagem discutida no começo desta seção no sentido que, em certas circunstâncias, é mais realista assumir que todos os indivíduos da população possuem células ou causas que podem produzir a falha e que essas células ou causas não foram ativadas, em vez de assumir simplesmente que alguns indivíduos não possuíam essas células ou causas. Similarmente, $E_{\theta(1-p)}[w(M + m | \phi)] = E_{\theta(1-p)}[M + m] = \theta(1 - p) + m$. Assim, segundo (2.22) tem-se que

$$\begin{aligned} P(D = d | \theta, p) &= \frac{(\theta p)^d e^{-\theta p} \{\theta(1 - p) + d\}}{d! \theta}, \\ &= \frac{(\theta p)^d e^{-\theta p}}{d!} \left(1 - p + \frac{d}{\theta}\right), \quad d = 0, 1, 2, \dots \end{aligned} \quad (2.24)$$

É simples verificar que a distribuição em (2.24) tem uma representação estocástica muito conveniente: $D | \theta, p = \text{Po}(\theta p) + \text{Bern}(p)$. Dessa forma, é possível calcular os momentos desta distribuição de uma forma rápida.

$$\begin{aligned} E(D | \theta, p) &= \theta p + p = (\theta + 1)p \\ \text{Var}(D | \theta, p) &= \theta p + p(1 - p) = (\theta + 1 - p)p. \end{aligned}$$

Note que $\text{Var}(D | \theta, p) \leq E(D | \theta, p)$, $\forall p \in (0, 1]$. Dessa forma, a distribuição de D

apresenta subdispersão. Por outro lado, usando (2.23), a função de sobrevivência do modelo destrutivo de Poisson com viés de duração é dada por

$$S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p) = \frac{e^{-\theta p F(t \mid \boldsymbol{\lambda})} \theta (1 - p F(t \mid \boldsymbol{\lambda}))}{\theta} = e^{-\theta p F(t \mid \boldsymbol{\lambda})} (1 - p F(t \mid \boldsymbol{\lambda})),$$

e as funções de densidade e de risco da população são dadas por

$$\begin{aligned} f_{pop}(t \mid \boldsymbol{\lambda}, \theta, p) &= p f(t \mid \boldsymbol{\lambda}) e^{-\theta p F(t \mid \boldsymbol{\lambda})} \{ \theta (1 - p F(t \mid \boldsymbol{\lambda})) + 1 \} \quad e \\ h_{pop}(t \mid \boldsymbol{\lambda}, \theta, p) &= p f(t \mid \boldsymbol{\lambda}) \left\{ \theta + \frac{1}{1 - p F(t \mid \boldsymbol{\lambda})} \right\}. \end{aligned}$$

Note que a fração de cura do modelo é dada por

$$S(\infty \mid \boldsymbol{\lambda}, \theta, p) = \lim_{t \rightarrow \infty} S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p) = e^{-\theta p} (1 - p).$$

Usando argumentos similares a (2.15), pode-se calcular a função de sobrevivência dos suscetíveis, que neste caso é dada por

$$S_{sus}(t \mid \boldsymbol{\lambda}, \theta, p) = \frac{e^{-\theta p F(t \mid \boldsymbol{\lambda})} (1 - p F(t \mid \boldsymbol{\lambda})) - e^{-\theta p} (1 - p)}{1 - e^{-\theta p} (1 - p)}. \quad (2.25)$$

Além disso, verifica-se que

$$\begin{aligned} M - \delta \mid t, \delta &\sim DLBP \left(\theta (1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta), \frac{1 - p F(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta} \right), \\ D - \delta \mid t, \delta &\sim DLBP \left(\theta (1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta), \frac{p S(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta} \right), \\ D - \delta \mid M = m, t, \delta &\sim Bin \left(m - \delta, \frac{p S(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda})} \right). \end{aligned} \quad (2.26)$$

Portanto, dada a informação observada, tem-se que as esperanças condicionais de M, D e $D \mid M = m$ são respectivamente

$$\begin{aligned} E(M \mid \boldsymbol{D}_{obs}) &= \delta + \frac{\{ \theta (1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta) + 1 \} (1 - p F(t \mid \boldsymbol{\lambda}))}{1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta}, \\ E(D \mid \boldsymbol{D}_{obs}) &= \delta + \frac{\{ \theta (1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta) + 1 \} p S(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda}) + \delta / \theta}, \\ E(D \mid M = m, \boldsymbol{D}_{obs}) &= (m - \delta) \frac{p (1 - p) S(t \mid \boldsymbol{\lambda})}{\{ 1 - p F(t \mid \boldsymbol{\lambda}) \}^2}. \end{aligned}$$

Detalhes desses resultados encontram-se no Apêndice A.1.2. Para trabalhar com a densidade aumentada do modelo, note que se as censuras observadas são à direita e sob a suposição de censura não informativa, verifica-se que

$$f(t, \delta \mid D = d, \boldsymbol{\lambda}) = S(t \mid \boldsymbol{\lambda})^{d-\delta} [df(t \mid \boldsymbol{\lambda})]^\delta. \quad (2.27)$$

Assim, a densidade conjunta de $(\mathbf{T}, \boldsymbol{\delta}, \mathbf{D})$ pode ser escrita como

$$\begin{aligned} f(\mathbf{t}, \boldsymbol{\delta}, \mathbf{d} \mid \boldsymbol{\lambda}, \theta, p) &= \prod_{i=1}^n f(t_i, \delta_i, d_i \mid \boldsymbol{\lambda}, \theta, p), \\ &= \prod_{i=1}^n f(t_i, \delta_i \mid D_i = d_i, \boldsymbol{\lambda}) P(D_i = d_i \mid \theta, p), \\ &= \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda})^{d_i - \delta_i} [d_i f(t_i \mid \boldsymbol{\lambda})]^{\delta_i} \frac{(\theta p)^{d_i} e^{-\theta p}}{d_i!} \left\{ 1 - p + \frac{d_i}{\theta} \right\}. \end{aligned}$$

Portanto, a verossimilhança aumentada do modelo é dada por

$$L(\boldsymbol{\lambda}, \theta, p \mid \mathbf{D}_c) = \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda})^{D_i - \delta_i} [D_i f(t_i \mid \boldsymbol{\lambda})]^{\delta_i} \frac{(\theta p)^{D_i} e^{-\theta p}}{D_i!} \left(1 - p + \frac{D_i}{\theta} \right). \quad (2.28)$$

No entanto, a verossimilhança observada pode ser construída somando (2.28) sob o vetor $\mathbf{D} = (D_1, \dots, D_n)$ no conjunto $\{0, 1, 2, \dots\}^n$. Usando (2.26), verifica-se que esta última verossimilhança é dada por

$$L(\boldsymbol{\lambda}, \theta, p \mid \mathbf{D}_c) = \prod_{i=1}^n \left\{ e^{-\theta p F(t_i \mid \boldsymbol{\lambda})} (1 - p F(t_i \mid \boldsymbol{\lambda})) \left[p f(t_i \mid \boldsymbol{\lambda}) \left(\theta + \frac{1}{1 - p F(t_i \mid \boldsymbol{\lambda})} \right) \right]^{\delta_i} \right\}. \quad (2.29)$$

2. Modelo destrutivo de Poisson com pesos exponenciais (DEWPM)

Considere $w(m \mid \phi) = e^{\phi m}$, $\phi \in \mathbb{R}$. Neste caso, $E_\theta[w(M \mid \phi)] = E_\theta[e^{\phi M}] = e^{\theta(e^\phi - 1)}$. Assim, (2.20) assume a forma

$$\begin{aligned} P(M = m \mid \theta, \phi) &= \frac{e^{\phi m} \theta^m e^{-\theta}}{m! e^{\theta(e^\phi - 1)}}, \\ &= \frac{(\theta e^\phi)^m e^{-\theta e^\phi}}{m!}, \quad m = 0, 1, 2, \dots, \end{aligned}$$

i.e., $M \sim Po(\theta e^\phi)$. Similarmente, $E_{\theta(1-p)}[w(M+m \mid \phi)] = E_{\theta(1-p)}[e^{\phi(M+m)}] = e^{\phi m + \theta(1-p)(e^\phi - 1)}$ e usando (2.22) tem-se que

$$\begin{aligned} P(D = d \mid \theta, p, \phi) &= \frac{(\theta p)^d e^{-\theta p} e^{\phi d + \theta(1-p)(e^\phi - 1)}}{d! e^{\theta(e^\phi - 1)}}, \\ &= \frac{(\theta p e^\phi)^d e^{-\theta p e^\phi}}{d!}, \quad d = 0, 1, 2, \dots, \end{aligned}$$

Portanto, $D \sim Po(\theta p e^\phi)$. Finalmente, usando (2.23), tem-se que a função de sobrevivência populacional do modelo destrutivo de Poisson com pesos exponenciais é dada por

$$S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) = \frac{e^{-\theta p F(t \mid \boldsymbol{\lambda})} e^{\theta(1-p F(t \mid \boldsymbol{\lambda}))(e^\phi - 1)}}{e^{\theta(e^\phi - 1)}} = \exp\{-\theta p e^\phi F(t \mid \boldsymbol{\lambda})\},$$

e as funções de densidade e de risco são dadas, respectivamente, por

$$\begin{aligned} f_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \theta p e^\phi f(t \mid \boldsymbol{\lambda}) \exp\{-\theta p e^\phi F(t \mid \boldsymbol{\lambda})\} \quad \text{e} \\ h_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \theta p e^\phi f(t \mid \boldsymbol{\lambda}). \end{aligned}$$

A fração de cura do modelo é dada por

$$S(\infty | \boldsymbol{\lambda}, \theta, p) = \lim_{t \rightarrow \infty} S_{pop}(t | \boldsymbol{\lambda}, \theta, p) = \exp\{-\theta p e^\phi\}.$$

Entretanto, pode-se verificar que a função de sobrevivência dos indivíduos suscetíveis é dada por

$$S_{sus}(t | \boldsymbol{\lambda}, \theta, p, \phi) = \frac{\exp\{-\theta p e^\phi F(t | \boldsymbol{\lambda})\} - \exp\{-\theta p e^\phi\}}{1 - \exp\{-\theta p e^\phi\}}. \quad (2.30)$$

Note que (2.30) é idêntica a (2.15) mudando o parâmetro θ por $\theta p e^\phi$.

Além disso, usando cálculos similares aos indicados anteriormente, verifica-se que

$$\begin{aligned} M - \delta | t, \delta &\sim Po\left(\theta e^\phi (1 - pF(t | \boldsymbol{\lambda}))\right), \\ D - \delta | t, \delta &\sim Po\left(\theta p e^\phi S(t | \boldsymbol{\lambda})\right), \\ D - \delta | M = m, t, \delta &\sim Bin\left(m - \delta, \frac{pS(t | \boldsymbol{\lambda})}{1 - pF(t | \boldsymbol{\lambda})}\right). \end{aligned} \quad (2.31)$$

Dessa forma, dada a informação observada, tem-se que as esperanças condicionais de M, D e $D | M = m$ são respectivamente

$$\begin{aligned} E(M | \mathbf{D}_{obs}) &= \delta + \theta e^\phi (1 - pF(t | \boldsymbol{\lambda})), \\ E(D | \mathbf{D}_{obs}) &= \delta + \theta p e^\phi S(t | \boldsymbol{\lambda}), \\ E(D | M = m, \mathbf{D}_{obs}) &= (m - \delta) \frac{p(1 - p)S(t | \boldsymbol{\lambda})}{\{1 - pF(t | \boldsymbol{\lambda})\}^2}. \end{aligned}$$

Por outro lado, usando (2.27), tem-se que

$$\begin{aligned} f(\mathbf{t}, \boldsymbol{\delta}, \mathbf{d} | \boldsymbol{\lambda}, \theta, p, \phi) &= \prod_{i=1}^n f(t_i, \delta_i, d_i | \boldsymbol{\lambda}, \theta, p, \phi), \\ &= \prod_{i=1}^n f(t_i, \delta_i | D_i = d_i, \boldsymbol{\lambda}) P(D_i = d_i | \theta, p, \phi), \\ &= \prod_{i=1}^n S(t_i | \boldsymbol{\lambda})^{d_i - \delta_i} [d_i f(t_i | \boldsymbol{\lambda})]^{\delta_i} \frac{(\theta p e^\phi)^{d_i} e^{-\theta p e^\phi}}{d_i!}. \end{aligned}$$

Portanto, a verossimilhança aumentada do modelo é dada por

$$L(\boldsymbol{\lambda}, \theta, p, \phi | \mathbf{D}_c) = \prod_{i=1}^n S(t_i | \boldsymbol{\lambda})^{D_i - \delta_i} [D_i f(t_i | \boldsymbol{\lambda})]^{\delta_i} \frac{(\theta p e^\phi)^{D_i} e^{-\theta p e^\phi}}{D_i!}. \quad (2.32)$$

Similarmente ao caso anterior, a verossimilhança observada pode ser construída somando (2.32) sob o vetor $\mathbf{D} = (D_1, \dots, D_n) \in \{0, 1, 2, \dots\}^n$. Usando (2.31), verifica-se que esta última verossimilhança é dada por

$$L(\boldsymbol{\lambda}, \theta, p, \phi | \mathbf{D}_c) = \prod_{i=1}^n \left\{ \exp\{-\theta p e^\phi F(t_i | \boldsymbol{\lambda})\} \left[\theta p e^\phi f(t_i | \boldsymbol{\lambda}) \right]^{\delta_i} \right\}.$$

3. Modelo destrutivo binomial negativo (DBN)

Considere $w(m \mid \phi) = \Gamma(\phi^{-1} + m)$, $\phi > 0$ e defina $\theta_1 = \frac{\phi\theta}{1+\phi\theta}$. Assim,

$$\begin{aligned} E_{\theta_1}[\Gamma(\phi^{-1} + m)] &= \sum_{j=0}^{\infty} \frac{e^{-\theta_1} \theta_1^j}{j!} \Gamma(\phi^{-1} + j), \\ &= \Gamma(\phi^{-1}) e^{-\theta_1} (1 - \theta_1)^{-\phi^{-1}} \sum_{j=0}^{\infty} \frac{\Gamma(\phi^{-1} + j)}{\Gamma(\phi^{-1}) j!} \theta_1^j (1 - \theta_1)^{\phi^{-1}}, \\ &= \Gamma(\phi^{-1}) e^{-\theta_1} (1 - \theta_1)^{-\phi^{-1}}. \end{aligned}$$

Portanto, usando (2.20) temos que

$$\begin{aligned} P(M = m \mid \theta, p, \phi) &= \frac{e^{-\theta_1} \theta_1^m \Gamma(\phi^{-1} + m)}{m! \Gamma(\phi^{-1}) e^{-\theta_1} (1 - \theta_1)^{-\phi^{-1}}}, \\ &= \frac{\Gamma(\phi^{-1} + m)}{\Gamma(\phi^{-1}) m!} \theta_1^m (1 - \theta_1)^{\phi^{-1}}. \end{aligned}$$

Desta forma, $M \sim BN(\phi, \theta_1)$. Ou seja, $E(M) = \phi^{-1} \frac{1-\theta_1}{\theta_1} = \theta$ e $\text{Var}(M) = \phi^{-1} \frac{1-\theta_1}{\theta_1^2} = \theta(1 + \phi\theta)$. Similarmente,

$$E_{\theta_1(1-p)}[\Gamma(\phi^{-1} + m + M)] = \frac{\Gamma(\phi^{-1} + m) e^{-\theta_1(1-p)}}{(1 - \theta_1(1-p))^{\phi^{-1}+m}}.$$

Note que

$$\begin{aligned} 1 - \theta_1(1-p) &= \frac{1 + \phi\theta p}{1 + \phi\theta}, \\ 1 - \theta_1 &= \frac{1}{1 + \phi\theta}, \\ \theta_1 p &= \frac{\phi\theta p}{1 + \phi\theta}. \end{aligned}$$

Portanto, utilizando (2.22), conclui-se que

$$\begin{aligned} P(D = d \mid \theta_1, p, \phi) &= \frac{(\theta_1 p)^d e^{-\theta_1 p} \Gamma(\phi^{-1}) e^{-\theta_1(1-p)} (1 - \theta)^{\phi^{-1}}}{d! \Gamma(\phi^{-1}) (1 - \theta_1(1-p))^{\phi^{-1}+d} e^{-\theta_1}}, \\ &= \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1}) d!} \left(\frac{\phi\theta p / (1 + \phi\theta)}{(1 + \phi\theta p) / (1 + \phi\theta)} \right)^d \left(\frac{1 / (1 + \phi\theta)}{(1 + \phi\theta p) / (1 + \phi\theta)} \right)^{-\phi^{-1}}, \\ &= \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1}) d!} \left(\frac{\phi\theta p}{1 + \phi\theta p} \right)^d (1 + \phi\theta p)^{-\phi^{-1}}. \end{aligned}$$

Assim, $D \sim BN\left(\phi, \frac{\phi\theta p}{1+\phi\theta p}\right)$, i.e., $E(D) = \theta p$ e $\text{Var}(D) = \theta p(1 + \phi\theta p)$. Note que como $\phi > 0$, então este modelo (diferentemente do modelo Binomial Negativo proposto em [Rodrigues et al. \[2011\]](#)) só admite que a distribuição de M e de D apresente superdispersão, limitando sua aplicação. Finalmente, usando (2.23) tem-se que

$$\begin{aligned} S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \frac{e^{-\theta_1 p F(t \mid \boldsymbol{\lambda})} \Gamma(\phi^{-1}) e^{-\theta_1(1-p F(t \mid \boldsymbol{\lambda}))} (1 - \theta_1(1 - p F(t \mid \boldsymbol{\lambda})))^{-\phi^{-1}}}{\Gamma(\phi^{-1}) e^{-\theta_1} (1 - \theta_1)^{-\phi^{-1}}}, \\ &= \{1 + \phi\theta p F(t \mid \boldsymbol{\lambda})\}^{-\phi^{-1}}, \end{aligned}$$

e as funções de densidade e de risco estão dadas respectivamente por

$$\begin{aligned} f_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \theta p f(t \mid \boldsymbol{\lambda}) \{1 + \phi \theta p F(t \mid \boldsymbol{\lambda})\}^{-\phi^{-1}-1} \quad \text{e} \\ h_{pop}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \frac{\theta p f(t \mid \boldsymbol{\lambda})}{1 + \phi \theta p F(t \mid \boldsymbol{\lambda})}. \end{aligned}$$

A fração de cura do modelo é dada por

$$S(\infty \mid \boldsymbol{\lambda}, \theta, p) = \lim_{t \rightarrow \infty} S_{pop}(t \mid \boldsymbol{\lambda}, \theta, p) = \{1 + \phi \theta p\}^{-\phi^{-1}}.$$

A função de sobrevivência dos não curados é dada por

$$S_{sus}(t \mid \boldsymbol{\lambda}, \theta, p, \phi) = \frac{\{1 + \phi \theta p F(t \mid \boldsymbol{\lambda})\}^{-\phi^{-1}} - \{1 + \phi \theta p\}^{-\phi^{-1}}}{1 - \{1 + \phi \theta p\}^{-\phi^{-1}}}.$$

Como anteriormente, pode-se verificar que

$$\begin{aligned} M - \delta \mid t, \delta &\sim BN \left((\phi^{-1} + \delta)^{-1}, \frac{\phi \theta (1 - p F(t \mid \boldsymbol{\lambda}))}{1 + \phi \theta} \right), \\ D - \delta \mid t, \delta &\sim BN \left((\phi^{-1} + \delta)^{-1}, \frac{\phi \theta p S(t \mid \boldsymbol{\lambda})}{1 + \phi \theta} \right), \\ D - \delta \mid M = m, t, \delta &\sim Bin \left(m - \delta, \frac{p S(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda})} \right). \end{aligned} \quad (2.33)$$

Detalhes destes resultados encontram-se no Apêndice A.1.2. Finalmente, dada a informação observada, tem-se que

$$\begin{aligned} E(M \mid \mathbf{D}_{obs}) &= \frac{\delta(1 + \phi \theta) + \theta(1 - p F(t \mid \boldsymbol{\lambda}))}{1 + \phi \theta p F(t \mid \boldsymbol{\lambda})}, \\ E(D \mid \mathbf{D}_{obs}) &= \frac{\delta(1 + \phi \theta) + \theta p S(t \mid \boldsymbol{\lambda})}{1 + \phi \theta (1 - p S(t \mid \boldsymbol{\lambda}))}, \\ E(D \mid M = m, \mathbf{D}_{obs}) &= (m - \delta) \frac{p(1 - p) S(t \mid \boldsymbol{\lambda})}{(1 - p F(t \mid \boldsymbol{\lambda}))^2}. \end{aligned}$$

Por outro lado, (2.27) implica que

$$\begin{aligned} f(\mathbf{t}, \boldsymbol{\delta}, \mathbf{d} \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \prod_{i=1}^n f(t_i, \delta_i, d_i \mid \boldsymbol{\lambda}, \theta, p, \phi), \\ &= \prod_{i=1}^n f(t_i, \delta_i \mid D_i = d_i, \boldsymbol{\lambda}) P(D_i = d_i \mid \theta, p, \phi), \\ &= \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda})^{d_i - \delta_i} [d_i f(t_i \mid \boldsymbol{\lambda})]^{\delta_i} \frac{\Gamma(\phi^{-1} + d_i)}{\Gamma(\phi^{-1}) d_i!} \left(\frac{\phi \theta p}{1 + \phi \theta p} \right)^{d_i} (1 + \phi \theta p)^{-\phi^{-1}}. \end{aligned}$$

Portanto, a verossimilhança aumentada do modelo é dada por

$$L(\boldsymbol{\lambda}, \theta, p, \phi \mid \mathbf{D}_c) = \prod_{i=1}^n S(t_i \mid \boldsymbol{\lambda})^{D_i - \delta_i} [D_i f(t_i \mid \boldsymbol{\lambda})]^{\delta_i} \frac{\Gamma(\phi^{-1} + D_i)}{\Gamma(\phi^{-1}) D_i!} \left(\frac{\phi \theta p}{1 + \phi \theta p} \right)^{D_i} (1 + \phi \theta p)^{-\phi^{-1}}. \quad (2.34)$$

Finalmente, a verossimilhança observada pode ser construída somando (2.34) sob o vetor $\mathbf{D} = (D_1, \dots, D_n)$ no conjunto $\{0, 1, 2, \dots\}^n$. Usando (2.33), verifica-se que esta última

verossimilhança é dada por

$$L(\boldsymbol{\lambda}, \theta, p, \phi \mid \mathbf{D}_c) = \prod_{i=1}^n \left\{ \{1 + \phi \theta p F(t_i \mid \boldsymbol{\lambda})\}^{-\phi^{-1}} \left[\frac{\theta p f(t_i \mid \boldsymbol{\lambda})}{1 + \phi \theta p F(t_i \mid \boldsymbol{\lambda})} \right]^{\delta_i} \right\}.$$

Em todos os modelos destrutivos apresentados é possível incluir um vetor de covariáveis $\mathbf{Z}$ que explicam a fração de cura, particionando esse vetor em $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ (em que $\mathbf{Z}_1$ e $\mathbf{Z}_2$ não compartilham elementos em comum) e fazendo

$$\log(\theta_i) = \vartheta_{1i} = \mathbf{z}_{1i}^\top \boldsymbol{\beta}_1 \quad \text{e} \quad \log\left(\frac{p_i}{1 - p_i}\right) = \vartheta_{2i} = \mathbf{z}_{2i}^\top \boldsymbol{\beta}_2, \quad i = 1, \dots, n, \quad (2.35)$$

em que $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são vetores desconhecidos de dimensão $s_1 \times 1$ e $s_2 \times 1$ respectivamente.

Adicionalmente, um conjunto $\mathbf{X}$ de covariáveis observadas no i -ésimo indivíduo também podem ser incorporadas através da função $F(t \mid \boldsymbol{\lambda})$ (ou da função $S(t \mid \boldsymbol{\lambda})$) para modelar a sobrevivência dos indivíduos suscetíveis como em (2.8), mas interpretando estes coeficientes em termos da função de sobrevivência das causas concorrentes que não foram destruídas.

Em relação à distribuição utilizada para modelar os tempos de vida das variáveis Y_k , $k = 1, \dots, D$, [Rodrigues et al. \[2011\]](#) utilizaram a distribuição Weibull.

Por outro lado, como apontado em [Rodrigues et al. \[2011\]](#), nos modelos DEWPM e DBN, o vetor $\mathbf{Z}_1$ não pode conter intercepto para não produzir problemas de identificabilidade no sentido de [Li et al. \[2001\]](#). Por exemplo, no caso do modelo DEWPM, se $\mathbf{Z}_1$ incluir intercepto, então a função de sobrevivência populacional do i -ésimo indivíduo assumiria a forma

$$\begin{aligned} S_{pop}(t \mid \boldsymbol{\lambda}, \eta_i, \vartheta_{1i}, \vartheta_{2i}) &= \exp\{-e^\phi \theta_i p_i F(t \mid \boldsymbol{\lambda})\}, \\ &= \exp\{-p_i F(t \mid \boldsymbol{\lambda}) \exp\{\phi + \beta_{10} + z_{11i}\beta_{11} + \dots + z_{1s_1i}\beta_{1s_1i}\}\}. \end{aligned}$$

Desta forma, só é possível identificar a quantidade $\phi + \beta_{10}$, mas não ϕ ou β_{10} separadamente (a menos que um deles for conhecido). Analogamente, pode-se verificar que este problema ocorre para o modelo DBN.

Finalmente, embora não comentado por seus autores, no modelo DLBP também não é possível incluir intercepto simultaneamente nos vetores $\mathbf{Z}_1$ e $\mathbf{Z}_2$, pois do contrário, a fração de cura do modelo não é identificável, o que é uma condição necessária para tornar o modelo identificável. Para exemplifica-lo, suponha o caso mais simples:

- $\mathbf{z}_{1i}^\top = (1, z_{1i})$;
- $\mathbf{z}_{2i}^\top = (1, z_{2i})$;
- z_{1i} e z_{2i} são variáveis binárias. Dessa forma, só existem quatro possíveis combinações para o par $(z_{1i}, z_{2i}) : (0, 0), (0, 1), (1, 0)$ e $(1, 1)$.

No Apêndice A.1.3 encontra-se uma prova de que sob estas condições, existem infinitos vetores $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ que permitem obter a mesma fração de cura do modelo.

2.2.5 Outros modelos de cura

As seções 2.2.1 a 2.2.4 discutem em detalhes os principais modelos que serão utilizados neste trabalho. Porém, o desenvolvimento destes modelos tem sido bastante mais amplo na literatura,

tendo contribuído muitos exponentes brasileiros nesta linha de pesquisa.

Zhang e Peng [2009] propõem a classe de modelos de cura com risco acelerado, o qual introduz as covariáveis $\mathbf{X}$ associadas à sobrevivência dos suscetíveis fazendo

$$S(t | \boldsymbol{\lambda}, \eta_i) = S_0(te^{\eta_i}), \quad i = 1, \dots, n,$$

ao invés de utilizar a estrutura em (2.8).

Cancho et. al [2013c] propõem outra classe de modelos utilizando para a variável latente M que representa o número de células do indivíduo, a classe de modelos baseada em séries de potência na forma

$$P(M = m | \theta) = \frac{a_m \theta^m}{A(\theta)}, \quad m = 0, 1, 2, \dots,$$

em que $a_m > 0$ e $A(\theta) = \sum_{m=0}^{\infty} a_m \theta^m$. Esta classe de modelos inclui como caso particular os modelos de Berkson e Gage [1952] ($A(\theta) = 1 + \theta$), Chen et al. [1999a] ($A(\theta) = e^\theta$) e o modelo geométrico ($A(\theta) = (1 - \theta)^{-1}$), além de incluir um novo modelo que é denominado o modelo logarítmico ($A(\theta) = -\log(1 - \theta)/\theta$).

Rodrigues et al. [2009b] utilizam uma abordagem similar à proposta em Rodrigues et al. [2009a], modificando a distribuição usada para a variável M usando a distribuição de Poisson de Conway-Maxwell (COM-Poisson), cuja função de probabilidade é dada por

$$P(M = m; \theta, \phi) = \frac{1}{Z(\theta, \phi)} \frac{\theta^m}{(m!)^\phi}, \quad m = 0, 1, 2, \dots,$$

em que $Z(\theta, \phi) = \sum_{j=0}^{\infty} \theta^j / (j!)^\phi$. Este modelo também tem como caso particular os modelos de Berkson e Gage [1952] ($\phi = 1$), Chen et al. [1999a] ($\phi \rightarrow \infty$) e o modelo Geométrico ($\phi = 0$ e $\theta < 1$). Rodrigues et al. [2009b] discute uma abordagem clássica deste modelo, enquanto que a versão bayesiana pode ser vista em Cancho et. al [2012a]. O principal inconveniente de utilizar esta distribuição é o fato que $Z(\theta, \phi)$ não ter forma fechada, o que aumenta o custo computacional consideravelmente.

Uma característica compartilhada pelos modelos expostos em detalhe nas seções 2.2.1 à 2.2.4, é o fato de supor que a falha do indivíduo é dada quando uma das componentes falha, i.e, o mínimo dos tempos de falha associados às causas concorrentes (esquema de activação mínima). Um esquema de activação mais geral pode ser encontrado em Cancho et. al [2013a] e Cancho et. al [2013b], em que supõe-se que a falha do indivíduo é dada quando todas as componentes deles falham, i.e., o máximo dos tempos de falha associados às causas concorrentes (esquema de activação máxima). Finalmente, também é considerado um esquema em que o número de componentes necessárias para produzir a falha do esquema é aleatório (esquema de activação aleatório).

Uma abordagem de cura ainda mais complexa pode ser encontrada em Borges et al. [2012]. No contexto do câncer, é assumido que existem três tipos de células: células iniciais, células malignas e células tumorais, em que N_{1i}, N_{2i} e N_{3i} denotam o número de cada uma delas para o i -ésimo indivíduo, em que $N_{3i} \leq N_{2i} \leq N_{1i}$. Sob algumas suposições razoáveis, monta-se um esquema que permite introduzir covariáveis para explicar cada uma das componentes do modelo, fazer inferência sobre a distribuição de cada uma das variáveis latentes envolvidas e explicar a função de sobrevivência para um indivíduo com características determinadas.

2.2.6 Comentários adicionais sobre os modelos de cura

Há alguns cuidados que devem ser tomados quando trabalha-se com os modelos de cura. Um deles é o fato de ter-se a certeza de que deve ser utilizado um modelo de cura ao invés de um modelo de sobrevivência usual. Por exemplo, no modelo de mistura precisa-se testar a hipótese $\pi = 1$. Note que esta hipótese pertence à borda do espaço paramétrico para π , portanto se for utilizada a estatística da razão de verossimilhanças deve-se modificar a distribuição usual dessa estatística (Qui-quadrado) e usar uma mistura entre uma v.a. degenerada em 0 e uma Qui-quadrado com um grau de liberdade (Maller e Zhou [1996]). Estes autores também apresentam um outro teste para esta hipótese baseado em estudos de simulação.

Outra questão importante destes modelos é se o tempo de seguimento do estudo é suficientemente longo, pois do contrário podem-se produzir problemas de estimação. Concretamente, para tempos insuficientes de seguimento, pode ser difícil distinguir entre uma chance de cura elevada com uma cauda leve para $S_0(t | \boldsymbol{\lambda})$ e uma baixa chance de cura com uma cauda pesada para $S_0(t | \boldsymbol{\lambda})$. Desta forma, diferentes escolhas para $S_0(t | \boldsymbol{\lambda})$ podem levar a estimativas muito diferentes para as frações de cura (Yu et al. [2004]), além do que as variâncias dos parâmetros estimados tendem a ser grandes (Farewell [1986]).

Por outro lado, como comentado em Klein et al. [2013], no caso de incluir covariáveis na parte suscetível do modelo (por exemplo, como em (2.8)), a estimação dos parâmetros deve ser feita em duas etapas: primeiro fixando $S_0(t | \boldsymbol{\lambda})$ e estimando os demais parâmetros do modelo e depois estimando $S_0(t | \boldsymbol{\lambda})$ com o demais parâmetros remanescentes. Se a maximização não for feita desta forma, o algoritmo de maximização pode produzir estimativas errôneas (máximos locais).

Finalmente, uma questão muito relevante nesta classe de modelos é a identificabilidade dos parâmetros. Li et al. [2001] discutem a identificabilidade do modelo de mistura e do modelo de Chen et al. [1999a], concluindo que em ambos modelos é necessário incluir covariáveis $\mathbf{Z}$ na cura para torná-lo identificável (equações (2.10) e equações (3.21)). Mas não discutem a identificabilidade do modelo caso tiver também covariáveis $\mathbf{X}$ para os indivíduos suscetíveis como em (2.8). Peng e Zhang [2008] discutem a identificabilidade do modelo de mistura com fragilidade quando são incluídas covariáveis $\mathbf{X}$ e $\mathbf{Z}$ no modelo para modelar a heterogeneidade dos suscetíveis e curados respectivamente. Verifica-se nesse cenário (com certas suposições sobre a distribuição da fragilidade, entre outras, que a esperança dessa distribuição seja 1) que o modelo é identificável, mesmo que $\mathbf{X} = \mathbf{Z}$. Dessa forma (e como mencionado em Peng e Zhang [2008]), garante-se a identificabilidade do modelo de mistura quando são incluídas covariáveis $\mathbf{X}$ e $\mathbf{Z}$ no modelo como explicado anteriormente, pois o modelo de mistura com fragilidade tem como caso particular o modelo de mistura quando todas as variáveis de fragilidade são v.a.'s degeneradas em 1.

Desta forma, pode-se garantir a identificabilidade no modelo de Chen et al. [1999a] incluindo covariáveis $\mathbf{X}$ e $\mathbf{Z}$ para os indivíduos suscetíveis e curados usando a identificabilidade do modelo de mistura. Pelas equações (2.16) e (2.10), a fração de cura do modelo de Chen et al. [1999a] é dada por $\pi_i = \exp\{-\theta_i\} = \exp\{-\exp\{\mathbf{z}_i^\top \boldsymbol{\beta}\}\}$, entanto que a fração de cura do modelo de mistura é dada por $\pi_i = \exp\{\mathbf{z}_i^\top \boldsymbol{\beta}^*\} / (1 + \exp\{\mathbf{z}_i^\top \boldsymbol{\beta}^*\})$. Note que há uma relação um a um entre π_i e θ_i para um $\mathbf{z}_i$ fixo, pelo que se garanta uma relação biunívoca entre $\boldsymbol{\beta}$ e $\boldsymbol{\beta}^*$ para um $\mathbf{z}_i$ fixo.

Por outro lado, note que a expressão da função de sobrevivência (equação (2.14)) no modelo de Chen et al. [1999a] pode ser escrita na forma

$$S_{pop}(t | \theta_i, \eta_i, \boldsymbol{\lambda}) = e^{-\theta_i F(t | \boldsymbol{\lambda}, \eta_i)} = e^{-\theta_i} + (1 - e^{-\theta_i}) \left(\frac{e^{-\theta_i F(t | \boldsymbol{\lambda}, \eta_i)} - e^{-\theta_i}}{1 - e^{-\theta_i}} \right), \quad (2.36)$$

em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\gamma}$, $F(t | \boldsymbol{\lambda}, \eta_i) = 1 - S(t | \boldsymbol{\lambda}, \eta_i)$ e $S(t | \boldsymbol{\lambda}, \eta_i)$ provém de (2.8). Comparando a equação

(2.36) com a função de sobrevivência populacional no modelo de mistura em (2.6), conclui-se que a função de sobrevivência dos indivíduos suscetíveis é dada por $(e^{-\theta_i F(t|\boldsymbol{\lambda}, \eta_i)} - e^{-\theta_i}) / (1 - e^{-\theta_i})$, como visto na equação (2.15). Além disso, suponha que $S(t | \boldsymbol{\lambda}^*, \eta_i^*)$ representa a função de sobrevivência dos indivíduos suscetíveis no modelo de mistura, em que $\eta_i^* = \mathbf{x}_i^\top \boldsymbol{\gamma}^*$.

Neste caso, para θ_i e $\mathbf{x}_i$ fixos, há uma relação um a um entre as funções $S_0(t | \boldsymbol{\lambda}, \eta_i)$ e $S_0(t | \boldsymbol{\lambda}^*, \eta_i^*)$. Finalmente, o fato de $S_0(t | \boldsymbol{\lambda}, \eta_i)$ pertencer à classe de riscos proporcionais garante a relação um a um entre $\boldsymbol{\gamma}$ e $\boldsymbol{\gamma}^*$ e entre $\boldsymbol{\lambda}$ e $\boldsymbol{\lambda}^*$.

Dessa forma, a relação bijectiva entre as funções de sobrevivência populacional dos modelos de mistura e de Chen garante a identificabilidade do último modelo quando são introduzidas covariáveis na componente curada e suscetível do modelo. Argumentos similares podem ser usados para garantir a identificabilidade dos demais modelos de cura apresentados nesta seção.

2.3 Estimadores BLUP e REML

Os estimadores BLUP (*Best linear unbiased predictor*) foram generalizados por Kuk [1994] no contexto dos modelos mistos. Seja $\mathbf{Y}$ a variável de interesse e $\mathbf{K}$ o vetor de efeitos aleatórios do modelo, em que $f_{\mathbf{Y}|\mathbf{K}}(\cdot | \mathbf{K})$ é a densidade de $\mathbf{Y}$ uma vez fixados os efeitos aleatórios, $f_{\mathbf{K}}(\cdot)$ é a densidade dos efeitos aleatórios e $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)$ é o vetor de parâmetros desconhecidos.

Na prática, é usual poder obter uma partição do vetor $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$ tal que $f(\mathbf{y} | \mathbf{K})$ dependa só de $\boldsymbol{\beta}$ e $f(\mathbf{K})$ dependa só de $\boldsymbol{\Sigma}$. Assuma que $\hat{\boldsymbol{\psi}}$ é uma estimativa inicial para $\boldsymbol{\psi}$, a qual é solução do conjunto de p equações de estimação que escreveremos na forma

$$\varphi(\boldsymbol{\psi}; \mathbf{y}) = 0.$$

Por exemplo, pode-se escolher $\hat{\boldsymbol{\psi}}$ que maximize a função $l^*(\boldsymbol{\psi}; \mathbf{y})$, em que

$$l^*(\boldsymbol{\psi}; \mathbf{y}) = \log l(\boldsymbol{\psi}; \mathbf{y}, \mathbf{K}^*),$$

$\log l(\boldsymbol{\psi}; \mathbf{y}, \mathbf{K})$ é a função de log-verossimilhança baseada na distribuição conjunta de $\mathbf{y}$ e $\mathbf{K}$ e $\mathbf{K}^* = \mathbf{K}^*(\mathbf{y}, \boldsymbol{\psi})$ é algum valor imputado para o vetor $\mathbf{K}$ não observado e que pode depender tanto de $\mathbf{y}$ quanto de $\boldsymbol{\psi}$. Dessa forma, a correspondente equação de estimação é dada pela solução do sistema de equações

$$l'^*(\boldsymbol{\psi}; \mathbf{y}) = 0 \quad (2.37)$$

em que l'^* denota o vetor de derivadas de l^* em relação às componentes do vetor $\boldsymbol{\psi}$. Evidentemente, diferentes escolhas para $\mathbf{K}^*(\mathbf{y}; \boldsymbol{\psi})$ produzirão diferentes equações de estimação e, portanto, diferentes estimativas para o vetor de parâmetros $\boldsymbol{\psi}$.

Em particular, os estimadores BLUP são obtidos maximizando $l(\boldsymbol{\psi}; \mathbf{y}, \mathbf{K})$ com respeito a $\mathbf{K}$. Isto corresponde a escolher $\mathbf{K}^* = \hat{\mathbf{K}}(\mathbf{y}, \boldsymbol{\Sigma})$ em (2.37) para um $\boldsymbol{\Sigma}$ fixo. Seja $\hat{\boldsymbol{\psi}}$ o estimador de $\boldsymbol{\psi}$ que é solução de (2.37) e defina $\hat{\mathbf{A}} = \left(-\frac{\partial l'^*(\boldsymbol{\psi}; \mathbf{y})}{\partial \boldsymbol{\psi}} \Big|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} \right)^{-1}$, em que $\hat{\mathbf{A}}$ pode ser particionado conforme a ordenação $\boldsymbol{\beta} | \mathbf{K}$ em

$$\hat{\mathbf{A}} = \begin{pmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{12} & \hat{\mathbf{A}}_{22} \end{pmatrix}. \quad (2.38)$$

Como mencionado em McGilchrist e Yau [1995], a distribuição assintótica de $\hat{\boldsymbol{\beta}}$ é dada por

$$\hat{\mathbf{A}}_{11}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}), \quad \text{quando } n \rightarrow \infty. \quad (2.39)$$

É importante destacar que o resultado anterior é considerado no contexto dos modelos lineares generalizados, i.e., em um contexto que não envolve observações censuradas como será estudado neste trabalho. Porém, no contexto em que tem-se observações censuradas, pode-se verificar com estudos de simulação que o resultado em (2.39) continua sendo válido.

Por outro lado, os estimadores REML (*Restricted Maximum Likelihood*) foram unificados por McGilchrist e Yau [1995] em conjunto com os estimadores BLUP e de máxima verossimilhança no contexto de modelos lineares generalizados com efeitos aleatórios normais, mas essa mesma ideia pode ser utilizada em outros contextos de modelos com efeitos aleatórios, desde que a distribuição utilizada para os efeitos aleatórios seja a distribuição Normal. Suponha que o vetor de efeitos aleatórios pode ser particionado em $\mathbf{K} = (\mathbf{K}_1, \dots, \mathbf{K}_J)$, em que $\mathbf{K}_j \sim N_s(0, \mathbf{\Sigma})$, $j = 1, \dots, J$, $\mathbf{\Sigma}$ depende de t parâmetros caracterizados por $\phi_1, \dots, \phi_t$ e $\mathbf{K}_i$ é independente de $\mathbf{K}_j$, se $1 \leq i \neq j \leq J$. Desta forma, o vetor $(\mathbf{K}_1, \dots, \mathbf{K}_J)$ terá distribuição $N_{sJ}(0, \mathbf{\Omega})$, em que $\mathbf{\Omega} = \mathbf{\Sigma} \otimes I_J$ ($\otimes$ representa o produto de kronecker). Além disso, suponha que o estimador BLUP de $\mathbf{K}$, digamos $\widehat{\mathbf{K}}$, tem matriz de variância-covariância estimada dada por $\widehat{\mathbf{A}}$ como em (2.38). Nessas circunstâncias, McGilchrist e Yau [1995] provaram que os estimadores REML para $\phi_1, \dots, \phi_t$ são solução das equações de estimação

$$\text{tr} \left(\mathbf{\Omega}^{-1} \frac{\partial \mathbf{\Omega}}{\partial \phi_i} \right) + \text{tr} \left(\left(\widehat{\mathbf{A}}_{22} + \widehat{\mathbf{K}} \widehat{\mathbf{K}}^\top \right) \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_i} \right) = 0, \quad i = 1, \dots, t.$$

Em alguns contextos, utilizar este procedimento de estimação em duas etapas pode ser muito conveniente em termos computacionais, sobretudo quando eliminar por integração os e.a.'s $\mathbf{K}$ é altamente custoso. Por exemplo, mesmo no caso mais simples como é a distribuição Normal para os e.a.'s, calcular a expressão $\int \log l(\boldsymbol{\psi}; \mathbf{y}, \mathbf{K}) d\mathbf{K}$ pode ser analiticamente intratável (com exceção de alguns modelos), o que obriga a utilizar métodos de aproximação para essas integrais, como a quadratura de Gauss-Hermite (Fahrmeir e Tutz [1994]) ou a quadratura de Laplace (Pinheiro [2006]). Porém, há erros de estimação associados a estes métodos, os que certamente podem comprometer as estimativas obtidas. Dessa forma, os estimadores REML fornecem uma alternativa viável, além de ter a vantagem de produzir estimadores menos viciados para as componentes da variância em relação aos estimadores de máxima verossimilhança (Lee e Nelder [2005]).

Por outro lado, a informação observada para os estimadores REML é dada por $2W$, em que os elementos da matriz W são dados por

$$W_{ij} = \text{tr} \left(\widehat{\mathbf{A}} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_i} \widehat{\mathbf{A}} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_j} + \mathbf{\Omega} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_i} \mathbf{\Omega} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_j} - 2 \widehat{\mathbf{A}} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_i} \mathbf{\Omega} \frac{\partial \mathbf{\Omega}^{-1}}{\partial \phi_j} \right), \quad 1 \leq i, j \leq t. \quad (2.40)$$

Se a matriz $\mathbf{\Sigma}$ é da forma $\mathbf{\Sigma} = \text{diag}\{\phi_1, \phi_2, \dots, \phi_s\}$ e a matriz $\widehat{\mathbf{A}}$ é particionada conforme à ordenação $(K_{11}, \dots, K_{1J}) \mid (K_{21}, \dots, K_{2J}) \mid \dots \mid (K_{s1}, \dots, K_{sJ})$ na forma

$$\widehat{\mathbf{A}} = \begin{pmatrix} \widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12} & \dots & \widehat{\mathbf{A}}_{1s} \\ \widehat{\mathbf{A}}_{21} & \widehat{\mathbf{A}}_{22} & \dots & \widehat{\mathbf{A}}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{A}}_{s1} & \widehat{\mathbf{A}}_{s2} & \dots & \widehat{\mathbf{A}}_{ss} \end{pmatrix},$$

então a matriz W se reduz a

$$W_{ij} = \phi_i^{-2} \{J - 2\phi_i^{-1} \text{tr}(\widehat{\mathbf{A}}_{ii})\} \delta_{ij}^\Delta + \phi_i^{-2} \phi_j^{-2} \text{tr}(\widehat{\mathbf{A}}_{ij} \widehat{\mathbf{A}}_{ji}), \quad 1 \leq i, j \leq s, \quad (2.41)$$

em que δ_{ij}^Δ é a função delta de Kronecker.

Por outro lado, a seleção de modelos é um ponto importante a ser tratado. O critério de Akaike

(AIC) e de Schwartz (BIC) tem sido bastante usados. Estes dois critérios assumem a forma

$$q(\boldsymbol{\beta}) = -2 \log l(\boldsymbol{\beta}) + \sharp \rho_c, \quad (2.42)$$

em que $\boldsymbol{\beta}$ representa o conjunto de parâmetros a estimar, ρ_c representa o número efectivo de parâmetros do modelo e $\sharp$ representa a penalização associada a cada critério. Particularmente, para o critério AIC $\sharp = 2$ e para o critério BIC $\sharp = \log(n)$. Desta forma, uma estimativa para os critérios AIC e BIC são dadas modificando $\boldsymbol{\beta}$ por sua respectiva estimativa em (2.42).

Porém, no contexto de modelos mistos, quantificar ρ_c não é tão intuitivo como no caso usual. Vaida e Blanchard [2013] fazem uma boa discussão ao respeito. Seguindo Donohue et. al [2011], para $\boldsymbol{\Sigma}$ fixada (por exemplo, aquela obtida via o uso dos estimadores REML explicados anteriormente) pode-se escrever o critério de Akaike condicional como

$$q(\boldsymbol{\psi}) = -2 \log l_1(\boldsymbol{\psi}) + \sharp \rho_c, \quad (2.43)$$

em que $\boldsymbol{\psi} = (\boldsymbol{\beta}, \mathbf{K})$, $\boldsymbol{\beta}$ e $\mathbf{K}$ representam os efeitos fixos e aleatórios do modelo como definido no começo desta seção, l_1 representa a verossimilhança construída considerando $\mathbf{K}$ como fixos, $\sharp = 2$ e ρ_c é dado por

$$\rho_c = \text{tr} \left[\mathbf{X}^{\top*} \boldsymbol{\Lambda} \mathbf{X}^* \left(\mathbf{X}^{\top*} \boldsymbol{\Lambda} \mathbf{X}^* + \text{diag} \left(0, -\frac{\partial^2 l_2}{\partial \mathbf{K} \partial \mathbf{K}^{\top}} \right) \right)^{-1} \right], \quad (2.44)$$

em que $\boldsymbol{\Lambda} = (-\partial^2 l_1(\boldsymbol{\psi}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\top})$ e l_2 representa o logaritmo da densidade $N(0, \boldsymbol{\Sigma})$ avaliada no vetor $\mathbf{K}$. Fazendo a analogia com o critério de Schwartz, poderia ser utilizado $\sharp = \log n$ em (2.43) para criar uma versão condicional deste critério. As matrizes $\boldsymbol{\Lambda}$ e $(-\partial^2 l_2 / \partial \mathbf{K} \partial \mathbf{K}^{\top})$ serão especificadas totalmente para cada modelo desenvolvido neste trabalho.

2.4 Processos de Dirichlet

No contexto dos modelos mistos, um dos principais problemas é a escolha da distribuição imposta para os efeitos aleatórios. Uma solução muito utilizada nos últimos anos para este problema, é o uso de distribuições *a priori* não paramétricas, que por sua vez, produzem também distribuições *a posteriori* não paramétricas. Diferente da inferência bayesiana paramétrica, essas distribuições são processos estocásticos. Esta seção busca apresentar ao leitor a esses processos específicos, denominados processos de Dirichlet (P.D.'s) e alguns métodos para amostrar deles.

Como mencionado por Ferguson [1973], um P.D. é uma classe de distribuições. Seja $\alpha \in \mathbb{R}^+$ e $\mathcal{G}_0$ uma medida de probabilidade (M.P.) sobre $(\mathcal{Q}, \mathcal{K})$, em que $\mathcal{Q}$ é um conjunto e $\mathcal{K}$ é uma σ -álgebra sobre $\mathcal{Q}$. Um P.D. sob $(\mathcal{Q}, \mathcal{K})$ com parâmetro $\alpha \mathcal{G}_0$ é uma medida de probabilidade aleatória (M.P.A.) $\mathcal{G}$ que atribui um número $\mathcal{G}(B)$ para todo $B \in \mathcal{K}$ tal que

- $\mathcal{G}(B) \in [0, 1]$, $\forall B \in \mathcal{K}$,
- cada realização de $\mathcal{G}$ é uma M.P. sobre $(\mathcal{Q}, \mathcal{K})$,
- para cada partição finita e mensurável $(B_1, \dots, B_k)$ de $\mathcal{Q}$, tem-se que

$$(\mathcal{G}(B_1), \dots, \mathcal{G}(B_k)) \sim \text{Dir}(\alpha \mathcal{G}_0(B_1), \dots, \alpha \mathcal{G}_0(B_k)),$$

em que $\text{Dir}(\alpha_1, \dots, \alpha_k)$ representa a distribuição Dirichlet.

Na literatura, α é conhecido como parâmetro de precisão, $\mathcal{G}_0$ é chamada de medida central e o produto $\alpha\mathcal{G}_0$ é a medida basal do processo. Será denotado simplesmente $\mathcal{G} \sim \text{PD}(\alpha\mathcal{G}_0)$. Seguindo Jara [2008], foram construídas as Figuras 2.7, 2.8, 2.9 e 2.10 usando como medida basal do processo as distribuições normal padrão, t de Student com 3 graus de liberdade, exponencial de taxa 1 e qui-quadrado com 1 grau de liberdade. Para cada distribuição, foi considerado o parâmetro de precisão α no conjunto $\{1, 10, 100, 1000\}$.

Das figuras, verifica-se que quando $\alpha \rightarrow \infty$, o processo converge para a medida basal $\mathcal{G}_0$. No entanto, quando $\alpha \rightarrow 0^+$, o processo converge para uma v.a. degenerada em um ponto K_0 , em que $K_0 \sim \mathcal{G}_0$.

Figura 2.7: Gráfico da f.d.a. para 100 amostras de tamanho 100 de $\text{PD}(\alpha\mathcal{G}_0)$, com $\alpha = 1$ e $\mathcal{G}_0$ a respectiva medida basal.

Blackwell e MacQueen [1973] fornecem uma representação alternativa dos P.D.'s. Nesta definição,

- $K_1 \mid \mathcal{G}_0 \sim \mathcal{G}_0$,
- $K_i \mid K_1, \dots, K_{i-1}, \alpha, \mathcal{G}_0 \sim \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{K_j}^* + \frac{\alpha}{\alpha + i - 1} \mathcal{G}_0, \quad i = 2, 3, \dots,$

em que δ_x^* é a função delta de Dirac no ponto x .

Note que se $\mathcal{G}$ e $\mathcal{H}$ são distribuições, então a notação $p\mathcal{G} + (1-p)\mathcal{H}$ representa aquela distribuição que é mistura de $\mathcal{G}$ e $\mathcal{H}$, com proporções de mistura p e $1-p$ respectivamente. Em outras palavras, a definição apresentada indica que K_i é gerado da medida $\mathcal{G}_0$ com probabilidade $\alpha/(\alpha + i - 1)$ e com probabilidade $1/(\alpha + j - 1)$ tem-se que $K_i = K_j$, $1 \leq j \leq i - 1$.

Esta representação é muito útil, sobretudo quando precisa-se amostrar de um P.D.

Figura 2.8: Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha \mathcal{G}_0)$, com $\alpha = 10$ e $\mathcal{G}_0$ a respectiva medida basal.

Figura 2.9: Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha \mathcal{G}_0)$, com $\alpha = 100$ e $\mathcal{G}_0$ a respectiva medida basal.

Figura 2.10: Gráfico da f.d.a. para 100 amostras de tamanho 100 de $PD(\alpha\mathcal{G}_0)$, com $\alpha = 1000$ e $\mathcal{G}_0$ a respectiva medida basal.

Por outro lado, [Korwar e Hollander \[1973\]](#) mostram que assintoticamente k^* , o número esperado de valores distintos de K_i 's em uma amostra aleatória de tamanho n de um P.D. é dado por $k^* = \alpha \log \frac{n}{\alpha}$, que é muito menor que n conforme aumenta o tamanho da amostra. Isto tem a ver com o fato de que um P.D. é uma medida aleatória discreta q.c. ([Ghosh e Ramamoorthi \[2003\]](#)).

Finalmente, [Sethuraman \[1994\]](#) fornece uma terceira construção alternativa para os P.D.'s. Sejam $\nu_1, \nu_2, \dots$ v.a.'s i.i.d. $\text{Beta}(1, \alpha)$ e $K_1, K_2, \dots$ v.a.'s i.i.d. $\mathcal{G}_0$. Defina $\omega_1 = \nu_1$ e $\omega_i = \nu_i \prod_{j=1}^{i-1} (1 - \nu_j)$, $j = 2, 3, \dots$. Então,

$$\sum_{i=1}^{\infty} \omega_i \delta_{K_i}^* \sim PD(\alpha\mathcal{G}_0). \quad (2.45)$$

A discretização dos P.D.'s fica evidente na expressão (2.45), pois a sucessão dos ω_i 's é decrescente. Note também que como caso particular da expressão anterior, se $\mathcal{G} \sim PD(\alpha\mathcal{G}_0)$, $K \sim \mathcal{G}_0$ e $\omega \sim \text{Beta}(1, \alpha)$, todas elas independentes, então

$$\omega \delta_K^* + (1 - \omega) \mathcal{G} \sim PD(\alpha\mathcal{G}_0).$$

[Ferguson \[1973\]](#) mostra que para $\mathcal{G}$ P.D. sob $(\mathcal{Q}, \mathcal{K})$ com parâmetro $\alpha\mathcal{G}_0$ e $K_1, \dots, K_n$ uma a.a. de tamanho n de $\mathcal{G}$, então

$$\mathcal{G} \mid K_1, \dots, K_n, \alpha, \mathcal{G}_0 \sim PD \left(\alpha\mathcal{G}_0 + \sum_{i=1}^n \delta_{K_i}^* \right). \quad (2.46)$$

O resultado em (2.46) implica que os P.D.'s possuem a propriedade de conjugação.

Na prática, a distribuição $\mathcal{G}_0$ pode estar indexada por um conjunto de parâmetros, digamos $\boldsymbol{\mu}$. Há duas alternativas: Assumir que $\boldsymbol{\mu}$ é conhecido, ou mais interessante ainda, assumir uma certa distribuição *a priori* $\pi(\boldsymbol{\mu})$ e introduzir $\boldsymbol{\mu}$ no processo de estimação. Porém, para não carregar ainda mais a notação, será omitido este fato nesta seção introdutória.

Por outro lado, a escolha do parâmetro α pode ser muito complexa. Como mencionado anteriormente, valores muito altos de α resultam que o processo seja essencialmente $\mathcal{G}_0$, entanto que valores baixos produzem uma distribuição degenerada em um só valor. Porém, também é possível introduzir α no procedimento de estimação, incluindo uma priori $\pi(\alpha)$ para ele. [Antoniak \[1974\]](#) prova que a distribuição condicional de k^* , o número de valores distintos de K_i 's em uma amostra aleatória de tamanho n de um P.D., é dada por

$$P(k^* | \alpha) = c_n(k^*) n! \alpha^{k^*} \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)}, \quad k^* = 1, \dots, n,$$

em que $c_n(k^*) = P(k^* | \alpha = 1)$, que não depende de α . Dessa forma, utilizando uma priori $\text{Gama}(\nu_1, \nu_2)$ para α , então a distribuição *a posteriori* de α é dada por

$$\begin{aligned} \pi(\alpha | k^*) &\propto \pi(\alpha) \pi(k^* | \alpha), \\ &\propto \alpha^{\nu_1-1} \exp\{-\alpha\nu_2\} \alpha^{k^*} \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)}. \end{aligned}$$

Note que por propriedades da função gama, $\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} = \frac{(\alpha+n)\beta(\alpha+1,n)}{\alpha\Gamma(n)}$, em que $\beta(\cdot, \cdot)$ é a função beta usual. Dessa forma,

$$\begin{aligned} \pi(\alpha | k^*) &\propto \alpha^{\nu_1-1} \exp\{-\alpha\nu_2\} \alpha^{k^*-1} (\alpha + n) \beta(\alpha + 1, n), \\ &\propto \alpha^{\nu_1-1} \exp\{-\alpha\nu_2\} \alpha^{k^*-1} (\alpha + n) \int_0^1 s^\alpha (1-s)^{n-1} ds. \end{aligned} \quad (2.47)$$

[Escobar e West \[1995\]](#) propõem introduzir duas variáveis latentes para auxiliar no passo de amostragem de α : φ cujo domínio é o intervalo $(0, 1)$ e v que é uma variável binária, assumindo valores no conjunto $\{0, 1\}$. Desta forma,

$$\pi(\alpha, \varphi, v | k^*) \propto \alpha^{\nu_1-1} \exp\{-\alpha\nu_2\} \alpha^{k^*-1} \varphi^\alpha (1-\varphi)^{n-1} \left(\frac{n}{\alpha}\right)^v. \quad (2.48)$$

Note que, como requerido, quando integrado e somado (2.48) sobre os respectivos domínios de φ e v , recupera-se a equação (2.47). Desse jeito, é fácil verificar que as distribuições condicionais de α , φ e v são dadas por

$$\begin{aligned} \alpha | \varphi, v, k^* &\sim \text{Gama}(\nu_1 + k^* - v, \nu_2 - \log \varphi), \\ \varphi | \alpha, v, k^* &\sim \text{Be}(\alpha + 1, n) \quad \text{e} \\ v | \alpha, \varphi, k^* &\sim \text{Bern}\left(\frac{n/\alpha}{1 + n/\alpha}\right), \end{aligned} \quad (2.49)$$

em que $\text{Be}(\cdot, \cdot)$ é a distribuição Beta e $\text{Bern}(\cdot)$ é a distribuição Bernoulli. Note que (2.49) permite amostrar facilmente valores de α .

2.4.1 Métodos de amostragem para processos de Dirichlet

Uma limitante dos P.D.'s é o alto custo computacional que eles demandam. Por esta razão, é importante utilizar algoritmos que permitam simular valores destes processos de uma forma eficiente. A seguir, faz-se um resumo dos principais métodos existentes para simular dos P.D.'s, que podem ser agrupados em métodos para distribuições *a priori* conjugadas e não conjugadas (em relação ao

modelo usado) e que podem ser vistos com mais detalhes em Neal [2000].

Para contextualizar o problema, suponha que

$$\begin{aligned} T_i | K_i &\stackrel{i.i.d.}{\sim} \mathcal{F}_T(K_i), \quad i = 1, \dots, n, \\ K_i &\stackrel{i.i.d.}{\sim} \text{PD}(\alpha \mathcal{G}_0), \quad i = 1, \dots, n. \end{aligned}$$

em que $\mathcal{F}_T(K)$ tem forma conhecida. Eventualmente, $\mathcal{F}_T(K)$ e $\mathcal{G}_0$ podem depender de outros conjuntos de parâmetros, mas por comodidade serão omitidos nesta seção. Na literatura, diz-se que a distribuição de $T_i, i = 1, \dots, n$ (marginalizando em relação a K_i) é uma mistura de P.D.'s.

- **Amostrador de Gibbs para P.D.'s quando são usadas distribuições *a priori* conjugadas.**

O fato de trabalhar com um modelo em que é possível obter a conjugação do modelo em relação à medida basal $\mathcal{G}_0$ ajuda bastante na simulação dos P.D.'s, pois permite utilizar algoritmos mais eficientes e sobretudo, mais rápidos.

Note que se for combinada a função de verossimilhança com o resultado em (2.46), tem-se que a distribuição condicional das variáveis $K_1, \dots, K_n$ é dada por

$$K_i | \mathbf{K}^{(i)}, T_i \sim b \sum_{j \neq i} \mathcal{F}(T_i, K_j) \delta_{K_j}^* + b \mathcal{F}(T_i, K_i) \alpha \mathcal{G}_0, \quad i = 1, \dots, n, \quad (2.50)$$

em que $\mathcal{F}(T_i, K_j)$ representa a função de densidade da variável aleatória T dado K_j e avaliada em T_i , $b^{-1} = \sum_{j \neq i} \mathcal{F}(T_i, K_j) + \mathcal{Y}^i$ é a constante que faz com que as probabilidades anteriores somem 1, $\mathcal{Y}^i = \alpha \int \mathcal{F}(T_i, K) \mathcal{G}_0 dK$ e $\mathbf{K}^{(i)}$ representa o vetor $(K_1, \dots, K_n)$ sem a i -ésima observação. Note que se a medida basal $\mathcal{G}_0$ for conjugada em relação a $\mathcal{F}_T(K)$, então a integral em $\mathcal{Y}^i$ pode ser calculada analiticamente. Dessa forma, pode-se amostrar do processo como segue:

- Suponha que o estado atual da cadeia de Markov consiste de $\mathbf{K} = (K_1, \dots, K_n)$. Para $i = 1, \dots, n$, simule um novo valor para K_i a partir da distribuição de $K_i | \mathbf{K}^{(i)}, T_i$ definida em (2.50).

Este é o algoritmo proposto em Blackwell e MacQueen [1973], que é muito simples de implementar. Porém, a convergência da distribuição *a posteriori* pode ser lenta, devido a que frequentemente há grupos associados a um mesmo K_i com uma alta probabilidade. Como o algoritmo não permite mudar mais de um dos K_i 's simultaneamente, então novos valores para os K_i 's acontecem só ocasionalmente.

Para resolver este problema, note que o modelo $T_i | K_i \stackrel{i.i.d.}{\sim} \mathcal{F}_T(K_i)$ pode ser obtido considerando $R \rightarrow \infty$ no seguinte modelo:

$$\begin{aligned} T_i | c_i, \boldsymbol{\psi} &\sim \mathcal{F}(\psi_{c_i}), \\ c_i | \mathbf{q} &\sim \text{Discreta}(q_1, \dots, q_R), \\ \psi_c &\sim \mathcal{G}_0, \\ \mathbf{q} &\sim \text{Dirichlet}(\alpha/R, \dots, \alpha/R), \end{aligned}$$

em que $\mathbf{q} = (q_1, \dots, q_R)$, $\sum_{i=1}^R q_i = 1$ e $c \sim \text{Discreta}(q_1, \dots, q_R)$ indica que $c = c_i$ com probabilidade q_i . c_i indica qual “classe latente” está associada com a observação T_i , em que a etiqueta para c_i não é relevante (i.e., pode-se usar valores para os c_i 's que facilitem a implementação computacional). Para cada classe c , os parâmetros ψ_c determinam a distribuição de

observações da classe e o conjunto de tais ψ_c é denotado por $\boldsymbol{\psi}$. Integrando sobre $\mathbf{q}$, Neal [2000] mostra que

$$P(c_i = c \mid c_1, \dots, c_{i-1}) = \frac{n_{i,c} + \alpha/R}{i - 1 + \alpha},$$

em que $n_{i,c}$ é o número de c_j para $j < i$ que são iguais a c . Assim, tem-se que

$$P(c_i = c \mid c^{(i)}, T_i, \boldsymbol{\psi}) = bF(T_i, \psi_c) \frac{n_{-i,c} + \alpha/R}{i - 1 + \alpha}.$$

$n_{-i,c}$ é o número de c_j para $i \neq j$ que são iguais a c , $c^{(i)}$ é o vetor $(c_1, \dots, c_n)$ sem a i -ésima componente e b é a constante normalizadora. Note que quando um novo valor é escolhido para ψ_c , mudam simultaneamente os valores de K para todas as observações associadas à componente c .

Porém, quando $R \rightarrow \infty$, não é possível representar explicitamente os infinitos valores de ψ_c . Em vez disso, pode-se realizar o amostrador de Gibbs só para aqueles valores que estão atualmente associados a alguma observação, pois

$$\begin{aligned} \text{Se } c = c_j \text{ para algum } j \neq i : P(c_i = c \mid c^{(i)}, T_i, \boldsymbol{\psi}) &= bF(T_i, \psi_c) \frac{n_{-i,c}}{n - 1 + \alpha}, \quad \text{e} \\ P(c_i \neq c_j, \forall j \neq i \mid c^{(i)}, T_i, \boldsymbol{\psi}) &= b \frac{\alpha}{n - 1 + \alpha} \int F(T_i, \psi) d\mathcal{G}_0(\psi). \end{aligned} \quad (2.51)$$

Quando o algoritmo escolhe para c_i um valor diferente de outro c_j , um valor para ψ_{c_i} é simulado da distribuição *a posteriori* baseado na medida basal *a priori* $\mathcal{G}_0$ e na única observação T_i . Desta forma, algoritmo para amostrar valores do processo pode ser resumido como segue:

- Seja o estado atual da cadeia de Markov composta por $\mathbf{c} = (c_1, \dots, c_n)$ e $\boldsymbol{\psi} = \{\psi_c : c \in \{c_1, \dots, c_n\}\}$. Simular um novo valor repetidamente da seguinte forma:
 - * Para $i = 1, \dots, n$: Se $n_{-i,c} = 0$, eliminar ψ_{c_i} do vetor $\boldsymbol{\psi}$ e simular um novo valor para c_i da distribuição $c_i = c \mid c^{(i)}, T_i, \boldsymbol{\psi}$ dada em (2.51). Se o novo valor de c_i não está associado a nenhuma outra observação, simule o valor de ψ_{c_i} da distribuição *a posteriori* baseado na medida basal *a priori* $\mathcal{G}_0$ e na única observação T_i .
 - * Para todo $c \in \{c_1, \dots, c_n\}$: Simular um novo valor de $\psi_c \mid$ todo y_i para os quais $c_i = c$, ou seja, da distribuição *a posteriori* baseada na priori $\mathcal{G}_0$ e em todas as observações associadas à classe latente c .

Este é o algoritmo discutido em Bush e MacEachern [1996].

Existe mais um algoritmo em que é usada a mesma ideia anterior, mas com ψ_c sendo eliminado por integração. Detalhes sobre este método podem ser encontrados em MacEachern [1994].

- **Amostrador de Gibbs para P.D.'s quando não são usadas distribuições *a priori* conjugadas.**

Os algoritmos apresentados anteriormente não podem ser aplicados quando $\mathcal{G}_0$ não é conjugado em relação a distribuição $\mathcal{F}_T(K_i)$, para $i = 1, \dots, n$, basicamente porque o fato de obter esta conjugação permite calcular analiticamente as integrais envolvidas, por exemplo, aquela em (2.50).

West et al. [1994] sugerem aproximar essas integrais utilizando métodos numéricos como a quadratura de Gauss ou métodos de Monte Carlo. Desta forma, também pode-se utilizar um método aproximado de simulação para a densidade *a posteriori* $\mathcal{F}(T_i, K_i)\mathcal{G}_0$, considerando um conjunto de valores para K_i , com probabilidades proporcionais a sua densidade.

Porém, este método é bastante inexato, sobretudo quando os K_i 's são multidimensionais. Se um número pequeno de valores para $K_i, i = 1, \dots, n$, for utilizado para essa aproximação, então a probabilidade de escolher um nova componente para c_i pode ser muito menor que se as probabilidades reais em (2.51) fossem usadas. Em consequência, o algoritmo além de ser ineficiente, pode produzir uma distribuição estacionária errada para ψ , e portanto, para todos os demais parâmetros do modelo.

O algoritmo “no gaps” proposto por MacEachern e Muller [1998] surge quando os c_i 's, para $i = 1, \dots, n$ são necessários para cobrir o conjunto dos inteiros de 1 a k^* , em que k^* é o número de diferentes c_i 's, mas não há outra restrição. Considerando Gibbs sampling nesta representação, eles deduziram o seguinte algoritmo.

- Seja o estado atual da cadeia de Markov que consiste em $\mathbf{c} = (c_1, \dots, c_n)$ e $\psi = \{\psi_c : c \in \{c_1, \dots, c_n\}\}$. Amostrar sucessivamente segundo os seguintes passos.
 - * Para $i = 1, \dots, n$: Defina k^- como o número de diferentes c_j 's para $j \neq i$, e assuma que estes c_j 's tem valores em $\{1, \dots, k^-\}$. Se $c_i \neq c_j$ para todo $j \neq i$, então com probabilidade $k^-/(k^- + 1)$ deixar c_i sem mudanças. Caso contrário, a etiqueta c_i é definida como $k^- + 1$ se $c_i \neq c_j$ para todo $j \neq i$, ou simule um valor para ψ_{k^-+1} de $\mathcal{G}_0$ se $c_i = c_j$ para algum $j \neq i$. Logo, simule um novo valor para c_i do conjunto $\{1, \dots, k^- + 1\}$ usando as seguintes probabilidades:

$$P(c_i = c \mid c^{(i)}, T_i, \psi_1, \dots, \psi_{k^-+1}) = \begin{cases} bn_{-i,c}F(T_i, \psi_c) & , \text{ se } 1 \leq c \leq k^-, \\ b[\alpha/(k^- + 1)]F(T_i, \psi_c) & , \text{ se } c = k^- + 1, \end{cases}$$

em que b é a constante normalizadora.

Note que este algoritmo pode ser aplicado em qualquer modelo no qual possa ser amostrado de $\mathcal{G}_0$ e calculado $F(T, K)$, independente de se $\mathcal{G}_0$ é conjugado para $\mathcal{F}_T(K)$ ou não. Porém, há uma ineficiência no mecanismo do algoritmo para mudar c_i a um valor diferentes de todos os outros c_j 's, pois a probabilidade de tal mudança é menor do que pode-se esperar por conta do factor $k^- + 1$.

Para melhorar este problema, Neal [2000] propõe uma pequena modificação no algoritmo anterior. A ideia é que se $c_i \neq c_j$ para todo $i \neq j$, então ele deve estar associado com uns dos ℓ parâmetros auxiliares. Dessa forma, verifica-se que

$$P(c_i = c \mid c^{(i)}, T_i, \psi_1, \dots, \psi_{k^-+1}) = \begin{cases} b \frac{n_{-i,c}}{n-1+\alpha} \mathcal{F}(T_i, \psi_c) & , \text{ se } 1 \leq c \leq k^-, \\ b \frac{\alpha/\ell}{n-1+\alpha} \mathcal{F}(T_i, \psi_c) & , \text{ se } c = k^- + 1, \dots, k^- + \ell, \end{cases}$$

em que b é a constante normalizadora. Note que $\ell = 1$, este algoritmo corresponde ao algoritmo de “no gaps”.

Adicionalmente, Neal [2000] apresenta mais dois algoritmos baseados no método de Metropolis-Hastings (Hastings [1970]). Estes algoritmos são muito fáceis de implementar em termos computacionais, mas também não podem competir com este último algoritmo apresentado.

2.4.2 Processos de Dirichlet centrados

O fato de utilizar P.D.'s no contexto de efeitos aleatórios pode ocasionar alguns problemas, pois assumindo $\mathbf{K} \sim \text{PD}(\alpha \mathcal{G}_0)$, não garante que a priori $E(\mathbf{K}) = \boldsymbol{\mu}_{\mathcal{G}} = \int \mathbf{K} d\mathcal{G}_0(\mathbf{K}) = 0$, mesmo que a medida basal estiver centrada em zero (Por exemplo, $\mathcal{G}_0 = N(0, 1)$).

Uma possível solução, é dada em [Li et al. \[2011\]](#). Suponha o contexto em que o estudo envolve J grupos ou clusters com n_j observações cada. Suponha ainda que o preditor linear associado ao i -ésimo indivíduo do j -ésimo cluster é definido por $\eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{G}_{ij}^\top \mathbf{K}$, em que $\mathbf{x}_{ij}$ e $\mathbf{G}_{ij}$ representam as matrizes de desenho associadas à parte fixa e aleatória do modelo. Sem perda de generalidade, suponha as partições $\mathbf{x}_{ij} = (\mathbf{x}_{ij}^F, \mathbf{x}_{ij}^R)$ e $\boldsymbol{\beta} = (\boldsymbol{\beta}^F, \boldsymbol{\beta}^R)$.

Se diz que os efeitos fixos $\boldsymbol{\beta}^R$ estão pareados com os efeitos aleatórios $\mathbf{K}$ se o subvetor $\mathbf{x}_{ij}^R$ é idêntico a $\mathbf{G}_{ij}$ para $i = 1, \dots, n_j$, por exemplo, ambos contem um intercepto. Dessa forma, ao invés de fazer inferência diretamente para $\boldsymbol{\beta}^R$, é mais apropriado fazer inferência para $\boldsymbol{\beta}_{par} \equiv \boldsymbol{\beta}^R + \boldsymbol{\mu}_{\mathcal{G}}$. Seguindo esse argumento, [Li et al. \[2011\]](#) propõem modelar a distribuição de $\boldsymbol{\beta}^R + \mathbf{K}_j$ como

$$\mathbf{K}_j^* = \boldsymbol{\beta}^R + \mathbf{K}_j \stackrel{i.i.d.}{\sim} \mathcal{G}, \quad \mathcal{G} \sim \text{PD}(\alpha \mathcal{G}_0), \quad \mathcal{G}_0 = N(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}), \quad (2.52)$$

em que $\boldsymbol{\beta}_k$ e $\boldsymbol{\Sigma}$ são um vetor e uma matriz respectivamente, que podem ser assumidos conhecidos ou melhor ainda, assumir distribuições *a priori* para eles e entrar no processo de estimação. O esquema proposto em (2.52) é conhecido como Processo Dirichlet Centrado (P.D.C.).

Defina $\mathbf{K}^* = (\mathbf{K}_1^*, \dots, \mathbf{K}_J^*)$ e

$$\mathcal{G}_\star = \frac{\sum_{j=1}^J \delta^* \mathbf{K}_j^* + \alpha N(\boldsymbol{\beta}_k, \boldsymbol{\Sigma})}{J + \alpha}.$$

A esperança e variância de $\mathcal{G}_\star$ são dados respectivamente por $\mu_{\mathcal{G}_\star} = (\alpha \boldsymbol{\beta}_k + \sum_{j=1}^J \mathbf{K}_j^*) / (J + \alpha)$ e $\text{Var}_{\mathcal{G}_\star} = \{\alpha(\boldsymbol{\beta}_k \boldsymbol{\beta}_k^\top + \boldsymbol{\Sigma}) + \sum_{j=1}^J \mathbf{K}_j^* \mathbf{K}_j^{*\top}\} / (J + \alpha) - \mu_{\mathcal{G}_\star} \mu_{\mathcal{G}_\star}^\top$.

Um aspecto computacional muito importante na utilização dos P.D.C.'s é o fato de não aumentar o custo computacional, pois precisa-se gerar exatamente o mesmo número de cadeias que no modelo original. O principal problema desta abordagem é que já não é possível realizar inferência sobre o vetor $\boldsymbol{\beta}^R$. Porém, neste trabalho o vetor $\boldsymbol{\beta}^R$ corresponderá aos interceptos dos preditores lineares no modelo. Dessa forma, a inferência sobre $\boldsymbol{\beta}^R$ não é muito relevante. Em geral, nos modelos mistos é mais relevante a inferência sobre $\boldsymbol{\beta}^R + \boldsymbol{\mu}_{\mathcal{G}}$ ou uma função desse termo.

Finalmente, em relação aos critérios de seleção de modelos para os processos Dirichlet, não é possível utilizar os critérios de seleção de modelos usuais (como AIC e BIC ou seus análogos condicionais) dado que a verossimilhança do modelo não pode ser escrita em forma fechada. Porém, pode-se utilizar o Fator de Bayes para comparar modelos dois a dois. [Basu e Chib \[2003\]](#) explicam em forma muito detalhada como estimar a verossimilhança marginal em modelos que envolvem processos Dirichlet.

Capítulo 3

Modelos de sobrevivência com fração de cura e efeitos aleatórios

Nos estudos nos quais os dados dos pacientes foram coletados em diferentes clínicas pode acontecer que as respostas dos indivíduos em uma mesma clínica (*cluster*) estejam correlacionados. Mais ainda, um dos interesses do estudo pode ser avaliar a diferença devido ao tratamento entre as diferentes clínicas. Este capítulo visa fazer uma revisão dos modelos com fração de cura vistos na seção 2.2 introduzindo uma componente aleatória para cada clínica, que representa o efeito de cada clínica sobre o paciente. Uma vez feita a revisão, propomos algumas extensões nesta mesma linha.

3.1 Modelo de mistura com efeitos aleatórios

O primeiro modelo que aparece no contexto de fração de cura com indivíduos que pertencem a diferentes clínicas (*clusters*) foi proposto em Kalbfleisch e Prentice [2002], o qual é uma aplicação do modelo em (2.6) incluindo o efeito das clínicas como efeitos fixos. Posteriormente, Yau e Ng [2001] apresentam uma abordagem de efeitos aleatórios deste mesmo modelo.

Suponha que há J clínicas e que T_{ij} representa o tempo observado (falha/censura) para o i -ésimo indivíduo da j -ésima clínica. Seja ainda n_j o número de observações na j -ésima clínica. No modelo de mistura com efeitos aleatórios, a função de sobrevivência para o i -ésimo indivíduo da j -ésima clínica é dada por

$$S_{pop}(t \mid \eta_{ij}, \pi_{ij}, \boldsymbol{\lambda}) = 1 - \pi_{ij} + \pi_{ij} \cdot S(t \mid \eta_{ij}, \boldsymbol{\lambda}).$$

Neste caso, assume-se que a chance de cura de cada indivíduo depende tanto de um conjunto de covariáveis $\mathbf{z}_{ij}$, quanto da clínica à qual pertence. O efeito aleatório da j -ésima clínica sobre a chance de cura dos pacientes tratados nela será denotado por V_j . Dessa forma, como em (2.10), Yau e Ng [2001] utilizam a ligação logito para modelar a probabilidade de cura fazendo

$$\pi_{ij} = \frac{1}{1 + \exp\{\vartheta_{ij}\}}, \quad (3.1)$$

e incluindo o vetor $\mathbf{z}_{ij}$ e o efeito aleatório V_j através de

$$\vartheta_{ij} = \mathbf{z}_{ij}^\top \boldsymbol{\beta} + V_j, \quad i = 1, \dots, n_j, j = 1, \dots, J,$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)^\top$ é vetor de parâmetros.

Além disso, acredita-se que a sobrevivência dos indivíduos suscetíveis depende tanto de um conjunto de covariáveis $\mathbf{x}_{ij}$, quanto da clínica na qual foi tratado. O efeito aleatório da j -ésima clínica sobre a sobrevivência dos indivíduos suscetíveis será denotado por U_j , e que determina o risco de formação de tumor para os indivíduos suscetíveis (termo também conhecido na literatura como fragilidade). Assim, como em (2.8), Yau e Ng [2001] fazem

$$S(t \mid \boldsymbol{\lambda}, \eta_{ij}) = [S_0(t \mid \boldsymbol{\lambda})]^{\exp(\eta_{ij})}, \quad (3.2)$$

em que $\eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\gamma} + U_j$ para $i = 1, \dots, n_j, j = 1, \dots, J, \boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_r)^\top$ é vetor de parâmetros e $S_0(t | \boldsymbol{\lambda})$ define a função de sobrevivência basal para os indivíduos suscetíveis e que deve pertencer à classe de modelos de riscos proporcionais.

Valores altos (baixos) de U_j indicam que os pacientes suscetíveis tratados na j -ésima clínica possuem um risco de morte maior (menor). Por outro lado, valores altos (baixos) de V_j indicam que os pacientes tratados na j -ésima clínica têm uma chance de cura menor (maior).

Yau e Ng [2001] assumem que os vetores $(U_j, V_j)^\top$ são i.i.d. tais que

$$\begin{pmatrix} U_j \\ V_j \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right), \quad j = 1, \dots, J.$$

Defina $\boldsymbol{\Sigma} = \text{diag}(\sigma_u^2, \sigma_v^2)$. Dessa forma, definindo $\mathbf{U}^\top = (U_1, \dots, U_J)$ e $\mathbf{V}^\top = (V_1, \dots, V_J)$, tem-se que o vetor $\mathbf{K} = (\mathbf{U}^\top, \mathbf{V}^\top)^\top \sim N(\mathbf{0}, \boldsymbol{\Omega})$, em que

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_u^2 \mathbf{I}_J & 0 \\ 0 & \sigma_v^2 \mathbf{I}_J \end{pmatrix},$$

e $\mathbf{I}_J$ representa a matriz identidade de dimensão J .

Nesta abordagem, os dados observados são $\mathbf{D}_{obs} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{Z})$, em que $\mathbf{t} = (t_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$, $\boldsymbol{\delta} = (\delta_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$ e $\mathbf{X}$ e $\mathbf{Z}$ são as matrizes de covariáveis para a modelagem da função de sobrevivência dos indivíduos suscetíveis e a fração de cura, respectivamente. Os dados completos são $\mathbf{D}_c = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{K}, \mathbf{X}, \mathbf{Z})$ que incluem os dados observados e o vetor de efeitos aleatórios $\mathbf{K}$. Yau e Ng [2001] utilizam uma abordagem baseada em estimadores REML [McGilchrist e Yau [1995]], a qual requer fixar $\boldsymbol{\Sigma}$ em um primeiro passo e obter o estimador BLUP de $\boldsymbol{\zeta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{K})$. Para $\boldsymbol{\Sigma}$ fixo, a função de log-verossimilhança BLUP do modelo é dada por $l = l_1 + l_2$, em que

$$\begin{aligned} l_1 &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \delta_{ij} \log f_{pop}(t_{ij} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}_{ij}, \mathbf{z}_{ij}) + (1 - \delta_{ij}) \log S_{pop}(t_{ij} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}_{ij}, \mathbf{z}_{ij}) \right\}, \\ &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \delta_{ij} \log [\pi_{ij} f(t_{ij} | \eta_{ij}, \boldsymbol{\lambda})] + (1 - \delta_{ij}) \log [1 - \pi_{ij} + \pi_{ij} S(t_{ij} | \eta_{ij}, \boldsymbol{\lambda})] \right\}, \quad \text{e} \quad (3.3) \\ l_2 &= (-1/2)[J \log(2\pi\sigma_u^2) + (1/\sigma_u^2)\mathbf{U}^\top \mathbf{U}] + (-1/2)[J \log(2\pi\sigma_v^2) + (1/\sigma_v^2)\mathbf{V}^\top \mathbf{V}], \end{aligned}$$

com $S(t | \eta_{ij}, \boldsymbol{\lambda})$ definida em (3.2), $f(t | \eta_{ij}, \boldsymbol{\lambda}) = -(\partial/\partial t)S(t | \eta_{ij}, \boldsymbol{\lambda})$ e π_{ij} como em (3.1). Para $S_0(t | \boldsymbol{\lambda})$, Yau e Ng [2001] consideraram a função de sobrevivência da distribuição Weibull.

Como mencionado em Klein et al. [2013], para não causar problemas de identificabilidade, o procedimento de estimação precisa fixar $S_0(t | \boldsymbol{\lambda})$ para estimar as demais componentes do modelo e vice-versa. Portanto, para $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ fixos, o estimador BLUP de $\boldsymbol{\zeta}$, digamos $\hat{\boldsymbol{\zeta}}$, pode ser encontrado maximizando l em relação a $\boldsymbol{\zeta}$. Defina $\boldsymbol{\zeta}_0$ como um valor inicial para $\hat{\boldsymbol{\zeta}}$, utilizando o algoritmo de Newton-Raphson iterativamente pode-se encontrar $\hat{\boldsymbol{\zeta}}$ da seguinte expressão

$$\boldsymbol{\zeta}_\ell = \boldsymbol{\zeta}_{\ell-1} + \mathbf{B}_{\ell-1}^{-1} \left(\frac{\partial l}{\partial \boldsymbol{\zeta}} \right) \bigg|_{\boldsymbol{\zeta}=\boldsymbol{\zeta}_{\ell-1}}, \quad \ell = 1, 2, \dots, \quad (3.4)$$

em que $\mathbf{B}$ é menos a segunda derivada de l em relação a $\boldsymbol{\zeta}$. Denote a matriz particionada $\mathbf{B}$ e sua inversa, avaliada em $\hat{\boldsymbol{\zeta}}$ de acordo à ordenação $\boldsymbol{\gamma} | \boldsymbol{\beta} | \mathbf{U} | \mathbf{V}$, respectivamente como

$$\widehat{\mathbf{B}} = \begin{bmatrix} \widehat{\mathbf{B}}_{11} & \widehat{\mathbf{B}}_{12} & \widehat{\mathbf{B}}_{13} & \widehat{\mathbf{B}}_{14} \\ \widehat{\mathbf{B}}_{21} & \widehat{\mathbf{B}}_{22} & \widehat{\mathbf{B}}_{23} & \widehat{\mathbf{B}}_{24} \\ \widehat{\mathbf{B}}_{31} & \widehat{\mathbf{B}}_{32} & \widehat{\mathbf{B}}_{33} & \widehat{\mathbf{B}}_{34} \\ \widehat{\mathbf{B}}_{41} & \widehat{\mathbf{B}}_{42} & \widehat{\mathbf{B}}_{43} & \widehat{\mathbf{B}}_{44} \end{bmatrix} \quad \text{e} \quad \widehat{\mathbf{B}}^{-1} = \begin{bmatrix} \widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12} & \widehat{\mathbf{A}}_{13} & \widehat{\mathbf{A}}_{14} \\ \widehat{\mathbf{A}}_{21} & \widehat{\mathbf{A}}_{22} & \widehat{\mathbf{A}}_{23} & \widehat{\mathbf{A}}_{24} \\ \widehat{\mathbf{A}}_{31} & \widehat{\mathbf{A}}_{32} & \widehat{\mathbf{A}}_{33} & \widehat{\mathbf{A}}_{34} \\ \widehat{\mathbf{A}}_{41} & \widehat{\mathbf{A}}_{42} & \widehat{\mathbf{A}}_{43} & \widehat{\mathbf{A}}_{44} \end{bmatrix}. \quad (3.5)$$

Note que, se for observado $\delta_{ij} = 1, \forall j = 1, \dots, J, \forall i = 1, \dots, n_j$, então usando (2.39) tem-se que a distribuição assintótica do vetor $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})^\top$ é dada por

$$\begin{pmatrix} \widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12} \\ \widehat{\mathbf{A}}_{21} & \widehat{\mathbf{A}}_{22} \end{pmatrix}^{-1} \left[\begin{pmatrix} \widehat{\boldsymbol{\gamma}} \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\beta} \end{pmatrix} \right] \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}), \quad n_1 + \dots + n_J \rightarrow \infty. \quad (3.6)$$

Na presença de observações censuradas, não pode ser garantido analiticamente o resultado anterior, mas será visto nos estudos de simulação que mesmo assim o resultado parece ser válido.

Para estimar as componentes de variância do modelo $(\boldsymbol{\Sigma})$ precisa-se resolver as equações de estimação dadas em (2.37), que neste caso assumem as formas

$$\text{tr} \left(\boldsymbol{\Omega}^{-1} \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_u^2} \right) + \text{tr}(\widehat{\mathbf{A}}_{33} + \widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top) \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2} = 0 \quad \text{e} \quad \text{tr} \left(\boldsymbol{\Omega}^{-1} \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_v^2} \right) + \text{tr}(\widehat{\mathbf{A}}_{44} + \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top) = 0. \quad (3.7)$$

Como mencionado em Yau e Ng [2001], a solução de (3.7) é dada por

$$\widehat{\sigma}_u^2 = (1/J) \text{tr}(\widehat{\mathbf{A}}_{33} + \widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top) \quad \text{e} \quad \widehat{\sigma}_v^2 = (1/J) \text{tr}(\widehat{\mathbf{A}}_{44} + \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top). \quad (3.8)$$

Note que pelo resultado em (2.41), a matriz de variâncias-covariâncias do vetor $(\widehat{\sigma}_u^2, \widehat{\sigma}_v^2)^\top$ é dada por $(2\mathbf{W})^{-1}$, em que os elementos da matriz $\mathbf{W}$ são dados por

$$\begin{aligned} W_{11} &= \sigma_u^{-4} (J - 2\sigma_u^{-2} \text{tr} \widehat{\mathbf{A}}_{33}) + \sigma_u^{-8} \text{tr} \widehat{\mathbf{A}}_{33}^2, \\ W_{22} &= \sigma_v^{-4} (J - 2\sigma_v^{-2} \text{tr} \widehat{\mathbf{A}}_{44}) + \sigma_v^{-8} \text{tr} \widehat{\mathbf{A}}_{44}^2, \quad \text{e} \\ W_{12} &= \sigma_u^{-4} \sigma_v^{-4} \text{tr}(\widehat{\mathbf{A}}_{34} \widehat{\mathbf{A}}_{43}). \end{aligned} \quad (3.9)$$

Até agora, o vetor $\boldsymbol{\lambda}$ foi assumido constante, mas ele também deve ser estimado. Com os valores de $\widehat{\boldsymbol{\zeta}}$ e $\widehat{\boldsymbol{\Sigma}}$ obtidos das expressões (3.4) e (3.8), pode-se construir uma versão perfilada l_p da função de verossimilhança l em (3.3) e encontrar o valor de $\boldsymbol{\lambda}$ que maximize essa expressão. Se $\boldsymbol{\lambda}_0$ é um valor inicial para esse valor, então pode-se encontrar $\widehat{\boldsymbol{\lambda}}$, o estimador de $\boldsymbol{\lambda}$, também usando o algoritmo de Newton-Raphson iterativamente na expressão

$$\boldsymbol{\lambda}_\ell = \boldsymbol{\lambda}_{\ell-1} + \mathbf{C}_{\ell-1}^{-1} \left(\frac{\partial l_p}{\partial \boldsymbol{\lambda}} \right) \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_{\ell-1}}, \quad \ell = 1, 2, \dots, \quad (3.10)$$

em que $\mathbf{C}$ é o negativo da segunda derivada de l_p em relação a $\boldsymbol{\lambda}$. Como apontado em Yau e Ng [2001], a variância assintótica do estimador obtido em (3.10) não é dada pela inversa de $\mathbf{C}$, pois o valor esperado das derivadas parciais em relação a $\boldsymbol{\lambda}$, considerando os outros parâmetros como fixos, não é zero. Porém, um método de reamostragem como o Jackknife (Miller [1974]) pode ser considerado para calcular os desvios padrão para $\widehat{\boldsymbol{\lambda}}$.

Em relação aos valores iniciais, pode-se considerar como estimativas iniciais para $\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\lambda}}$ as estimativas produzidas pelo modelo de mistura sem efeitos aleatórios apresentado em (2.12). Entretanto para os efeitos aleatórios, podemos considerar como valor inicial a média da distribuição imposta, i.e., $\widehat{\mathbf{U}}_0 = \mathbf{0}$ e $\widehat{\mathbf{V}}_0 = \mathbf{0}$. Finalmente, para a matriz das componentes da variância, precisa-se iniciar o algoritmo com algum valor escolhido ao acaso. Neste caso, pode-se testar vários valores iniciais para $\boldsymbol{\Sigma}$ para avaliar se há influência ou não nas estimativas obtidas.

Outra alternativa de valores iniciais que pode ser considerada, é o modelo de mistura com efeitos fixos (Kalbfleisch e Prentice [2002]). Dessa forma, pode-se obter um valor inicial para $\boldsymbol{\zeta}$ e $\boldsymbol{\lambda}$. Neste caso, um valor inicial para as componentes de variância pode ser obtido pela variância dos efeitos aleatórios estimados a partir do modelo com efeitos fixos.

Desta forma, considerando valores iniciais como descrito anteriormente para $\hat{\boldsymbol{\zeta}}$, $\hat{\boldsymbol{\Sigma}}$ e $\hat{\boldsymbol{\lambda}}$, pode-se iterar entre as expressões (3.4), (3.8) e (3.10) até obter a convergência do algoritmo.

Posteriormente, Lai e Yau [2008] propõem uma extensão do modelo de Yau e Ng [2001], modificando alguns aspectos do modelo. O principal deles é o fato de considerar que os e.a.'s U_j e V_j são correlacionados, o que faz sentido no contexto do problema, pois todos os pacientes envolvidos numa mesma clínica deveriam ter algum tipo de correlação, independente de se eles são suscetíveis ou curados. Pontualmente, assumiram que os pares (U_j, V_j) , para $j = 1, \dots, J$ são independentes com distribuição comum $N(0, \boldsymbol{\Sigma})$, em que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}. \quad (3.11)$$

Nesta parametrização, ρ representa o coeficiente de correlação entre U e V . Se $\rho = 0$, então os efeitos aleatórios são independentes. A interpretação dos efeitos aleatórios é a mesma que no caso dos efeitos independentes. Portanto, uma correlação positiva é indicativo de que as clínicas com pacientes suscetíveis com maior (menor) risco também tendem a ter frações de cura menores (maiores). Pelo contrário, uma correlação negativa sugere que as clínicas com pacientes suscetíveis com maior (menor) risco tendem a ter maiores (menores) frações de cura. Em outras palavras, se ρ é positivo, as ações tomadas pelas clínicas poderiam ter o mesmo efeito (positivo ou negativo) em ambos grupos, suscetíveis e curados; entretanto ρ negativo, indica que qualquer ação que a clínica possa tomar e que tiver um efeito positivo nos indivíduos suscetíveis, prejudicaria aos indivíduos curados dessa mesma clínica e viceversa.

Por outro lado, ao invés de trabalhar com a função de log-verossimilhança observada do modelo em (2.12), Lai e Yau [2008] trabalharam com a verossimilhança aumentada do modelo em (2.11), i.e., incluindo no modelo as variáveis latentes M_{ij} que indicam se o i -ésimo indivíduo da j -ésima clínica é suscetível ou não. Esta função de log-verossimilhança pode ser escrita como $l_a = l_1^a + l_2^a$, em que

$$l_1^a = \sum_{j=1}^J \sum_{i=1}^{n_j} \left[M_{ij} \left\{ \delta_{ij} \left[\log \hat{h}_0(t_{ij}) + \eta_{ij} \right] - \exp\{\eta_{ij}\} \hat{S}_0(t_{ij}) + \log \pi_{ij} \right\} \right. \\ \left. + (1 - M_{ij}) \log(1 - \pi_{ij}) \right], \quad \text{e} \quad (3.12)$$

$$l_2^a = -\frac{1}{2(1 - \rho^2)} \left[\frac{\mathbf{U}^\top \mathbf{U}}{\sigma_u^2} + \frac{\mathbf{V}^\top \mathbf{V}}{\sigma_v^2} - \frac{2\rho \mathbf{U}^\top \mathbf{V}}{\sigma_u \sigma_v} \right] - (J/2) \log\{4\pi^2 \sigma_u^2 \sigma_v^2 (1 - \rho^2)\},$$

em que $\hat{h}_0 = \log \hat{S}_0$, η_{ij} provém de (3.2), mas para π_{ij} foi utilizada a parametrização

$$\pi_{ij} = \frac{\exp(\vartheta_{ij})}{1 + \exp(\vartheta_{ij})}, \quad (3.13)$$

em que $\vartheta_{ij} = \mathbf{z}_{ij}^\top \boldsymbol{\beta} + V_j$, de forma similar a expressão em (3.1).

Nesta abordagem, foi utilizado o algoritmo EM (Dempster et al. [1977]), mas no passo da maximização foram usados os estimadores REML (McGilchrist e Yau [1995]). Adicionalmente, ao invés

de trabalhar com uma forma paramétrica para $S_0(t)$, [Lai e Yau \[2008\]](#) adotaram um estimador tipo [Breslow \[1974\]](#) para a função de sobrevivência basal $S_0(\cdot)$ na forma

$$\widehat{S}_0(t) = \exp \left\{ - \sum_{ij:t_{(ij)} \leq t} \frac{y_{ij}}{\sum_{l \in \mathcal{R}_{ij}} M_l \exp \eta_l} \right\}, \quad (3.14)$$

em que y_{ij} é o número de falhas que aconteceram em $t_{(ij)}$ e $\mathcal{R}_{ij}$ é o conjunto dos indivíduos em risco no instante imediatamente anterior a $t_{(ij)}$.

Desta forma, o passo E (da esperança) do algoritmo pode ser computado utilizando o resultado em (2.7) e adaptando-o ao fato de ter-se covariáveis $\mathbf{X}$ e $\mathbf{Z}$ para explicar a sobrevivência dos indivíduos suscetíveis e a chance de cura. Assim, com valores de $\boldsymbol{\zeta}$ e $\boldsymbol{\Sigma}$ fixados no passo anterior, pode-se calcular o valor esperado do vetor $\mathbf{M} = (M_{11}, \dots, M_{Jn_j})$ na $(\ell+1)$ -ésima iteração utilizando

$$E \left(M_{ij}^{(\ell+1)} \mid \boldsymbol{\zeta}^{(\ell)} \right) = \delta_{ij} + (1 - \delta_{ij}) \frac{\pi_{ij}^{(\ell)} \widehat{S}_0(t_{ij})^{\exp \eta_{ij}^{(\ell)}}}{1 - \pi_{ij}^{(\ell)} + \pi_{ij}^{(\ell)} \widehat{S}_0(t_{ij})^{\exp \eta_{ij}^{(\ell)}}}, \quad \ell = 0, 1, 2, \dots, \quad (3.15)$$

em que $\eta_{ij}^{(\ell)}$ pode ser computado com os valores de $\boldsymbol{\gamma}^{(\ell)}$ e $\mathbf{U}^{(\ell)}$ em (3.2), $\pi_{ij}^{(\ell)}$ pode ser calculado com os valores de $\boldsymbol{\beta}^{(\ell)}$ e $\mathbf{V}^{(\ell)}$ em (3.13) e $\widehat{S}_0(t)$ provém de (3.14) usando os valores de $\mathbf{M}^{(\ell)}$ e $\boldsymbol{\eta}^{(\ell)}$.

O passo M do algoritmo (da maximização) foi implementado em conjunto com os estimadores REML. Assim, para $\boldsymbol{\Sigma}^{(\ell)}$ fixo e $\mathbf{M}^{(\ell)}$ computado na expressão anterior, pode-se maximizar a expressão em (3.12) em relação a $\boldsymbol{\zeta}^{(\ell)}$ usando o algoritmo de Newton-Raphson com algum valor inicial como segue:

$$\boldsymbol{\zeta}_l^{(\ell)} = \boldsymbol{\zeta}_{l-1}^{(\ell)} + \left(\mathbf{B}_{l-1}^{(\ell)} \right)^{-1} \left(\frac{\partial l^a}{\partial \boldsymbol{\zeta}} \right) \bigg|_{\boldsymbol{\zeta} = \boldsymbol{\zeta}_{l-1}^{(\ell)}}, \quad l = 1, 2, \dots, \quad (3.16)$$

em que $\mathbf{B}^{(\ell)}$ é o negativo da segunda derivada de l^a em relação a $\boldsymbol{\zeta}$ e avaliada em $\boldsymbol{\zeta}^{(\ell)}$. A matriz particionada $\mathbf{B}$ e sua inversa, avaliada na solução de (3.16), digamos $\widehat{\boldsymbol{\zeta}}^{(\ell)}$, conforme à ordenação $\boldsymbol{\gamma} \mid \boldsymbol{\beta} \mid \mathbf{U} \mid \mathbf{V}$ é a mesma que em (A.2.2). A variância do vetor $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})^\top$ pode ser estimada como em (3.6).

Defina $\widehat{\mathbf{A}}^*$ como a submatriz de $\widehat{\mathbf{B}}^{-1}$ correspondente só aos efeitos aleatórios, i.e.,

$$\widehat{\mathbf{A}}^* = \begin{pmatrix} \widehat{\mathbf{A}}_{33} & \widehat{\mathbf{A}}_{34} \\ \widehat{\mathbf{A}}_{43} & \widehat{\mathbf{A}}_{44} \end{pmatrix}.$$

Assim, as equações de estimação para estimar os elementos da matriz $\boldsymbol{\Sigma}^{(\ell)}$ das componentes de variância, com $\mathbf{M}^{(\ell)}$ e $\boldsymbol{\zeta}^{(\ell)}$ já atualizados, são dadas por

$$\text{tr} \left(\{ \boldsymbol{\Omega}^{(\ell)} \}^{-1} \frac{\partial \boldsymbol{\Omega}^{(\ell)}}{\partial \boldsymbol{\tau}_i^{(\ell)}} \right) + \text{tr} \left(\left(\widehat{\mathbf{A}}^{*(\ell)} + \widehat{\mathbf{K}}^{(\ell)} \{ \widehat{\mathbf{K}}^{(\ell)} \}^\top \right) \frac{\partial \{ \boldsymbol{\Omega}^{(\ell)} \}^{-1}}{\partial \boldsymbol{\tau}_i^{(\ell)}} \right) = 0, \quad i = 1, 2, 3, \quad (3.17)$$

em que $\mathbf{K} = (\mathbf{U}, \mathbf{V})$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes I_J$ e $\boldsymbol{\tau} = (\sigma_u^2, \sigma_v^2, \rho)$. Seguindo [Lai e Yau \[2008\]](#), a solução de (3.17)

é dada por

$$\begin{aligned}\hat{\sigma}_u^{2(\ell)} &= \frac{1}{J} \text{tr} \left(\hat{\mathbf{A}}_{33}^{(\ell)} + \hat{\mathbf{U}}^{(\ell)} \{ \hat{\mathbf{U}}^{(\ell)} \}^\top \right), & \hat{\sigma}_v^{2(\ell)} &= \frac{1}{J} \text{tr} \left(\hat{\mathbf{A}}_{44}^{(\ell)} + \hat{\mathbf{V}}^{(\ell)} \{ \hat{\mathbf{V}}^{(\ell)} \}^\top \right) \quad \text{e} \\ \hat{\rho}^{(\ell)} &= \frac{\text{tr} \left(\hat{\mathbf{A}}_{34}^{(\ell)} + \hat{\mathbf{U}}^{(\ell)} \{ \hat{\mathbf{V}}^{(\ell)} \}^\top \right)}{\sqrt{\hat{\sigma}_u^{2(\ell)} \times \hat{\sigma}_v^{2(\ell)}}}.\end{aligned}\quad (3.18)$$

Note que a solução para $\hat{\sigma}_u^2$ e $\hat{\sigma}_v^2$ é a mesma que em (3.8) e (3.18). Por outro lado, usando o resultado em (2.40) pode-se estimar a variância do vetor $\boldsymbol{\tau}$, a qual é dada por $(2\mathbf{W})^{-1}$, em que os elementos da matriz $\mathbf{W}$ são dados por

$$W_{ij} = \text{tr} \left(\hat{\mathbf{A}}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_i} \hat{\mathbf{A}}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_j} + \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_i} \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_j} - 2 \hat{\mathbf{A}}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_i} \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{\tau}_j} \right), \quad 1 \leq i, j \leq 3. \quad (3.19)$$

Pode ser verificado que

$$\begin{aligned}\frac{\partial \boldsymbol{\Omega}}{\partial \sigma_u^2} &= \mathbf{Q}_1 + \frac{\rho \sigma_v}{2 \sigma_u} \mathbf{Q}_2, \\ \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_v^2} &= \mathbf{Q}_3 + \frac{\rho \sigma_u}{2 \sigma_v} \mathbf{Q}_2, \\ \frac{\partial \boldsymbol{\Omega}}{\partial \rho} &= \sigma_u \sigma_v \mathbf{Q}_2, \\ \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2} &= \frac{-1}{\sigma_u^4 (1 - \rho^2)} \mathbf{Q}_1 + \frac{\rho}{2 \sigma_u^3 \sigma_v (1 - \rho^2)} \mathbf{Q}_2, \\ \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2} &= \frac{-1}{\sigma_v^4 (1 - \rho^2)} \mathbf{Q}_3 + \frac{\rho}{2 \sigma_u \sigma_v^3 (1 - \rho^2)} \mathbf{Q}_2, \quad \text{e} \\ \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho} &= \frac{2\rho}{(1 - \rho^2)^2} \left[\frac{1}{\sigma_u^2} \mathbf{Q}_1 + \frac{1}{\sigma_v^2} \mathbf{Q}_3 \right] - \frac{(1 + \rho^2)}{\sigma_u \sigma_v (1 - \rho^2)^2} \mathbf{Q}_2,\end{aligned}$$

em que $\mathbf{Q}_1 = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}$, $\mathbf{Q}_2 = \begin{pmatrix} 0 & I_J \\ I_J & 0 \end{pmatrix}$ e $\mathbf{Q}_3 = \begin{pmatrix} 0 & 0 \\ 0 & I_J \end{pmatrix}$. Defina

$$\begin{aligned}D_1 &= \mathbf{A}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2}, & D_2 &= \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2}, & D_3 &= \mathbf{A}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2}, \\ D_4 &= \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2}, & D_5 &= \mathbf{A}^* \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho}, & D_6 &= \boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho}.\end{aligned}$$

Dessa forma, os elementos da matriz $\mathbf{W}$ em (3.19) são dados por

$$W_{ij} = \text{tr}(D_{(2i-1)} D_{(2j-1)} + D_{2i} D_{2j} - 2 D_{(2i-1)} D_{2j}), \quad 1 \leq i, j \leq 3. \quad (3.20)$$

Finalmente, valores iniciais podem ser obtidos baseados no modelo de mistura sem e.a. ou no modelo de mistura com efeitos fixos como explicado anteriormente. Desta forma, com valores iniciais para $\boldsymbol{\zeta}^{(\ell)}$ e $\boldsymbol{\Sigma}^{(\ell)}$, precisa-se iterar entre os procedimentos explicados em (3.15), (3.16) e (3.18) até obter a convergência.

3.2 Modelo de Chen com efeitos aleatórios

Como no caso do modelo de Berkson e Gage [1952], no modelo de Chen et al. [1999a] também pode-se fazer uma extensão incluindo efeitos aleatórios. Similarmente ao caso anterior, suponha que

foram estudados casos em J clínicas e que para a j -ésima clínica havia n_j indivíduos, $j = 1, \dots, J$. Para $i = 1, \dots, n_j$, considere M_{ij} como o número de células suscetíveis a terem câncer que o i -ésimo indivíduo da j -ésima clínica possui. Suponha ainda que M_{ij} são independentes com distribuição de Poisson com média θ_{ij} .

Seja Y_{cij} a variável aleatória que denota o tempo até que a c -ésima célula carcinogênica do i -ésimo indivíduo da j -ésima clínica produza câncer. Suponha que as variáveis Y_{cij} são condicionalmente i.i.d. dado M_{ij} com f.d.a. comum $F(t \mid \boldsymbol{\lambda}, \eta_{ij}) = 1 - S(t \mid \boldsymbol{\lambda}, \eta_{ij})$ e independentes de M_{ij} . O tempo para o aparecimento do câncer pode ser definido pela variável aleatória $T_{ij}^* = \min\{Y_{cij}, 1 \leq c \leq M_{ij}\}$, mas $T_{ij}^* = \infty$ com probabilidade 1 se $M_{ij} = 0$. Na prática, estes tempos são observados sujeitos a uma variável C_{ij} de censura, implicando que os valores observados (sob censura à direita) são dados por $T_{ij} = \min(T_{ij}^*, C_{ij})$ e $\delta_{ij} = I(T_{ij}^* \leq C_{ij})$.

Como visto em (2.16), a fração de cura é obtida por $e^{-\theta_{ij}}$, em que é possível incluir covariáveis $\mathbf{z}_{ij}$ do i -ésimo indivíduo da j -ésima clínica e um efeito aleatório V_j que está relacionado à fração de cura da j -ésima clínica fazendo

$$\theta_{ij} = \exp(\mathbf{z}_{ij}^\top \boldsymbol{\beta} + V_j), \quad i = 1, \dots, n_j, j = 1, \dots, J. \quad (3.21)$$

Em particular, Lopes e Bolfarine [2012] consideram que $V_j \sim N(0, \sigma_v^2)$, $j = 1, \dots, J$. Valores baixos (altos) para V_j indicam que os pacientes tratados na j -ésima clínica experimentam uma maior (menor) chance de cura.

Particularmente para o caso de Lopes e Bolfarine [2012], os dados observados são $\mathbf{D}_{obs} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{Z})$, em que $\mathbf{t} = (t_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$, $\boldsymbol{\delta} = (\delta_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$ e $\mathbf{Z}$ é a matriz de covariáveis associadas à fração de cura. Os dados completos são $\mathbf{D}_c = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{V}, \mathbf{Z})$ que incluem os dados observados e os dados não observados $\mathbf{V}$. A função de log-verossimilhança BLUP do modelo (somando sob os possíveis valores do vetor $\mathbf{M}$) é dada por $l = l_1 + l_2$, em que

$$l_1 = \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \delta_{ij} [\log(\theta_{ij}) + \log f(t_{ij} \mid \boldsymbol{\lambda})] - \theta_{ij} [1 - S(t_{ij} \mid \boldsymbol{\lambda})] \right\}, \quad \text{e}$$

$$l_2 = (-1/2) [J \log(2\pi\sigma_v^2) + (1/\sigma_v^2) \mathbf{V}^\top \mathbf{V}].$$

Lopes e Bolfarine [2012] consideraram que $T_{ij} \sim \text{Weibull}(\lambda_{ij}, \nu)$, em que $\lambda_{ij} = \exp(\mathbf{x}_{ij}^\top \boldsymbol{\gamma})$, i.e., uma forma paramétrica para $S_0(t \mid \boldsymbol{\lambda})$ e desenvolveram abordagens clássica e bayesiana para este modelo. No caso da estimação clássica, foi considerada uma abordagem muito similar à utilizada por Yau e Ng [2001] e explicada anteriormente; no caso bayesiano, foram utilizadas distribuições *a priori* próprias para as componentes do vetor $\boldsymbol{\lambda}$ e distribuições *a priori* próprias para as componentes dos vetores $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$.

Uma primeira contribuição deste trabalho de tese é estender o modelo de Lopes e Bolfarine [2012], incluindo efeitos aleatórios em cada clínica tanto para os indivíduos curados V_j quanto para os suscetíveis U_j , para $j = 1, \dots, J$. Para o i -ésimo indivíduo da j -ésima clínica, o vetor $\mathbf{z}_{ij}$ e o efeito aleatório V_j podem ser incluídos como em (3.21). Por outro lado, o vetor $\mathbf{x}_{ij}$ e o efeito aleatório U_j para explicar o efeito da j -ésima clínica na sobrevivência dos indivíduos suscetíveis podem ser incluídos como em (3.2). Em um primeiro caso, podemos considerar que $(U_j, V_j)^\top$, para $j = 1, \dots, J$ são i.i.d. com distribuição comum $N(0, \boldsymbol{\Sigma})$, em que $\boldsymbol{\Sigma}$ é dada como em (3.11).

Diferentemente do modelo de Berkson e Gage, as fragilidades U_j do modelo são incluídos através da função de sobrevivência das células carcinogênicas e não diretamente através da função de sobrevivência dos indivíduos suscetíveis. Contudo, e embora as fragilidades dos dois modelos não são comparáveis, a interpretação é a mesma, pois valores altos (baixos) para U_j indicam que as

células dos pacientes suscetíveis tratados na j -ésima clínica experimentam um maior (menor) risco de formação de tumores carcinogênicos e portanto, o risco desses pacientes será maior. Entretanto, a interpretação de V_j é a mesma que no modelo de [Lopes e Bolfarine \[2012\]](#). Este último modelo aparece como um caso particular deste modelo quando $\sigma_u^2 = 0$, i.e., $P(U_j = 0) = 1$, $j = 1, \dots, J$.

Entretanto, para $S_0(t \mid \boldsymbol{\lambda})$ pode-se utilizar uma abordagem paramétrica utilizando-se a distribuição Weibull e a distribuição Exponencial Segmentada, esta última entregando vantagens de flexibilidade na modelagem discutidas na seção 2.1.2.

Como na subseção anterior, pode-se utilizar uma abordagem clássica baseada nos estimadores REML. Para $\boldsymbol{\Sigma}$ fixa, a função de log-verossimilhança BLUP do modelo é dada por $l = l_1 + l_2$, em que

$$l_1 = \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \delta_{ij} [\log \theta_{ij} + \log f(t_{ij} \mid \eta_{ij}, \boldsymbol{\lambda})] - \theta_{ij} F(t_{ij} \mid \eta_{ij}, \boldsymbol{\lambda}) \right\}, \quad \text{e}$$

$$l_2 = -\frac{1}{2(1-\rho^2)} \left[\frac{\mathbf{U}^\top \mathbf{U}}{\sigma_u^2} + \frac{\mathbf{V}^\top \mathbf{V}}{\sigma_v^2} - \frac{2\rho \mathbf{U}^\top \mathbf{V}}{\sigma_u \sigma_v} \right] - (J/2) \log \{4\pi^2 \sigma_u^2 \sigma_v^2 (1-\rho^2)\}, \quad (3.22)$$

em que $F(t_{ij} \mid \eta_{ij}, \boldsymbol{\lambda})$ e $f(t_{ij} \mid \eta_{ij}, \boldsymbol{\lambda})$ provém de (3.2).

Como no modelo de mistura, quando incorporado efeitos aleatórios tanto na sobrevivência dos indivíduos suscetíveis quanto na fração de cura, não pode ser estimado o vetor de parâmetros $\boldsymbol{\lambda}$ da função de sobrevivência basal dos indivíduos suscetíveis e os efeitos aleatórios U_j , $j = 1, \dots, J$ conjuntamente, pois

$$\begin{aligned} S_0(t \mid \boldsymbol{\lambda})^{\exp \eta_{ij}} &= S_0(t \mid \boldsymbol{\lambda})^{\exp \{\mathbf{x}_{ij}^\top \boldsymbol{\gamma} + U_j\}}, \\ &= \left\{ S_0(t \mid \boldsymbol{\lambda})^{\exp U_j} \right\}^{\exp \mathbf{x}_{ij}^\top \boldsymbol{\gamma}}, \\ &= S_0(t \mid \boldsymbol{\lambda}^*)^{\exp \mathbf{x}_{ij}^\top \boldsymbol{\gamma}}. \end{aligned} \quad (3.23)$$

em que $\boldsymbol{\lambda}^* = g(\boldsymbol{\lambda}, U_j)$. O fato de $S_0(\cdot \mid \boldsymbol{\lambda})$ pertencer à família de riscos proporcionais garante que $S_0(\cdot \mid \boldsymbol{\lambda})$ e $S_0(\cdot \mid \boldsymbol{\lambda})^{\exp \eta_{ij}}$ pertencem à mesma família de distribuições. (Ver a equação (2.9) para o caso particular da distribuição Weibull). Desta forma, para $\boldsymbol{\Sigma}$ fixada, o trio $(\boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{U})$ produzirá a mesma verossimilhança perfilada em (3.22) que o trio $(\boldsymbol{\gamma}, \boldsymbol{\lambda}^*, 0)$. (Na verdade existem infinitos trios que produzem a mesma verossimilhança perfilada, só apresentou-se dois).

Valores iniciais para o algoritmo podem ser dados baseados em um método similar ao descrito na subseção 3.1. Desta forma, o procedimento de estimação novamente deve ser dividido em 3 passos.

Para $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ fixos, o estimador BLUP para $\boldsymbol{\zeta}$ pode ser calculado maximizando (3.22) em relação a $\boldsymbol{\zeta}$. Se $\boldsymbol{\zeta}_0$ é um valor inicial para $\boldsymbol{\zeta}$ e usando o método iterativo de Newton-Raphson, temos que

$$\boldsymbol{\zeta}_{\ell+1} = \boldsymbol{\zeta}_\ell + \mathbf{B}_\ell^{-1} \left(\frac{\partial l}{\partial \boldsymbol{\zeta}} \right) \bigg|_{\boldsymbol{\zeta}=\boldsymbol{\zeta}_\ell}, \quad \ell = 0, 1, \dots, \quad (3.24)$$

em que $\mathbf{B} = -\partial^2 l / \partial \boldsymbol{\zeta}^2$. Denote $\hat{\boldsymbol{\zeta}}$ o estimador BLUP de $\boldsymbol{\zeta}$ obtido da convergência em (3.24) e suponha a partição da matrix $\mathbf{B}$ avaliada em $\hat{\boldsymbol{\zeta}}$ exatamente como em (A.2.2). Detalhes do procedimento em (3.24) podem ser encontrados no Apêndice A.2.1. A variância assintótica do vetor $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}})^\top$ pode ser calculada usando (3.6).

Por outro lado, as componentes da matriz das componentes da variância $\boldsymbol{\Sigma}$ podem ser estimada

usando a equação (3.18). Entretanto, se for assumida a independência dos efeitos aleatórios dentro de cada cluster (caso em que $\rho = 0$), então as componentes de Σ podem ser estimadas usando (3.8). A variância dessas componentes podem ser calculadas usando (3.9) ou (3.19) dependendo de se for assumida a dependência ou não dos efeitos aleatórios dentro de cada clínica.

Uma vez obtidos $\hat{\zeta}$ e $\hat{\Sigma}$, é construída uma função de log-verossimilhança perfilada para o vetor λ de (3.22), a qual será denominada l_p . O estimador de λ também pode ser encontrado utilizando o algoritmo de Newton-Raphson. Assim, se λ_0 é um valor inicial para λ , então o procedimento iterativo fica

$$\lambda_{\ell+1} = \lambda_{\ell} + C_{\ell}^{-1} \left(\frac{\partial l_p}{\partial \lambda} \right) \bigg|_{\lambda=\lambda_{\ell}}, \quad \ell = 0, 1, \dots, \quad (3.25)$$

em que $C = -\partial^2 l_p / \partial \lambda^2$. O estimador para λ é obtido da convergência em (3.25).

As expressões para os vetores $\partial l / \partial \zeta$, $\partial l_p / \partial \lambda$ e as matrizes B e C em (3.24) e (3.25) para o modelo de Chen com efeitos aleatórios podem ser encontradas no Apêndice A.2.1, considerando como função de sobrevivência basal a distribuição Weibull e a distribuição exponencial segmentada.

Desta forma, deve-se iterar entre as equações (3.24), (3.18) e (3.25) até obter a convergência para todos os parâmetros. Se for assumido $\rho = 0$, então a segunda equação é substituída por (3.8).

3.2.1 Estudo de simulação: Recuperação de parâmetros

Nesta seção serão apresentados estudos de simulação feitos para o modelo de Chen com efeitos aleatórios com a finalidade avaliar o desempenho dos estimadores propostos. Nesta linha, será avaliado o vício dos estimadores propostos e se a variância de cada estimador está ou não sendo bem estimada.

Para simular do modelo em questão, suponha que os pacientes estão divididos em J clínicas, em que a j -ésima clínica tem n_j pacientes. Suponha também que o i -ésimo indivíduo da j -ésima clínica tem covariáveis x_{ij} e z_{ij} de dimensão r e s , respectivamente, para explicar a sobrevivência dos indivíduos suscetíveis e a fração de cura. Para Σ considerado como em (3.11), foram simulados os efeitos aleatórios $(U_1, V_1), \dots, (U_J, V_J)$ da distribuição $N(0, \Sigma)$. Defina η_{ij} e θ_{ij} para $j = 1, \dots, J$, $i = 1, \dots, n_j$ como em (3.2) e (3.21), respectivamente em que γ e β são vetores conhecidos de tamanho r e s respectivamente. Foi simulado M_{ij} da distribuição Poisson com média θ_{ij} . Se M_{ij} é zero, então defina o tempo de falha do i -ésimo indivíduo da j -ésima clínica, T_{ij}^* , como ∞ . Se $M_{ij} > 0$, então simule $Y_{1ij}, \dots, Y_{M_{ij}ij}$ tempos da densidade $h_0(t | \lambda) e^{\eta_{ij} [S_0(t | \lambda)]^{\exp(\eta_{ij})}}$, em que $S_0(t | \lambda)$ é conhecida, que neste trabalho está sendo considerada como a distribuição Weibull ou a distribuição Exponencial Segmentada e defina $T_{ij}^* = \min(Y_{a ij}, a = 1, \dots, M_{ij})$. Finalmente, para um tempo de censura C fixado, defina as variáveis $T_{ij} = \min(T_{ij}^*, C)$ e $\delta_{ij} = I(T_{ij}^* \leq C)$, que representam os tempos observados (de falha ou censura) e o indicador de falha para cada indivíduo.

Em um primeiro estudo, será considerada $S_0(t | \lambda)$ proveniente da distribuição Weibull de parâmetros $\lambda = (\lambda, \nu)$ como em (2.1) e as mesmas condições de Li et al. [2001], i.e., considerando:

- $J = 10$ clínicas;
- $n_j = 15$, $j = 1, \dots, J$ pacientes em cada clínica;
- em cada clínica, 5 pacientes no grupo de tratamento ($x_{ij} = 1$) e 10 pacientes no grupo de controle ($x_{ij} = 0$);
- as covariáveis para explicar a fração de cura como $z_{ij} = (1, x_{ij})$;

- $\gamma = -0.5$ e $\beta = (0.5, -0.5)^\top$ e ρ variando em $\{-0.2, -0.4, -0.6, -0.8\}$, e para cada combinação de esses parâmetros, considerando a simulação com $\{\nu = 1, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000\}$ como caso basal e examinando o efeito ao mudar os valores de $\nu, \sigma_u^2, \sigma_v^2$ e C ;
- o valor de λ será 0.01 para todas as simulações; e
- o valor de C foi diminuído de $C = 1000$ para $C = 500$ seguindo [Lai e Yau \[2008\]](#).

Para cada grupo de parâmetros, foram feitas 500 réplicas, que são apresentadas na Tabela 3.1 e nas Figuras 3.1, 3.2, 3.3 e 3.4. Nas Figuras, a curva em azul representa a distribuição normal com média e desvio padrão de acordo aos resultados obtidos para cada parâmetro e a linha em vermelho representa a respectiva média amostral. Nas tabelas, SE_1 e SE_2 representam a média dos desvios padrão assintóticos e o desvio padrão das estimativas obtidas nas 500 simulações.

Dos resultados, tem-se que o vício médio do estimador de β_1 é aceitável, embora de forma geral o vício médio do estimador de β_0 é um pouco maior. As variâncias de ambos estimadores são levemente subestimadas.

O vício médio do estimador de γ também é um pouco elevado, superando às vezes o 5% em relação ao parâmetro estimado, mas ele diminui quando ν aumenta de 1 para 1.3 e quando C diminui de 1000 para 500. De forma geral, o desvio padrão deste estimador é bem estimado, sendo levemente subestimado em casos pontuais.

O vício médio do estimador de σ_u^2 também é aceitável. Mas, nos casos considerados e visto de forma geral, o vício médio de σ_v^2 é menor. O vício dos dois parâmetros aumenta quando eles aumentam de 0.5 para 1 nos conjuntos de parâmetros II e III. Os desvios dos dois parâmetros também são subestimados levemente de maneira geral. Só no caso em que $\sigma_u^2 = \sigma_v^2 = 1$ são subestimados consideravelmente.

Em relação a $\hat{\rho}$, ele sempre sobrestimou o verdadeiro valor, exceto no caso da simulação 6. Mas o vício médio deste estimador também pode ser considerado aceitável. De forma geral, o desvio padrão de $\hat{\rho}$ é bem estimado exceto no caso em que a variância dos efeitos aleatórios aumenta de 0.5 para 1, em cujo caso é notoriamente subestimada.

Visto de uma forma mais geral, quando ν aumenta de 1 para 1.3 melhorou tanto o desempenho dos estimadores envolvidos no sentido do vício quando na estimativa da sua respectiva variância. No entanto e de forma geral, o aumento da variância dos efeitos aleatórios piorou o desempenho de todos os parâmetros em ambos sentidos, por conta de ter-se uma maior variabilidade tanto no grupo dos curados quanto no grupo dos suscetíveis.

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(a) Grupo I ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.2$)				
Simulação 1 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.048	0.286	0.290
β_0	0.5	0.042	0.234	0.268
β_1	-0.5	-0.011	0.246	0.240
σ_u^2	0.5	-0.012	0.316	0.338
σ_v^2	0.5	0.027	0.316	0.361
ρ	-0.2	0.031	0.470	0.484
Simulação 2 ($\nu = 1.3, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.027	0.268	0.263
β_0	0.5	0.028	0.230	0.279
β_1	-0.5	-0.007	0.239	0.233
σ_u^2	0.5	0.046	0.334	0.325
σ_v^2	0.5	0.001	0.300	0.323
ρ	-0.2	0.016	0.424	0.473
Simulação 3 ($\nu = 1.0, \sigma_u^2 = 1.0, \sigma_v^2 = 1.0, C = 1000$)				
γ	-0.5	0.051	0.323	0.319
β_0	0.5	0.054	0.311	0.358
β_1	-0.5	0.007	0.269	0.284
σ_u^2	1.0	-0.007	0.577	0.614
σ_v^2	1.0	0.033	0.578	0.716
ρ	-0.2	0.017	0.376	0.437
Simulação 4 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.023	0.351	0.367
β_0	0.5	0.025	0.235	0.264
β_1	-0.5	0.004	0.276	0.268
σ_u^2	0.5	0.010	0.351	0.366
σ_v^2	0.5	0.033	0.326	0.389
ρ	-0.2	0.027	0.557	0.505
(a) Grupo II ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.4$)				
Simulação 5 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.028	0.286	0.302
β_0	0.5	0.019	0.219	0.263
β_1	-0.5	-0.001	0.245	0.249
σ_u^2	0.5	-0.013	0.313	0.328
σ_v^2	0.5	-0.006	0.298	0.327
ρ	-0.4	0.071	0.410	0.466
Simulação 6 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.012	0.269	0.262
β_0	0.5	0.023	0.217	0.255
β_1	-0.5	0.021	0.240	0.230
σ_u^2	0.5	0.034	0.331	0.337
σ_v^2	0.5	0.009	0.304	0.323
ρ	-0.4	-0.017	0.378	0.408
continua na próxima página.				

Tabela 3.1: Vício estimado dos estimadores REML no modelo de Chen com efeitos aleatórios dependentes: caso sobrevivência basal Weibull.

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(a) Grupo II ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.4$)				
Simulação 7 ($\nu = 1.0, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.021	0.323	0.343
β_0	0.5	0.050	0.282	0.343
β_1	-0.5	0.018	0.269	0.271
σ_u^2	1.0	-0.039	0.558	0.598
σ_v^2	1.0	-0.051	0.534	0.591
ρ	-0.4	0.039	0.344	0.414
Simulação 8 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.041	0.352	0.374
β_0	0.5	0.039	0.219	0.245
β_1	-0.5	-0.013	0.278	0.275
σ_u^2	0.5	-0.022	0.332	0.331
σ_v^2	0.5	-0.007	0.306	0.329
ρ	-0.4	0.076	0.492	0.480
(c) Grupo III ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.6$)				
Simulação 9 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.043	0.285	0.292
β_0	0.5	0.047	0.200	0.225
β_1	-0.5	-0.016	0.245	0.241
σ_u^2	0.5	0.032	0.335	0.362
σ_v^2	0.5	0.008	0.304	0.354
ρ	-0.6	0.040	0.338	0.376
Simulação 10 ($\nu = 1.3, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.015	0.268	0.287
β_0	0.5	0.033	0.201	0.240
β_1	-0.5	0.002	0.239	0.240
σ_u^2	0.5	0.044	0.335	0.361
σ_v^2	0.5	0.007	0.302	0.329
ρ	-0.6	0.060	0.387	0.395
Simulação 11 ($\nu = 1.0, \sigma_u^2 = 1.0, \sigma_v^2 = 1.0, C = 1000$)				
γ	-0.5	0.052	0.316	0.347
β_0	0.5	0.042	0.261	0.314
β_1	-0.5	-0.017	0.265	0.272
σ_u^2	1.0	-0.026	0.562	0.607
σ_v^2	1.0	-0.003	0.553	0.670
ρ	-0.6	0.070	0.284	0.371
Simulação 12 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	-0.007	0.404	0.484
β_0	0.5	0.041	0.199	0.220
β_1	-0.5	0.017	0.329	0.361
σ_u^2	0.5	-0.015	0.333	0.354
σ_v^2	0.5	0.000	0.307	0.347
ρ	-0.6	0.037	0.411	0.402

continua na próxima página.

Tabela 3.1 - continuação.

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(c) Grupo IV ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.8$)				
Simulação 13 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.039	0.281	0.290
β_0	0.5	0.054	0.177	0.187
β_1	-0.5	-0.016	0.243	0.254
σ_u^2	0.5	0.036	0.336	0.461
σ_v^2	0.5	0.008	0.302	0.353
ρ	-0.8	0.035	0.279	0.268
Simulação 14 ($\nu = 1.3, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.013	0.267	0.274
β_0	0.5	0.043	0.173	0.187
β_1	-0.5	-0.008	0.239	0.250
σ_u^2	0.5	0.023	0.325	0.380
σ_v^2	0.5	0.000	0.297	0.342
ρ	-0.8	0.026	0.260	0.278
Simulação 15 ($\nu = 1.0, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.032	0.309	0.330
β_0	0.5	0.054	0.212	0.253
β_1	-0.5	0.013	0.261	0.252
σ_u^2	1.0	-0.021	0.561	0.716
σ_v^2	1.0	-0.060	0.521	0.590
ρ	-0.8	0.039	0.195	0.244
Simulação 16 ($\nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.031	0.333	0.333
β_0	0.5	0.038	0.174	0.199
β_1	-0.5	0.013	0.264	0.268
σ_u^2	0.5	-0.008	0.333	0.352
σ_v^2	0.5	-0.026	0.292	0.321
ρ	-0.8	0.040	0.337	0.300

Tabela 3.1 - continuação.

Figura 3.1: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo I

Figura 3.2: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo II

(a) Simulação 9

(b) Simulação 10

(c) Simulação 11

(d) Simulação 12

Figura 3.3: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo III

(a) Simulação 13

(b) Simulação 14

(c) Simulação 15

(d) Simulação 16

Figura 3.4: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Weibull: Grupo IV

Também será apresentado um estudo de simulação considerando $S_0(t \mid \boldsymbol{\lambda})$ proveniente da distribuição exponencial segmentada com $L = 3$. A função de densidade deste modelo é dada em (2.3). As condições deste estudo são similares ao estudo de simulação feito anteriormente, i.e.,

- $J = 10$ clínicas;
- $n_j = 15$, $j = 1, \dots, J$ pacientes em cada clínica;
- em cada clínica, 5 pacientes no grupo de tratamento ($x_{ij} = 1$) e 10 pacientes no grupo de controle ($x_{ij} = 0$) na proporção 1:2;
- as covariáveis para explicar a fração de cura como $\mathbf{z}_{ij} = (1, x_{ij})$;
- $\boldsymbol{\gamma} = -0.5$ e $\boldsymbol{\beta} = (0.5, -0.5)^\top$ e ρ variando em $\{-0.2, -0.4, -0.6, -0.8\}$, e para cada combinação desses parâmetros, considerando a simulação com $\{\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000\}$ como base e examinando o efeito ao mudar os valores de $\lambda_2, \lambda_3, \sigma_u^2, \sigma_v^2$ e C ;
- o valor de λ_1 será 0.01 para todas as simulações.

Também foram feitas 500 réplicas para cada grupo de parâmetros, que são apresentadas na Tabela 3.2 e nas Figuras 3.5, 3.6, 3.7 e 3.8. Assim como nas Figuras anteriores, a curva em azul e a linha em vermelho representam a distribuição normal e a média amostral.

De acordo com os resultados obtidos, tem-se que para os estimadores de β_0 e β_1 , o vício médio é pequeno e mantém-se relativamente constante para todos os casos considerados. Os desvios padrões de ambos estimadores são levemente subestimados, embora de forma geral para $\hat{\beta}_0$ esse efeito é maior. Uma conclusão similar pode ser obtida para o estimador de γ .

O vício médio do estimador de σ_u^2 é maior, visto de maneira global, do que o vício do estimador de σ_v^2 . O aumento de λ_2 e λ_3 melhora o desempenho de ambos estimadores.

No entanto, a diminuição de C de 1000 a 500 melhora o desempenho de $\hat{\sigma}_u^2$, mas piora o de $\hat{\sigma}_v^2$. Contudo, os desvios dos dois estimadores são levemente subestimados, sendo mais grave a subestimação só no caso em que $\rho = -0.8$ e $\sigma_u^2 = \sigma_v^2 = 1$.

Em relação a ρ , este também é bem estimado em todos os casos, aumentando o vício do estimador só nos casos em que C diminui para 500. O desvio deste estimador é subestimado de forma geral, mas isto vai melhorando à medida que o verdadeiro valor de ρ aproxima-se de -1.

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(a) Grupo I ($\lambda_1 = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.2$)				
Simulação 1 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.031	0.272	0.290
β_0	0.5	0.024	0.229	0.278
β_1	-0.5	0.008	0.241	0.257
σ_u^2	0.5	0.053	0.339	0.342
σ_v^2	0.5	0.011	0.306	0.349
ρ	-0.2	-0.010	0.428	0.488
Simulação 2 ($\lambda_2 = 0.03, \lambda_3 = 0.05, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.035	0.267	0.282
β_0	0.5	0.030	0.230	0.267
β_1	-0.5	0.013	0.239	0.242
σ_u^2	0.5	-0.011	0.309	0.313
σ_v^2	0.5	0.007	0.303	0.346
ρ	-0.2	0.042	0.453	0.502
Simulação 3 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.024	0.300	0.300
β_0	0.5	0.036	0.302	0.372
β_1	-0.5	-0.001	0.258	0.250
σ_u^2	1.0	0.082	0.605	0.612
σ_v^2	1.0	-0.026	0.544	0.628
ρ	-0.2	0.019	0.362	0.429
Simulação 4 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.042	0.291	0.303
β_0	0.5	0.027	0.237	0.267
β_1	-0.5	-0.005	0.248	0.240
σ_u^2	0.5	-0.006	0.320	0.354
σ_v^2	0.5	0.044	0.324	0.377
ρ	-0.2	0.039	0.493	0.507
(a) Grupo II ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.4$)				
Simulação 5 ($\lambda_2 = 0.03, \lambda_3 = 0.05, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.028	0.271	0.275
β_0	0.5	0.019	0.220	0.247
β_1	-0.5	-0.018	0.241	0.245
σ_u^2	0.5	0.025	0.327	0.339
σ_v^2	0.5	0.027	0.312	0.349
ρ	-0.4	0.039	0.437	0.460
Simulação 6 ($\lambda_2 = 0.03, \lambda_3 = 0.05, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.031	0.269	0.282
β_0	0.5	0.045	0.220	0.263
β_1	-0.5	-0.018	0.240	0.258
σ_u^2	0.5	0.033	0.330	0.387
σ_v^2	0.5	0.011	0.305	0.338
ρ	-0.4	0.051	0.398	0.445
continua na próxima página.				

Tabela 3.2: Vício estimado dos estimadores REML no modelo de Chen com efeitos aleatórios dependentes: caso sobrevivência Exponencial Segmentada ($L = 3$).

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(a) Grupo II ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.4$)				
Simulação 7 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.016	0.298	0.314
β_0	0.5	0.017	0.282	0.335
β_1	-0.5	0.022	0.257	0.263
σ_u^2	1.0	0.033	0.580	0.609
σ_v^2	1.0	-0.044	0.533	0.676
ρ	-0.4	0.060	0.331	0.403
Simulação 8 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.036	0.293	0.282
β_0	0.5	0.037	0.223	0.253
β_1	-0.5	0.006	0.248	0.244
σ_u^2	0.5	0.031	0.337	0.334
σ_v^2	0.5	0.033	0.318	0.338
ρ	-0.4	0.022	0.409	0.446
(c) Grupo III ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.6$)				
Simulação 9 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.031	0.271	0.274
β_0	0.5	0.037	0.202	0.223
β_1	-0.5	-0.014	0.241	0.243
σ_u^2	0.5	0.054	0.340	0.352
σ_v^2	0.5	0.029	0.312	0.347
ρ	-0.6	0.017	0.338	0.356
Simulação 10 ($\lambda_2 = 0.03, \lambda_3 = 0.05, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.035	0.268	0.274
β_0	0.5	0.028	0.197	0.215
β_1	-0.5	-0.001	0.239	0.235
σ_u^2	0.5	0.033	0.329	0.328
σ_v^2	0.5	-0.002	0.297	0.363
ρ	-0.6	0.028	0.355	0.376
Simulação 11 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.037	0.299	0.302
β_0	0.5	0.030	0.256	0.295
β_1	-0.5	-0.007	0.258	0.260
σ_u^2	1.0	0.074	0.601	0.628
σ_v^2	1.0	-0.011	0.548	0.576
ρ	-0.6	0.017	0.275	0.331
Simulação 12 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.042	0.288	0.318
β_0	0.5	0.020	0.198	0.229
β_1	-0.5	0.008	0.246	0.245
σ_u^2	0.5	0.007	0.324	0.355
σ_v^2	0.5	-0.013	0.294	0.321
ρ	-0.6	0.044	0.418	0.399

continua na próxima página.

Tabela 3.2: continuação.

Parâmetros	Valor verdadeiro	Vício médio	SE_1	SE_2
(c) Grupo IV ($\lambda = 0.01, \gamma = -0.5, \beta_0 = 0.5, \beta_1 = -0.5, \rho = -0.8$)				
Simulação 13 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.021	0.268	0.274
β_0	0.5	0.028	0.172	0.189
β_1	-0.5	0.022	0.240	0.257
σ_u^2	0.5	0.052	0.339	0.400
σ_v^2	0.5	0.003	0.298	0.348
ρ	-0.8	0.032	0.266	0.271
Simulação 14 ($\lambda_2 = 0.03, \lambda_3 = 0.05, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000$)				
γ	-0.5	0.031	0.268	0.283
β_0	0.5	0.034	0.175	0.193
β_1	-0.5	0.008	0.240	0.253
σ_u^2	0.5	0.029	0.327	0.350
σ_v^2	0.5	0.001	0.297	0.325
ρ	-0.8	0.035	0.275	0.278
Simulação 15 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 1, \sigma_v^2 = 1, C = 1000$)				
γ	-0.5	0.051	0.293	0.298
β_0	0.5	0.057	0.213	0.251
β_1	-0.5	-0.015	0.256	0.254
σ_u^2	1.0	0.004	0.567	0.715
σ_v^2	1.0	-0.069	0.515	0.648
ρ	-0.8	0.037	0.192	0.223
Simulação 16 ($\lambda_2 = 0.02, \lambda_3 = 0.03, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 500$)				
γ	-0.5	0.035	0.285	0.285
β_0	0.5	0.029	0.172	0.188
β_1	-0.5	-0.005	0.244	0.244
σ_u^2	0.5	-0.004	0.317	0.336
σ_v^2	0.5	-0.033	0.282	0.309
ρ	-0.8	0.014	0.258	0.240

Tabela 3.2: *continuação.*

Figura 3.5: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo I

(a) Simulação 5

(b) Simulação 6

(c) Simulação 7

(d) Simulação 8

Figura 3.6: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo II

Figura 3.7: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo III

(a) Simulação 13

(b) Simulação 14

(c) Simulação 15

(d) Simulação 16

Figura 3.8: Histogramas das estimativas dos parâmetros para o modelo de Chen com efeitos aleatórios dependentes e sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo IV

3.2.2 Comparação do modelo de Chen com efeitos aleatórios normais dependentes *versus* independentes

Uma questão muito importante a responder e que não foi encontrado na literatura, é a questão de qual é o impacto nas estimativas do modelo se for utilizado, por engano, o modelo com efeitos aleatórios dependentes, i.e., se o modelo que deveria ser usado é aquele com efeitos aleatórios independentes. A pergunta inversa também é válida.

Com a finalidade de responder a estas perguntas, foram feitos dois estudos de simulação: o primeiro utilizando como modelo basal para os indivíduos suscetíveis o modelo Weibull e o segundo usando o modelo exponencial segmentado. Mantendo fixados os valores para os parâmetros do modelo γ , β , σ_u^2 , σ_v^2 e os parâmetros que definem a sobrevivência dos indivíduos suscetíveis, variou-se apenas o valor de ρ , procurando cenários nos quais é recomendável utilizar o modelo com efeitos aleatórios dependentes e nos quais é recomendável utilizar efeitos aleatórios independentes.

Os dados foram simulados do modelo com efeitos aleatórios dependentes considerando:

- $J = 10$ clínicas;
- $n_j = 15$, $j = 1, \dots, J$ pacientes em cada clínica;
- em cada clínica, 5 pacientes no grupo de tratamento ($x_{ij} = 1$) e 10 pacientes no grupo de controle ($x_{ij} = 0$);
- as covariáveis para explicar a fração de cura como $\mathbf{z}_{ij} = (1, x_{ij})$;
- três grupos de parâmetros para γ e β . Grupo I: $\gamma = -0.5$ e $\beta = (0, 0.5)^\top$; Grupo II: $\gamma = -0.5$ e $\beta = (-0.5, 0.5)^\top$; Grupo III: $\gamma = -1$ e $\beta = (-1, 1)^\top$;
- em todas as simulações, no caso de utilizar como sobrevivência basal para os indivíduos não curados a distribuição Weibull, será considerado $(\lambda, \nu) = (1, 1)$ e para a distribuição exponencial segmentada, $(\lambda_1, \lambda_2, \lambda_3) = (0.01, 0.02, 0.03)$;
- as variâncias dos efeitos aleatórios como $\sigma_u^2 = \sigma_v^2 = 0.5$ para todas as simulações;
- o valor de C será considerado como $C = 1000$; e
- o valor de ρ variando em $\{-0.2, -0.4, -0.6, -0.8\}$.

Posteriormente, foram estimados todos os parâmetros considerando o modelo de Chen com efeitos aleatórios normais em suas versões dependente e independente e a respectiva função de sobrevivência basal para os indivíduos suscetíveis. Os resultados registrados para cada modelo foram o vício médio de cada estimador e o respectivo desvio padrão das estimativas obtidas (i.e., o desvio padrão amostral dos estimadores obtido em cada simulação).

O estudo para a distribuição Weibull é apresentado na Tabela 3.3 e nas Figuras 3.9, 3.10 e 3.11 e para a distribuição exponencial segmentada na Tabela 3.4 e nas Figuras 3.12, 3.13 e 3.14.

Nas Figuras, é representado o histograma das estimativas obtidas no modelo de Chen com efeitos aleatórios dependentes, enquanto que as linhas em azul e vermelho estão associadas às distribuições normais do modelo com efeitos aleatórios dependentes e independentes respectivamente, baseado nas estimativas obtidas em cada réplica. Dos resultados obtidos, obtém-se conclusões similares considerando as duas distribuições basais para $S_0(t | \lambda)$.

De uma forma geral, verifica-se nos casos em que $\rho = -0.2$ que os estimadores do modelo com e.a.'s independentes são menos viciados do que os estimadores do modelo com e.a.'s dependentes,

além de que neste caso em particular, o desvio padrão dos estimadores do modelo com e.a.'s independentes é menor que o desvio padrão dos estimadores no modelo com e.a.'s dependentes. Mas, quando $\rho \leq -0.4$, as conclusões são invertidas, logrando-se um melhor desempenho no modelo com e.a.'s dependentes. Mas o impacto é sempre maior nos estimadores das variâncias dos e.a.'s que nos demais parâmetros, obtendo-se subestimações que em alguns casos inclusive superam, em média, o 20% em relação ao parâmetro estimado. De fato, quando ρ é mais próximo de -1 , maior será o efeito da subestimação sobre as variâncias dos efeitos aleatórios.

As conclusões que podem ser obtidas deste estudo são as seguintes:

- Se a correlação de U_j e V_j é fraca, então é recomendável utilizar o modelo com e.a.'s independentes, pois as estimativas dos parâmetros obtidas serão menos viciadas e mais precisas;
- Se a correlação de U_j e V_j é forte, então é recomendável utilizar o modelo com e.a.'s dependentes, pois as estimativas dos parâmetros obtidas serão menos viciadas;
- É importante destacar que, de forma geral, os desvios padrão do modelo considerando e.a.'s independentes são menores que os desvios padrões do modelo com e.a.'s dependentes, mas se os estimadores forem viciados isso pode levar a obter-se conclusões erradas;
- O efeito pontual sobre as estimativas das variâncias dos efeitos aleatórios ao considerar independência quando a correlação é forte é a subestimação, o que também produzirá uma subestimação (em módulo) sob os e.a.'s. Isto faz com que o efeito real que as clínicas possam ter sobre os pacientes seja desconsiderado.

Contudo, não é possível definir um valor de corte para ρ para o qual é recomendável utilizar um modelo ou outro, pois isso também vai depender dos demais parâmetros do modelo. Com exceção do caso em que tenha-se informação *a priori* sobre uma forte correlação entre os e.a.'s associados a uma mesma clínica, a recomendação é ajustar o modelo com e.a.'s dependentes e independentes e utilizar algum critério de seleção de modelos para selecionar o mais adequado.

Parâmetros	Valor verdadeiro	E.a. independentes	E.a. dependentes
		Vício Médio (SE)	Vício Médio (SE)
(a) Grupo I ($\lambda = 0.01, \nu = 1.0, \gamma = -0.5, \beta_0 = 0.0, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 1 ($\rho = -0.2$)			
γ	-0.5	0.019 (0.289)	0.006 (0.291)
β_0	0.0	0.020 (0.253)	0.034 (0.241)
β_1	0.5	0.001 (0.250)	0.010 (0.252)
σ_u^2	0.5	-0.038 (0.297)	0.012 (0.325)
σ_v^2	0.5	-0.022 (0.286)	0.028 (0.315)
ρ	-0.2	-	0.021 (0.486)
Simulação 2 ($\rho = -0.4$)			
γ	-0.5	0.060 (0.287)	0.037 (0.290)
β_0	0.0	0.014 (0.246)	0.048 (0.228)
β_1	0.5	-0.034 (0.249)	-0.020 (0.251)
σ_u^2	0.5	-0.045 (0.293)	0.039 (0.337)
σ_v^2	0.5	-0.055 (0.270)	0.024 (0.313)
ρ	-0.4	-	0.035 (0.409)
Simulação 3 ($\rho = -0.6$)			
γ	-0.5	0.073 (0.278)	0.042 (0.284)
β_0	0.0	-0.015 (0.237)	0.031 (0.211)
β_1	0.5	-0.030 (0.244)	-0.011 (0.248)
σ_u^2	0.5	-0.091 (0.270)	0.032 (0.331)
σ_v^2	0.5	-0.102 (0.246)	0.005 (0.301)
ρ	-0.6	-	0.040 (0.359)
Simulação 4 ($\rho = -0.8$)			
γ	-0.5	0.083 (0.271)	0.048 (0.278)
β_0	0.0	-0.015 (0.226)	0.029 (0.187)
β_1	0.5	-0.042 (0.241)	-0.019 (0.245)
σ_u^2	0.5	-0.167 (0.233)	-0.006 (0.311)
σ_v^2	0.5	-0.149 (0.223)	-0.021 (0.287)
ρ	-0.8	-	0.032 (0.268)
(a) Grupo II ($\lambda = 0.01, \nu = 1.0, \gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 1 ($\rho = -0.2$)			
γ	-0.5	0.023 (0.292)	0.017 (0.294)
β_0	-0.5	0.026 (0.262)	0.043 (0.248)
β_1	0.5	-0.009 (0.258)	-0.004 (0.259)
σ_u^2	0.5	-0.018 (0.319)	0.043 (0.349)
σ_v^2	0.5	-0.023 (0.297)	0.017 (0.318)
ρ	-0.2	-	-0.008 (0.494)
Simulação 2 ($\rho = -0.4$)			
γ	-0.5	0.038 (0.288)	0.031 (0.291)
β_0	-0.5	0.026 (0.256)	0.044 (0.236)
β_1	0.5	-0.024 (0.256)	-0.017 (0.258)
σ_u^2	0.5	-0.064 (0.296)	0.017 (0.334)
σ_v^2	0.5	-0.053 (0.282)	0.006 (0.312)
ρ	-0.4	-	0.002 (0.471)

continua na próxima página.

continua na próxima página.

Tabela 3.3: *Vício e Desvio Padrão estimado dos estimadores REML no modelo de Chen com efeitos aleatórios independentes e dependentes: caso sobrevivência Weibull.*

Parâmetros	Valor verdadeiro	E.a. independentes	E.a. dependentes
		Vício Médio (SE)	Vício Médio (SE)
(a) Grupo II ($\lambda = 0.01, \nu = 1.0, \gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 7 ($\rho = -0.6$)			
γ	-0.5	0.026 (0.286)	0.018 (0.290)
β_0	-0.5	0.030 (0.251)	0.049 (0.225)
β_1	0.5	-0.027 (0.255)	-0.019 (0.257)
σ_u^2	0.5	-0.091 (0.282)	0.016 (0.333)
σ_v^2	0.5	-0.075 (0.271)	-0.001 (0.307)
ρ	-0.6	-	0.021 (0.415)
Simulação 8 ($\rho = -0.8$)			
γ	-0.5	0.053 (0.281)	0.044 (0.286)
β_0	-0.5	0.021 (0.247)	0.038 (0.209)
β_1	0.5	-0.038 (0.254)	-0.026 (0.256)
σ_u^2	0.5	-0.122 (0.267)	0.014 (0.331)
σ_v^2	0.5	-0.100 (0.259)	-0.011 (0.301)
ρ	-0.8	-	0.055 (0.328)
(a) Grupo III ($\lambda = 0.01, \nu = 1.0, \gamma = -1.0, \beta_0 = -1.0, \beta_1 = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 9 ($\rho = -0.2$)			
γ	-1.0	0.054 (0.335)	0.043 (0.335)
β_0	-1.0	0.054 (0.273)	0.059 (0.262)
β_1	1.0	-0.022 (0.289)	-0.012 (0.289)
σ_u^2	0.5	0.000 (0.351)	0.040 (0.373)
σ_v^2	0.5	-0.059 (0.294)	-0.020 (0.316)
ρ	-0.2	-	0.024 (0.624)
Simulação 10 ($\rho = -0.4$)			
γ	-1.0	0.040 (0.338)	0.027 (0.337)
β_0	-1.0	0.054 (0.274)	0.070 (0.258)
β_1	1.0	-0.036 (0.291)	-0.022 (0.291)
σ_u^2	0.5	-0.011 (0.347)	0.065 (0.384)
σ_v^2	0.5	-0.046 (0.301)	0.014 (0.332)
ρ	-0.4	-	0.006 (0.568)
Simulação 11 ($\rho = -0.6$)			
γ	-1.0	0.047 (0.331)	0.039 (0.330)
β_0	-1.0	0.029 (0.269)	0.037 (0.250)
β_1	1.0	-0.031 (0.289)	-0.015 (0.290)
σ_u^2	0.5	-0.060 (0.322)	0.039 (0.370)
σ_v^2	0.5	-0.082 (0.284)	-0.010 (0.320)
ρ	-0.6	-	0.072 (0.530)
Simulação 12 ($\rho = -0.8$)			
γ	-1.0	0.066 (0.324)	0.050 (0.324)
β_0	-1.0	0.031 (0.266)	0.038 (0.238)
β_1	1.0	-0.038 (0.286)	-0.016 (0.287)
σ_u^2	0.5	-0.112 (0.297)	0.018 (0.359)
σ_v^2	0.5	-0.096 (0.277)	0.000 (0.324)
ρ	-0.8	-	0.060 (0.460)

Tabela 3.3 - continuação.

Figura 3.9: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo I

(a) Simulação 5

(b) Simulação 6

(c) Simulação 7

(d) Simulação 8

Figura 3.10: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo II

Figura 3.11: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Weibull: Grupo III

Parâmetros	Valor verdadeiro	E.a. independentes Vício Médio (SE)	E.a. dependentes Vício Médio (SE)
(a) Grupo I ($\lambda_1 = 0.01, \lambda_2 = 0.02, \lambda_3 = 0.03, \gamma = -0.5, \beta_0 = 0.0, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 1 ($\rho = -0.2$)			
γ	-0.5	0.022 (0.270)	0.014 (0.272)
β_0	0.0	0.007 (0.251)	0.017 (0.239)
β_1	0.5	0.005 (0.242)	0.010 (0.243)
σ_u^2	0.5	-0.038 (0.292)	0.004 (0.314)
σ_v^2	0.5	-0.029 (0.282)	0.013 (0.306)
ρ	-0.2	-	0.023 (0.448)
Simulação 2 ($\rho = -0.4$)			
γ	-0.5	0.024 (0.268)	0.003 (0.272)
β_0	0.0	0.041 (0.242)	0.059 (0.222)
β_1	0.5	-0.009 (0.241)	0.005 (0.243)
σ_u^2	0.5	-0.069 (0.276)	0.016 (0.320)
σ_v^2	0.5	-0.075 (0.258)	0.005 (0.302)
ρ	-0.4	-	0.002 (0.374)
Simulação 3 ($\rho = -0.6$)			
γ	-0.5	0.040 (0.267)	0.013 (0.272)
β_0	0.0	0.002 (0.236)	0.038 (0.211)
β_1	0.5	-0.014 (0.240)	0.003 (0.243)
σ_u^2	0.5	-0.081 (0.270)	0.035 (0.328)
σ_v^2	0.5	-0.102 (0.246)	0.005 (0.301)
ρ	-0.6	-	0.027 (0.329)
Simulação 4 ($\rho = -0.8$)			
γ	-0.5	0.050 (0.263)	0.013 (0.270)
β_0	0.0	-0.005 (0.229)	0.029 (0.190)
β_1	0.5	-0.026 (0.239)	-0.002 (0.243)
σ_u^2	0.5	-0.154 (0.236)	0.007 (0.314)
σ_v^2	0.5	-0.133 (0.230)	-0.006 (0.293)
ρ	-0.8	-	0.030 (0.286)
(a) Grupo II ($\lambda_1 = 0.01, \lambda_2 = 0.02, \lambda_3 = 0.03, \gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 1 ($\rho = -0.2$)			
γ	-0.5	0.032 (0.273)	0.026 (0.274)
β_0	-0.5	0.021 (0.265)	0.026 (0.252)
β_1	0.5	-0.011 (0.252)	-0.007 (0.253)
σ_u^2	0.5	-0.015 (0.314)	0.026 (0.334)
σ_v^2	0.5	-0.005 (0.305)	0.036 (0.327)
ρ	-0.2	-	0.004 (0.491)
Simulação 2 ($\rho = -0.4$)			
γ	-0.5	0.038 (0.288)	0.031 (0.291)
β_0	-0.5	0.026 (0.256)	0.044 (0.236)
β_1	0.5	-0.024 (0.256)	-0.017 (0.258)
σ_u^2	0.5	-0.064 (0.296)	0.017 (0.334)
σ_v^2	0.5	-0.053 (0.282)	0.006 (0.312)
ρ	-0.4	-	0.002 (0.471)

continua na próxima página.

continua na próxima página.

Tabela 3.4: Vício e Desvio Padrão estimado dos estimadores REML no modelo de Chen com efeitos aleatórios independentes e dependentes: caso sobrevivência Exponencial Segmentada ($L = 3$).

Parâmetros	Valor verdadeiro	E.a. independentes	E.a. dependentes
		Vício Médio (SE)	Vício Médio (SE)
(a) Grupo II ($\lambda_1 = 0.01, \lambda_2 = 0.02, \lambda_3 = 0.03, \gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 7 ($\rho = -0.6$)			
γ	-0.5	0.043 (0.270)	0.038 (0.273)
β_0	-0.5	0.036 (0.254)	0.051 (0.226)
β_1	0.5	-0.029 (0.251)	-0.021 (0.252)
σ_u^2	0.5	-0.035 (0.304)	0.062 (0.348)
σ_v^2	0.5	-0.064 (0.276)	0.004 (0.309)
ρ	-0.6	-	0.048 (0.406)
Simulação 8 ($\rho = -0.8$)			
γ	-0.5	0.026 (0.267)	0.016 (0.271)
β_0	-0.5	0.009 (0.248)	0.035 (0.207)
β_1	0.5	-0.011 (0.250)	-0.001 (0.252)
σ_u^2	0.5	-0.093 (0.277)	0.039 (0.337)
σ_v^2	0.5	-0.095 (0.261)	-0.004 (0.304)
ρ	-0.8	-	0.047 (0.321)
(a) Grupo III ($\lambda_1 = 0.01, \lambda_2 = 0.02, \lambda_3 = 0.03, \gamma = -1.0, \beta_0 = -1.0, \beta_1 = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$)			
Simulação 9 ($\rho = -0.2$)			
γ	-1.0	0.029 (0.284)	0.021 (0.286)
β_0	-1.0	0.030 (0.277)	0.034 (0.265)
β_1	1.0	-0.012 (0.271)	-0.005 (0.272)
σ_u^2	0.5	0.012 (0.344)	0.060 (0.367)
σ_v^2	0.5	-0.043 (0.300)	-0.010 (0.318)
ρ	-0.2	-	-0.041 (0.545)
Simulação 10 ($\rho = -0.4$)			
γ	-1.0	0.040 (0.282)	0.026 (0.285)
β_0	-1.0	0.032 (0.278)	0.043 (0.262)
β_1	1.0	-0.010 (0.270)	0.002 (0.271)
σ_u^2	0.5	-0.014 (0.333)	0.055 (0.364)
σ_v^2	0.5	-0.030 (0.306)	0.013 (0.328)
ρ	-0.4	-	0.055 (0.528)
Simulação 11 ($\rho = -0.6$)			
γ	-1.0	0.012 (0.278)	0.039 (0.283)
β_0	-1.0	0.024 (0.273)	0.037 (0.251)
β_1	1.0	0.002 (0.270)	-0.015 (0.272)
σ_u^2	0.5	-0.074 (0.304)	0.039 (0.346)
σ_v^2	0.5	-0.064 (0.289)	-0.010 (0.321)
ρ	-0.6	-	0.072 (0.468)
Simulação 12 ($\rho = -0.8$)			
γ	-1.0	0.037 (0.278)	0.017 (0.283)
β_0	-1.0	0.007 (0.269)	0.010 (0.237)
β_1	1.0	-0.001 (0.271)	0.019 (0.273)
σ_u^2	0.5	-0.071 (0.305)	0.046 (0.358)
σ_v^2	0.5	-0.088 (0.279)	-0.010 (0.316)
ρ	-0.4	-	0.065 (0.463)

Tabela 3.4 - continuação.

(a) Simulação 1

(b) Simulação 2

(c) Simulação 3

(d) Simulação 4

Figura 3.12: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo I

Figura 3.13: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo II

(a) Simulação 9

(b) Simulação 10

(c) Simulação 11

(d) Simulação 12

Figura 3.14: Histogramas das estimativas dos parâmetros na comparação dos modelos de Chen com efeitos aleatórios independentes e dependentes com sobrevivência basal Exponencial Segmentada ($L = 3$): Grupo III

3.3 Modelo de mistura com outras ligações na cura

Note que no modelo de [Chen et al. \[1999a\]](#), a ligação em (3.21) para relacionar as covariáveis $\mathbf{Z}$ e os efeitos aleatórios das clínicas V_j , para $j = 1, \dots, J$ com o parâmetro associado à cura é uma ligação assimétrica. Entretanto no modelo de mistura, a ligação logito usada por [Yau e Ng \[2001\]](#) e [Li et al. \[2001\]](#) é simétrica (veja a equação (3.1)). Esta ligação é muito conveniente pois permite a interpretação das covariáveis usadas na fração de cura em termos do aumento (ou diminuição) no risco do paciente.

Porém, a ligação logito é uma ligação simétrica, suposição que na prática pode não ser adequada. Por exemplo, suponha que o efeito da j -ésima clínica sobre a fração de cura é nulo, i.e., $V_j = 0$ e que a chance de cura depende de um coeficiente padronizado z_{ij} que reflete o grau de gravidade da doença em questão. Nessas condições, a ligação logito obriga que, se a chance de cura de um indivíduo com coeficiente z_{ij} é π_{ij} , então a chance de cura de indivíduo com coeficiente $-z_{ij}$ será $1 - \pi_{ij}$. Mas, para valores extremos de z_{ij} , essa relação pode não ser adequada.

Com a finalidade de testar outras ligações no modelo de mistura para modelar a fração de cura, as covariáveis z_{ij} e o efeito aleatório V_j da j -ésima clínica serão incluídos no modelo através de

$$\pi_{ij} = \psi(\mathbf{z}_{ij}^\top \boldsymbol{\beta} + V_j), \quad j = 1, \dots, J; i = 1, \dots, n_j, \quad (3.26)$$

em que $\psi(\cdot)$ corresponde a alguma das funções de ligação da Tabela 3.5.

Ligação	$\psi(u)$
Logito	$e^u / (1 + e^u)$
Probit	$\Phi(u)$
Complementar log-log	$1 - \exp\{-\exp(u)\}$

Tabela 3.5: *Ligações para modelar a proporção de não curados*

Os demais aspectos do modelo serão como no modelo de [Li et al. \[2001\]](#) apresentado anteriormente, com a exceção de utilizar-se uma forma paramétrica para $S_0(t | \boldsymbol{\lambda})$, especificamente as distribuições Weibull e exponencial segmentada, com a finalidade de poder realizar uma aplicação unificada do modelo de [Berkson e Gage \[1952\]](#) e do modelo de [Chen et al. \[1999a\]](#) com efeitos aleatórios. O procedimento de estimação é similar ao descrito na seção 3.1, mas como $S_0(t | \boldsymbol{\lambda})$ agora está sendo assumida com forma paramétrica, então deve-se adicionar um passo para estimar $\boldsymbol{\lambda}$. Com $\boldsymbol{\zeta}$, $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ fixados no passo anterior, pode-se calcular a esperança do vetor latente $\mathbf{M}$ para o passo E do algoritmo EM como segue:

$$E\left(M_{ij}^{(\ell+1)} \mid \boldsymbol{\zeta}^{(\ell)}, \boldsymbol{\lambda}^{(\ell)}\right) = \delta_{ij} + (1 - \delta_{ij}) \frac{\pi_{ij}^{(\ell)} S_0(t_{ij} \mid \boldsymbol{\lambda}^{(\ell)})^{\exp \eta_{ij}^{(\ell)}}}{1 - \pi_{ij}^{(\ell)} + \pi_{ij}^{(\ell)} S_0(t_{ij} \mid \boldsymbol{\lambda}^{(\ell)})^{\exp \eta_{ij}^{(\ell)}}}, \quad \ell = 0, 1, 2, \dots, \quad (3.27)$$

em que $\eta_{ij}^{(\ell)}$ pode ser calculado com os valores de $\boldsymbol{\gamma}^{(\ell)}$ e $\mathbf{U}^{(\ell)}$ em (3.2), $\pi_{ij}^{(\ell)}$ pode ser calculado com os valores de $\boldsymbol{\beta}^{(\ell)}$ e $\mathbf{V}^{(\ell)}$ em (3.26) dependendo da ligação utilizada na Tabela 3.5 e $S_0(t \mid \boldsymbol{\lambda}^{(\ell)})$ é a função de sobrevivência da distribuição Weibull em (2.2) ou a função de sobrevivência da distribuição exponencial segmentada em (2.4).

Com $\mathbf{M}$ atualizado em (3.27) e $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ do passo anterior, atualiza-se $\boldsymbol{\zeta}$ utilizando o algoritmo de Newton-Raphson como na equação (3.16). Similarmente, a matriz $\boldsymbol{\Sigma}$ é atualizada conforme as equações em (3.18). Para $\mathbf{M}$, $\boldsymbol{\zeta}$ e $\boldsymbol{\Sigma}$ já atualizados, constrói-se a função de verossimilhança perfilada para l_p^a em (3.12) e encontra-se o valor $\hat{\boldsymbol{\lambda}}$ que maximize l_p^a em relação a $\boldsymbol{\lambda}$ usando o algoritmo de

Newton-Raphson (usando λ_0 como um valor inicial) como segue

$$\lambda_\ell = \lambda_{\ell-1} + C_{\ell-1}^{-1} \left(\frac{\partial l_p}{\partial \lambda} \right) \bigg|_{\lambda=\lambda_{\ell-1}}, \quad \ell = 1, 2, \dots, \quad (3.28)$$

sendo C o negativo da segunda derivada de l_p^a em relação a λ .

Dessa forma, o processo de estimação consiste em fixar valores iniciais para ζ , Σ e λ e iterar entre os processos descritos em (3.27), (3.16), (3.18) e (3.28) até obter a convergência.

A variância de $(\hat{\gamma}, \hat{\beta})^\top$ pode ser estimada usando (3.6), enquanto que a variância do vetor $\hat{\tau}$ pode ser estimada usando (3.19). Finalmente, pelos argumentos usados na seção 3.1, a variância de $\hat{\lambda}$ pode ser estimada usando Jackknife [Miller [1974]].

Note que na Tabela 3.5 poderia ter sido utilizada uma classe mais ampla de ligações para a equação (3.26). Por exemplo, Chen et al. [1999b] discutem a classe de ligações

$$\pi_{ij} = \int_{-\infty}^{\infty} \psi(\mathbf{z}_{ij}^\top \beta + V_j + \varphi s) g(s) ds, \quad (3.29)$$

em que $\psi(\cdot)$ é a f.d.a. de alguma distribuição, $g(\cdot)$ é a função de densidade de alguma distribuição (não necessita coincidir com a densidade associada a ψ) e φ é o parâmetro de assimetria. Note que se $\varphi = 0$ ou se $g(\cdot)$ corresponde à função de probabilidade de uma v.a. degenerada, então a equação (3.29) fica exatamente igual a (3.26), recuperando-se os casos particulares das ligações discutidas anteriormente. Caso contrário, a expressão em (3.29) gera uma classe de distribuições assimétricas que incluem, entre outras, a ligação probito assimétrica (quando $\psi(\cdot)$ é a f.d.a. distribuição normal e $g(\cdot)$ a densidade da distribuição half-normal) e logito assimétrica (quando $\psi(\cdot)$ é a acumulada da distribuição logística e $g(\cdot)$ a densidade da distribuição logística), que corresponde às f.d.a. das distribuições Skew-Normal (Azzalini [1986]) e skew-logística (Nadarajah [2009]) respectivamente.

Porém, no contexto que se está trabalhando aqui o método de estimação já é complexo e computacionalmente custoso, por conta de dividir o procedimento de estimação nos quatro passos descritos anteriormente. Desta forma, incluir o cálculo de π_{ij} mediante uma integral não fechada (com exceção de alguns casos particulares) como em (3.29) para cada observação pode elevar consideravelmente o tempo de execução quando implementado o modelo computacionalmente.

Desta forma, esta seção tem apenas o espírito de ilustrar o que pode acontecer quando é utilizada só uma ligação simétrica como é a ligação logito, sobretudo quando o objetivo do estudo é prever, seja a chance de cura do paciente ou a probabilidade de sobrevida associada a ele.

3.3.1 Estudo de simulação: Recuperação de parâmetros

A seguir será apresentado um breve estudo de simulação feito para o modelo de mistura com efeitos aleatórios descrito na seção 3.3, mas só para a ligação complementar log-log da Tabela 3.5 que é a única ligação assimétrica proposta.

A simulação dos dados é similar à explicada na seção 3.2.1. Assuma que os pacientes estão divididos em J clínicas, em que a j -ésima clínica tem n_j pacientes. Assuma também que o i -ésimo indivíduo da j -ésima clínica tem covariáveis $\mathbf{x}_{ij}$ e $\mathbf{z}_{ij}$ de dimensão r e s respectivamente para explicar a sobrevivência dos indivíduos suscetíveis e a fração de cura. Considerando Σ como em (3.11), foram simulados os pares de efeitos aleatórios $(U_1, V_1), \dots, (U_J, V_J)$ da distribuição $N(0, \Sigma)$. η_{ij} e π_{ij} para $j = 1, \dots, J$; $i = 1, \dots, n_j$ foram calculados baseados em (3.2) e (3.26) respectivamente, em que γ e β são vetores conhecidos de tamanho r e s respectivamente. Cada M_{ij} foi simulado da

distribuição Bernoulli com sua respectiva probabilidade de sucesso π_{ij} . Se M_{ij} é zero, então tempo de falha do i -ésimo indivíduo da j -ésima clínica, T_{ij}^* , é definido como ∞ . Se $M_{ij} = 1$, então T_{ij}^* é simulado da densidade $h_0(t | \boldsymbol{\lambda}) \exp \eta_{ij} S_0(t | \boldsymbol{\lambda})^{\exp \eta_{ij}}$, em que $S_0(t | \boldsymbol{\lambda})$ provém da distribuição Weibull com parâmetros $\boldsymbol{\lambda} = (\lambda, \nu)$ como em (2.2). (Neste caso só foi utilizada esta distribuição). Finalmente, para um tempo de censura C conhecido, defina as variáveis $T_{ij} = \min(T_{ij}^*, C)$ e $\delta_{ij} = I(T_{ij}^* \leq C)$, que representam os tempos observados (de falha ou censura) e o indicador de falha para cada indivíduo.

O estudo foi feito considerando

- $J = 10$ clínicas;
- $n_j = 15$, $j = 1, \dots, J$ pacientes em cada clínica;
- em cada clínica, 5 pacientes no grupo de tratamento ($x_{ij} = 1$) e 10 pacientes no grupo de controle ($x_{ij} = 0$);
- as covariáveis para explicar a fração de cura como $\mathbf{z}_{ij} = (1, x_{ij})$;
- $\boldsymbol{\gamma} = -0.5$, $\boldsymbol{\beta} = (0.5, -0.5)^\top$, $\{\lambda = 0.01, \nu = 1.0, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000\}$ para todos os casos, só mudando o valor de ρ no conjunto $\{0.2, 0.4, 0.6, 0.8\}$.

Para cada grupo de parâmetros, foram feitas 500 réplicas, as que são apresentadas na Tabela 3.6 e nas Figuras 3.1, 3.2, 3.3 e 3.4. Nas Figuras, a curva em azul representa a distribuição normal com média e desvio padrão de acordo aos resultados obtidos para cada parâmetro e a linha em vermelho representa a respectiva média amostral. Nas tabelas, SE_1 e SE_2 representam a média dos desvios padrões e o desvio padrão das estimativas obtidas nas 500 simulações.

Como mostrado na seção 3.2.2, espera-se que o aumento de ρ melhore o desempenho dos estimadores em geral, pois a inclusão deste parâmetro no modelo é justificada pela existência da dependência entre os efeitos aleatórios de uma mesma clínica.

Em relação ao estimador de γ , percebe-se que o vício tende a aumentar quando ρ aumenta e a variância do estimador é levemente sobrestimada. Para $\hat{\beta}_0$ e $\hat{\beta}_1$, o vício é menor para todos os casos considerados, mas a variância de ambos também é sobrestimada de forma leve.

Por outro lado, $\hat{\sigma}_u^2$ e $\hat{\sigma}_v^2$ apresentam vícios aceitáveis, os quais melhoram a medida que a correlação ρ é mais forte. Para os casos considerados, a variância destes estimadores é levemente sobrestimada.

Finalmente, o desempenho de $\hat{\rho}$ também é bom em relação ao vício médio, embora a variância deste estimador seja sobrestimada em maior medida em relação aos outros parâmetros, mas tem uma pequena melhoria quando ρ aumenta.

Baseado neste estudo de simulação, podemos concluir que, pelo menos para os casos considerados, o desempenho dos estimadores propostos foi aceitável em relação ao vício e a sua variância estimada, o qual mostrou melhoria à medida que a correlação ρ era mais forte.

Parâmetro	Valor verdadeiro	Vício médio	SE_1	SE_2
Simulação 1 ($\rho = 0.2$)				
γ	-0.5	0.052	0.276	0.219
β_0	0.5	-0.044	0.269	0.225
β_1	-0.5	0.001	0.298	0.253
σ_u^2	0.5	-0.066	0.307	0.245
σ_v^2	0.5	-0.058	0.295	0.273
ρ	0.2	0.055	0.459	0.390
Simulação 2 ($\rho = 0.4$)				
γ	-0.5	0.033	0.287	0.217
β_0	0.5	-0.035	0.267	0.223
β_1	-0.5	0.028	0.294	0.252
σ_u^2	0.5	-0.044	0.314	0.254
σ_v^2	0.5	-0.027	0.309	0.288
ρ	0.4	0.051	0.426	0.356
Simulação 3 ($\rho = 0.6$)				
γ	-0.5	0.013	0.280	0.217
β_0	0.5	-0.039	0.272	0.213
β_1	-0.5	0.027	0.296	0.249
σ_u^2	0.5	-0.055	0.281	0.249
σ_v^2	0.5	-0.031	0.336	0.286
ρ	0.6	0.045	0.357	0.312
Simulação 4 ($\rho = 0.8$)				
γ	-0.5	0.013	0.278	0.213
β_0	0.5	-0.045	0.256	0.209
β_1	-0.5	0.009	0.288	0.248
σ_u^2	0.5	-0.033	0.283	0.260
σ_v^2	0.5	0.031	0.364	0.317
ρ	0.8	0.014	0.246	0.247

Tabela 3.6: Vício estimado para os estimadores REML baseados em 500 réplicas (Função de sobrevivência dos não curados considerada do modelo Weibull).

Figura 3.15: Histograma dos parâmetros estimados no modelo de mistura com efeitos aleatórios usando ligação complementar log-log para a cura e sobrevivência basal Weibull

3.3.2 Estimação da fração de cura usando as diferentes ligações

Como visto no estudo de simulação anterior, o processo de estimação parece ser bem comportado, inclusive quando a ligação usada é assimétrica na cura (Li et al. [2001] e Lai e Yau [2008] apresentam estudos de simulação utilizando a ligação logito na cura). Porém, fica a dúvida em relação a como pode influenciar nas estimativas uma escolha errada desta ligação. Para responder esta questão, fez-se um estudo de simulação. Mas o vetor β e os efeitos aleatórios V_j , para $j = 1, \dots, J$ não são comparáveis usando as três ligações da Tabela 3.5. Por esta razão, para fazer uma comparação unificada, ao invés de comparar as estimativas dos parâmetros usando cada ligação, foram comparadas as frações de cura estimadas. Esse será o único foco do estudo.

A simulação dos dados é feita de forma similar à seção 3.3.1 considerando

- em cada clínica, $1/3$ dos pacientes no grupo de tratamento ($x_{ij} = 1$) e $2/3$ dos pacientes no grupo de controle ($x_{ij} = 0$);
- as covariáveis para explicar a fração de cura como $z_{ij} = (1, z_{ij})$, em que z_{ij} foi simulado da distribuição $N(0, 1)I(-3, 3)$, i.e., da distribuição Normal padrão truncada no intervalo $(-3, 3)$;
- em cada caso, foram substituídos dois valores de z_{ij} como -3 e 2 valores como 3 em clínicas e tratamentos ao acaso, com a finalidade de garantir ter valores “extremos” observados na covariável;
- $\gamma = -0.5$, $\beta = (-0.5, 0.5)^\top$ e $C = 300$,
- para cada combinação dos parâmetros anteriores, considerou-se como caso basal $\sigma_u^2 = 0.5$, $\sigma_v^2 = 0.5$, $\rho = 0.5$, depois aumentou-se as variâncias dos efeitos aleatórios para $\sigma_u^2 = \sigma_v^2 = 1.0$ e posteriormente aumentou-se a correlação ρ para 0.8, ambas variações em relação ao caso basal,
- só foi considerado para $S_0(t | \lambda)$ a distribuição Weibull em (2.2) com $\lambda = 0.01$ e $\nu = 1.0$,
- para cada uma das especificações anteriores, foi considerado como caso basal $J = 5$ clínicas com $n_j = 6$ pacientes cada clínica e considerando depois aumentar o número de clínicas para $J = 10$ em um primeiro caso, e aumentar o número de pacientes para $n_j = 12$ em um segundo caso.

Em cada caso, foram feitas 500 réplicas. Resultados deste estudo estão na Tabela 3.7. Para as frações de cura estimadas no modelo com as três ligações, foi calculado o vício médio e o a média do vício máximo da fração de cura produzido nas 500 réplicas, com seu respectivo desvio padrão.

Em termos gerais, a escolha de uma ligação errada pode incrementar levemente o vício médio das estimativas nas frações de cura, mas o maior problema é que o vício máximo das frações de cura estimadas pode aumentar consideravelmente. Isto pode constituir um problema se o alvo do estudo é predizer, seja tanto a fração de cura em si ou o tempo de sobrevivência de um paciente com determinadas características, sobretudo quando essas características do paciente são valores extremos das covariáveis observadas. Este último fato será exemplificado no capítulo da aplicação.

Finalmente, um aumento do número de pacientes em cada clínica diminuiu o vício máximo das frações de cura estimadas, produzindo que estas estimativas sejam mais acuradas e um aumento do número de clínicas só produziu estimativas mais acuradas.

Conjunto de parâmetros: $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300$.									
Ligação usada para simular os dados: logito									
$J = 5, n_j = 6$									
$J = 5, n_j = 12$									
Ligação na cura	caso I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.004 (0.099)	0.244 (0.161)	-0.005 (0.071)	0.200 (0.120)	-0.001 (0.070)	0.246 (0.123)	-0.001 (0.070)	0.246 (0.123)
	Probit	-0.005 (0.095)	0.250 (0.159)	-0.012 (0.070)	0.209 (0.121)	-0.011 (0.070)	0.269 (0.124)	-0.011 (0.070)	0.269 (0.124)
Ligação na cura	caso II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	-0.010 (0.094)	0.268 (0.166)	-0.008 (0.070)	0.218 (0.134)	-0.010 (0.066)	0.274 (0.131)	-0.010 (0.066)	0.274 (0.131)
	Probit	-0.023 (0.091)	0.274 (0.164)	-0.018 (0.065)	0.222 (0.131)	-0.024 (0.066)	0.296 (0.134)	-0.024 (0.066)	0.296 (0.134)
Ligação na cura	caso III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.006 (0.093)	0.237 (0.146)	0.005 (0.064)	0.200 (0.124)	0.001 (0.063)	0.230 (0.115)	0.001 (0.063)	0.230 (0.115)
	Probit	-0.004 (0.093)	0.243 (0.146)	-0.004 (0.063)	0.206 (0.124)	-0.009 (0.062)	0.244 (0.118)	-0.009 (0.062)	0.244 (0.118)
Conjunto de parâmetros: $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300$.									
$J = 5, n_j = 6$									
$J = 5, n_j = 12$									
Ligação na cura	caso I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.007 (0.088)	0.286 (0.175)	0.002 (0.058)	0.238 (0.134)	0.003 (0.060)	0.298 (0.132)	0.003 (0.060)	0.298 (0.132)
	Probit	-0.005 (0.086)	0.257 (0.161)	-0.008 (0.056)	0.213 (0.123)	-0.010 (0.059)	0.275 (0.126)	-0.010 (0.059)	0.275 (0.126)
Ligação na cura	caso II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.008 (0.088)	0.311 (0.187)	0.002 (0.067)	0.264 (0.161)	0.001 (0.063)	0.326 (0.163)	0.001 (0.063)	0.326 (0.163)
	Probit	-0.010 (0.082)	0.277 (0.172)	-0.014 (0.060)	0.230 (0.144)	-0.016 (0.061)	0.296 (0.152)	-0.016 (0.061)	0.296 (0.152)
Ligação na cura	caso III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.015 (0.085)	0.310 (0.181)	0.005 (0.061)	0.259 (0.154)	0.011 (0.062)	0.322 (0.160)	0.011 (0.062)	0.322 (0.160)
	Cloglog								
Ligação na cura	caso I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.014 (0.086)	0.283 (0.166)	0.009 (0.066)	0.240 (0.138)	0.008 (0.061)	0.302 (0.136)	0.008 (0.061)	0.302 (0.136)
	Probit	0.001 (0.081)	0.263 (0.154)	-0.003 (0.063)	0.206 (0.124)	-0.005 (0.060)	0.272 (0.133)	-0.005 (0.060)	0.272 (0.133)
Ligação na cura	caso II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.021 (0.083)	0.288 (0.159)	0.013 (0.064)	0.236 (0.134)	0.018 (0.063)	0.302 (0.142)	0.018 (0.063)	0.302 (0.142)
	Cloglog								
Conjunto de parâmetros: $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300$.									
$J = 5, n_j = 6$									
$J = 5, n_j = 12$									
Ligação na cura	caso I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	-0.005 (0.100)	0.263 (0.163)	-0.004 (0.069)	0.229 (0.137)	-0.008 (0.068)	0.290 (0.125)	-0.008 (0.068)	0.290 (0.125)
	Probit	-0.015 (0.096)	0.277 (0.156)	-0.014 (0.066)	0.214 (0.129)	-0.021 (0.065)	0.284 (0.118)	-0.021 (0.065)	0.284 (0.118)
Ligação na cura	caso II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.005 (0.103)	0.309 (0.191)	-0.009 (0.072)	0.259 (0.148)	-0.009 (0.072)	0.319 (0.143)	-0.009 (0.072)	0.319 (0.143)
	Probit	-0.017 (0.099)	0.314 (0.183)	-0.024 (0.069)	0.241 (0.137)	-0.031 (0.064)	0.303 (0.131)	-0.031 (0.064)	0.303 (0.131)
Ligação na cura	caso III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.013 (0.099)	0.256 (0.160)	0.002 (0.064)	0.283 (0.126)	0.000 (0.072)	0.273 (0.124)	0.000 (0.072)	0.273 (0.124)
	Probit	0.019 (0.099)	0.257 (0.161)	-0.012 (0.062)	0.271 (0.117)	-0.012 (0.071)	0.272 (0.121)	-0.012 (0.071)	0.272 (0.121)
Ligação na cura	caso I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$.								
	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício médio (dp)	Vício Máx. (dp)	Vício Máx. (dp)
	Logito	0.002 (0.098)	0.241 (0.164)	0.008 (0.062)	0.257 (0.123)	0.007 (0.070)	0.265 (0.126)	0.007 (0.070)	0.265 (0.126)
	Cloglog								

Tabela 3.7: Vício estimado e vício máximo estimado para a fração de cura usando as 3 ligações propostas.

Capítulo 4

Modelos destrutivos de Poisson ponderado com fração de cura e efeitos aleatórios

Os modelos destrutivos de Poisson ponderado com fração de cura (MDPPFC, [Rodrigues et al. \[2011\]](#)) foram propostos recentemente. Portanto, não há uma proposta na literatura destes modelos que incluíam efeitos aleatórios.

Por outro lado, no contexto de modelos de sobrevivência com fração de cura e efeitos aleatórios, o uso extensivo da distribuição normal para a modelagem dos efeitos aleatórios explica-se devido à facilidade da implementação computacional do modelo sob um ponto de vista clássico, pois podem ser utilizados os estimadores REML ([McGilchrist e Yau \[1995\]](#)), os que também tem a vantagem de possuir propriedades desejáveis, como ser não viciados por exemplo. Embora isso, a suposição de normalidade para os efeitos aleatórios pode não ser adequada. Por exemplo, podem ter sido omitidas características relevantes das clínicas em relação ao efeito delas sobre os pacientes no seu cuidado. Desta forma, a suposição de simetria dos efeitos aleatórios poderia ser afetada. Melhor ainda, pode-se assumir uma abordagem através de P.D.'s ([Ferguson \[1973\]](#), seção 2.4). Nesta seção são apresentadas uma abordagem clássica e bayesiana para incorporar efeitos aleatórios nos MDPPFC.

4.1 O modelo

Similarmente ao contexto do capítulo anterior, suponha que foram estudados casos em J clínicas e que para a j -ésima clínica havia n_j indivíduos, $j = 1, \dots, J$. Para $i = 1, \dots, n_j$, considere M_{ij} e D_{ij} como o número de células suscetíveis de serem ativadas e o número de células que efetivamente podem produzir câncer que o i -ésimo indivíduo da j -ésima clínica possui. Suponha ainda que os M_{ij} 's são independentes e com distribuição de Poisson ponderada com parâmetro θ_{ij} e função de pesos $w(\cdot \mid \phi)$ como em (2.20) segundo cada modelo usado (Tabela 4.1), além de supor que $D_{ij} \mid M_{ij} = m_{ij} \sim \text{Bin}(m_{ij}, p_{ij})$ como discutido na seção 2.2.4. Defina ainda Y_{cij} a variável aleatória que representa o tempo até a c -ésima célula carcinogênica ativada do i -ésimo indivíduo da j -ésima clínica produza câncer. Suponha que as variáveis Y_{cij} são condicionalmente i.i.d. dado D_{ij} com f.d.a. comum $F(t \mid \boldsymbol{\lambda}, \eta_{ij}) = 1 - S(t \mid \boldsymbol{\lambda}, \eta_{ij})$ e independentes de D_{ij} . O tempo para o aparecimento do câncer pode ser definido pela variável aleatória $T_{ij}^* = \min(Y_{cij}, 1 \leq c \leq D_{ij})$, mas $T_{ij}^* = \infty$ com probabilidade 1 se $D_{ij} = 0$. Estes tempos são observados sujeitos a uma variável C_{ij} de censura, implicando que os valores observados (sob censura à direita) são dados por $T_{ij} = \min(T_{ij}^*, C_{ij})$ e $\delta_{ij} = I(T_{ij}^* \leq C_{ij})$.

Modelo	DLBP	EWPM	DBN
$w(m \mid \phi)$	m	$e^{\phi m}$	$\Gamma(\phi^{-1} + m)$

Tabela 4.1: Funções de peso para cada modelo Destrutivo.

Covariáveis $\mathbf{X}$ para modelar a heterogeneidade dos indivíduos suscetíveis podem ser incorpo-

radas conjuntamente com um efeito aleatório U_j associado à clínica a que pertence o indivíduo, através da função $S(t \mid \boldsymbol{\lambda}, \eta_{ij})$, que neste caso representa a função de sobrevivência das células que não foram destruídas, como em (3.2) fazendo

$$S(t \mid \boldsymbol{\lambda}, \eta_{ij}) = [S_0(t \mid \boldsymbol{\lambda})]^{\exp(\eta_{ij})}, \quad i = 1, \dots, n, j = 1, \dots, J, \quad (4.1)$$

em que $\eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\gamma} + U_j$.

Por outro lado, também pode-se incluir um vetor de covariáveis $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, em que $\mathbf{Z}_1$ está associado com o número de células iniciais do indivíduo e $\mathbf{Z}_2$ com a probabilidade de ativação de cada uma dessas células, em que $\mathbf{Z}_1$ e $\mathbf{Z}_2$ não podem ter elementos em comum. Um efeito aleatório V_j , $j = 1, \dots, J$ associado ao efeito da j -ésima clínica na fração de cura pode ser incluído através do parâmetro θ_{ij} fazendo

$$\log(\theta_{ij}) = \vartheta_{1ij} = \mathbf{z}_{1ij}^\top \boldsymbol{\beta}_1 + V_j \quad \text{e} \quad \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \vartheta_{2ij} = \mathbf{z}_{2ij}^\top \boldsymbol{\beta}_2, \quad 1 \leq i \leq n_j, 1 \leq j \leq J, \quad (4.2)$$

ou bem, através do parâmetro p_{ij} fazendo

$$\log(\theta_{ij}) = \vartheta_{1ij} = \mathbf{z}_{1ij}^\top \boldsymbol{\beta}_1 \quad \text{e} \quad \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \vartheta_{2ij} = \mathbf{z}_{2ij}^\top \boldsymbol{\beta}_2 + V_j, \quad 1 \leq i \leq n_j, 1 \leq j \leq J. \quad (4.3)$$

Note que não é possível pensar em incluir dois efeitos aleatórios V_{1j} e V_{2j} associados à cura e introduzidos através de θ_{ij} e p_{ij} respectivamente, pois isso produz problemas de identificabilidade. Isso é fácil de visualizar pensando em introduzir primeiro esses dois efeitos como efeitos fixos. Note que se isso for feito, então as matrizes de desenho $\mathbf{Z}_1$ e $\mathbf{Z}_2$ teriam como elemento em comum a coluna que indica à clínica que pertence o paciente. Mas como discutido no final da seção 2.2.4, o fato de $\mathbf{Z}_1$ e $\mathbf{Z}_2$ terem elementos em comum cria problemas de identificabilidade.

Por outro lado e como mencionado na seção 2.2.4, é possível trabalhar com a verossimilhança observada dos modelos destrutivos (vide equação (2.23)) somando a verossimilhança sobre os possíveis valores do vetor $\mathbf{M}$ e $\mathbf{D}$. Desta forma, os dados observados são $\mathbf{D}_{obs} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2)$, em que $\mathbf{t} = (t_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$, $\boldsymbol{\delta} = (\delta_{ij})_{i=1, \dots, n_j}^{j=1, \dots, J}$ e $\mathbf{X}, \mathbf{Z}_1$ e $\mathbf{Z}_2$ são as matrizes de covariáveis para a modelagem da função de sobrevivência dos indivíduos suscetíveis, células iniciais dos indivíduos e probabilidade de ativação de cada célula respectivamente. Os dados completos são $\mathbf{D}_c = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{K}, \mathbf{X}, \mathbf{Z})$ que incluem os dados observados e os efeitos aleatórios $\mathbf{K} = (\mathbf{U}, \mathbf{V})$, em que $\mathbf{U} = (U_1, \dots, U_J)$ e $\mathbf{V} = (V_1, \dots, V_J)$.

4.2 Abordagem clássica

De um ponto de vista clássico, pode-se utilizar os estimadores REML (McGilchrist e Yau [1995]), como visto extensivamente através do capítulo 3. Assim, assume-se que $(U_1, V_1), \dots, (U_J, V_J)$ são i.i.d. com distribuição idêntica $N(0, \boldsymbol{\Sigma})$, em que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix},$$

em que segundo esta parametrização, $\text{Cor}(U_j, V_j) = \rho$, para $j = 1, \dots, J$. Desta forma, para $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ fixos, constrói-se a função de log-verossimilhança BLUP do modelo, a qual é dada por $l = l_1 + l_2$, em que

$$l_1 = \sum_{j=1}^J \sum_{i=1}^{n_j} \left[\delta_{ij} \log h_{pop}(t_{ij} | \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij}) + \log S_{pop}(t_{ij} | \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij}) \right],$$

$$l_2 = -\frac{1}{2(1-\rho^2)} \left[\frac{\mathbf{U}^\top \mathbf{U}}{\sigma_u^2} + \frac{\mathbf{V}^\top \mathbf{V}}{\sigma_v^2} - \frac{2\rho \mathbf{U}^\top \mathbf{V}}{\sigma_u \sigma_v} \right] - (J/2) \log \{4\pi^2 \sigma_u^2 \sigma_v^2 (1-\rho^2)\}, \quad (4.4)$$

em que $S(t_{ij} | \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij})$ e $h(t_{ij} | \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij})$ são dados na Tabela 4.2. Nesta mesma tabela, $F(t | \boldsymbol{\lambda}, \eta_{ij}) = 1 - S(t | \boldsymbol{\lambda}, \eta_{ij})$, $f(t | \boldsymbol{\lambda}, \eta_{ij}) = (\partial/\partial t)F(t | \boldsymbol{\lambda}, \eta_{ij})$ e $S(t | \boldsymbol{\lambda}, \eta_{ij})$ e η_{ij} são dadas em (4.1). θ_{ij} e p_{ij} são dados em (4.2) ou em (4.3) dependendo de qual parâmetro são incluídos os efeitos aleatórios associados à cura.

Modelo	$S_{pop}(t \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij})$	$h_{pop}(t \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij})$
DLBP	$e^{-\theta_{ij} p_{ij} F(t \boldsymbol{\lambda}, \eta_{ij})} (1 - p_{ij} F(t \boldsymbol{\lambda}, \eta_{ij}))$	$p_{ij} f(t \boldsymbol{\lambda}, \eta_{ij}) \left\{ \theta_{ij} + \frac{1}{1 - p_{ij} F(t \boldsymbol{\lambda}, \eta_{ij})} \right\}$
DEWPM	$\exp\{-\theta_{ij} p_{ij} e^{\phi} F(t \boldsymbol{\lambda}, \eta_{ij})\}$	$\theta_{ij} p_{ij} e^{\phi} f(t \boldsymbol{\lambda}, \eta_{ij})$
DBN	$\left\{ 1 + \phi \theta_{ij} p_{ij} F(t \boldsymbol{\lambda}, \eta_{ij}) \right\}^{-\phi^{-1}}$	$\frac{\theta_{ij} p_{ij} f(t \boldsymbol{\lambda}, \eta_{ij})}{1 + \phi \theta_{ij} p_{ij} F(t \boldsymbol{\lambda}, \eta_{ij})}$

Tabela 4.2: Funções de sobrevivência e risco para os modelos DLBP, EWPM e DBN.

Desta forma, para $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ fixos, o estimador BLUP para $\boldsymbol{\zeta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{U}, \mathbf{V})$ pode ser encontrado maximizando $l = l_1 + l_2$ em (4.4). Se $\boldsymbol{\zeta}_0$ é um valor inicial para $\boldsymbol{\zeta}$, então pode-se usar o método de Newton-Raphson como segue:

$$\boldsymbol{\zeta}_\ell = \boldsymbol{\zeta}_{\ell-1} + \mathbf{B}_{\ell-1}^{-1} \left(\frac{\partial l}{\partial \boldsymbol{\zeta}} \right) \bigg|_{\boldsymbol{\zeta}=\boldsymbol{\zeta}_{\ell-1}}, \quad \ell = 1, 2, \dots, \quad (4.5)$$

em que $\mathbf{B}$ o negativo da segunda derivada de l em relação a $\boldsymbol{\zeta}$. Defina $\hat{\boldsymbol{\zeta}}$ como o estimador obtido em (4.5) e denote a matriz particionada $\mathbf{B}$ e sua inversa, avaliada em $\hat{\boldsymbol{\zeta}}$ conforme à ordenação $\boldsymbol{\gamma} | \boldsymbol{\beta}_1 | \boldsymbol{\beta}_2 | \mathbf{U} | \mathbf{V}$, respectivamente como

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{B}}_{12} & \hat{\mathbf{B}}_{13} & \hat{\mathbf{B}}_{14} & \hat{\mathbf{B}}_{15} \\ \hat{\mathbf{B}}_{21} & \hat{\mathbf{B}}_{22} & \hat{\mathbf{B}}_{23} & \hat{\mathbf{B}}_{24} & \hat{\mathbf{B}}_{25} \\ \hat{\mathbf{B}}_{31} & \hat{\mathbf{B}}_{32} & \hat{\mathbf{B}}_{33} & \hat{\mathbf{B}}_{34} & \hat{\mathbf{B}}_{35} \\ \hat{\mathbf{B}}_{41} & \hat{\mathbf{B}}_{42} & \hat{\mathbf{B}}_{43} & \hat{\mathbf{B}}_{44} & \hat{\mathbf{B}}_{45} \\ \hat{\mathbf{B}}_{51} & \hat{\mathbf{B}}_{52} & \hat{\mathbf{B}}_{53} & \hat{\mathbf{B}}_{54} & \hat{\mathbf{B}}_{55} \end{bmatrix} \quad \text{e} \quad \hat{\mathbf{B}}^{-1} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} & \hat{\mathbf{A}}_{13} & \hat{\mathbf{A}}_{14} & \hat{\mathbf{A}}_{15} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} & \hat{\mathbf{A}}_{23} & \hat{\mathbf{A}}_{24} & \hat{\mathbf{A}}_{25} \\ \hat{\mathbf{A}}_{31} & \hat{\mathbf{A}}_{32} & \hat{\mathbf{A}}_{33} & \hat{\mathbf{A}}_{34} & \hat{\mathbf{A}}_{35} \\ \hat{\mathbf{A}}_{41} & \hat{\mathbf{A}}_{42} & \hat{\mathbf{A}}_{43} & \hat{\mathbf{A}}_{44} & \hat{\mathbf{A}}_{45} \\ \hat{\mathbf{A}}_{51} & \hat{\mathbf{A}}_{52} & \hat{\mathbf{A}}_{53} & \hat{\mathbf{A}}_{54} & \hat{\mathbf{A}}_{55} \end{bmatrix}.$$

Deve-se levar em conta o fato de que a matriz $\mathbf{B}$ deve ser construída baseada em qual parâmetro são incluídos os efeitos aleatórios V_j , $j = 1, \dots, J$ (equações (4.2) e (4.3)). Para mais detalhes, vide o Apêndice A.2.3.

Se $\delta_{ij} = 1, \forall j = 1, \dots, J, \forall i = 1, \dots, n_j$, então usando (2.39) tem-se que a distribuição assintótica

para $(\hat{\gamma}, \hat{\beta}_1, \hat{\beta}_2)^\top$ é dada por

$$\begin{pmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} & \hat{\mathbf{A}}_{13} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} & \hat{\mathbf{A}}_{23} \\ \hat{\mathbf{A}}_{31} & \hat{\mathbf{A}}_{32} & \hat{\mathbf{A}}_{33} \end{pmatrix}^{-1} \left[\begin{pmatrix} \hat{\gamma} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} - \begin{pmatrix} \gamma \\ \beta_1 \\ \beta_2 \end{pmatrix} \right] \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}), \quad n_1 + \dots + n_J \rightarrow \infty. \quad (4.6)$$

Como mencionado anteriormente, quando são observadas censuras não pode ser garantido analiticamente o resultado em (4.6), mas estudos de simulação (não apresentados em este trabalho) mostram que essa convergência é razoável.

Por outro lado, a matriz Σ pode ser estimada usando o mesmo resultado em (3.18), a saber

$$\hat{\sigma}_u^2 = \frac{1}{J} \text{tr}(\hat{\mathbf{A}}_{44} + \hat{\mathbf{U}}\hat{\mathbf{U}}^\top), \quad \hat{\sigma}_v^2 = \frac{1}{J} \text{tr}(\hat{\mathbf{A}}_{55} + \hat{\mathbf{V}}\hat{\mathbf{V}}^\top) \quad \text{e} \quad \hat{\rho} = \frac{\text{tr}(\hat{\mathbf{A}}_{45} + \hat{\mathbf{U}}\hat{\mathbf{V}}^\top)}{\sqrt{\hat{\sigma}_u^2 \times \hat{\sigma}_v^2}}, \quad (4.7)$$

lembrando que como mencionado na seção 3.2, a solução para $\hat{\sigma}_u^2$ e $\hat{\sigma}_v^2$ mantém-se invariante se for assumido $\rho = 0$. A variância do vetor $\hat{\tau} = (\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho})$ é dada pela inversa da matriz $2\mathbf{W}$, em que os elementos de $\mathbf{W}$ são dados em (3.20) se for estimado ρ , e em (3.9) se for assumido que $\rho = 0$.

Finalmente, com ζ e Σ fixos, constrói-se a função de log-verossimilhança perfilada para λ , digamos l_p , e encontra-se o valor de λ que maximize essa verossimilhança ($\hat{\lambda}$). Se λ_0 é um valor inicial para $\hat{\lambda}$, então pode-se usar o algoritmo de Newton-Raphson como segue

$$\lambda_\ell = \lambda_{\ell-1} + \mathbf{C}_{\ell-1}^{-1} \left(\frac{\partial l_p}{\partial \lambda} \right) \bigg|_{\lambda=\lambda_{\ell-1}}, \quad \ell = 1, 2, \dots, \quad (4.8)$$

em que $\mathbf{C}$ é a segunda derivada de l_p em relação a λ , dependendo se for usado o modelo Weibull ou Exponencial Segmentado para $S_0(t \mid \lambda)$. Detalhes de $\partial l_p / \partial \lambda$ e da matriz $\mathbf{C}$ encontram-se no Apêndice A.2.3. A variância do vetor $\hat{\lambda}$ estima-se por Jackknife (Miller [1974]) pelas mesmas razões expostas no final da seção 3.1.

Desta forma, o processo de estimação consiste em iterar entre os procedimentos explicados em (4.5), (4.7) e (4.8) até obter a convergência de todos os estimadores e preditores dos efeitos aleatórios envolvidos.

4.3 Abordagem bayesiana

Nesta seção será assumido que a função $S_0(t \mid \lambda)$ é conhecida. Por exemplo, pode-se assumir o valor de λ estimado usando a abordagem clássica. Isto é feito pois o problema computacional que gera-se quando tenta-se estimar λ junto aos outros parâmetros é independente da abordagem usada, clássica ou bayesiana. Porém, o custo computacional de estimar os demais parâmetros fixando λ e vice-versa até obter a convergência aumenta enormemente de um ponto de vista Bayesiano, sobretudo pelo fato das estimativas bayesianas basear-se em amostras, o que implica que para garantir uma mesma estimativa em um procedimento dado requer-se cadeias de tamanho elevado, a diferença da abordagem clássica em que o máximo encontrado, para um determinado valor inicial, sempre será o mesmo.

Por outro lado, do ponto de vista bayesiano, tem-se a vantagem de poder utilizar os P.D.'s para os efeitos aleatórios, o que permite não ficar limitado a uma certa distribuição *a priori*. Porém, a diferença da abordagem clássica, em que assumir a normalidade dos e.a. garante que $E(\mathbf{K}_j) = 0$ para $j = 1, \dots, J$, deve-se fazer algumas reparametrizações no modelo para garantir essa proprie-

dade. Como em (2.52), pode-se assumir P.D.C.'s para os efeitos fixos que estiverem pareados com os e.a.'s (seção 2.4.2), que pela forma em que foram introduzidos no modelo só correspondem aos interceptos do modelo.

Em relação aos e.a.'s V_j , note que há algumas combinações possíveis em relação ao fato de incorporar intercepto só na matriz $\mathbf{Z}_1$ ou em $\mathbf{Z}_2$. Para simplificar, será assumido que o intercepto é incorporado junto ao e.a., i.e., o termo de intercepto e o e.a. V_j são incorporados através de θ_{ij} como em (4.2) ou ambos são incorporados através de p_{ij} como em (4.3). Desta forma, se os e.a.'s V_j são incorporado através de θ_{ij} , defina

$$V_j^* = \beta_{10} + V_j, \quad j = 1, \dots, J.$$

Pelo contrário, se são incorporados através de p_{ij} , defina

$$V_j^* = \beta_{20} + V_j, \quad j = 1, \dots, J.$$

Desta forma, com as reparametrizações mencionadas anteriormente, o vetor de efeitos aleatórios é definido como $\mathbf{K}^* = (\mathbf{K}_1^*, \dots, \mathbf{K}_J^*)$, em que $\mathbf{K}_j^* = (U_j, V_j^*)$, $j = 1, \dots, J$.

Além disso, defina os vetores β_1^F e β_2^F como os subvetores que contêm só aqueles elementos de β_1 e β_2 que não estão associados ao intercepto (independente da forma em que foram incluídos os e.a.'s V_j no modelo). Analogamente, defina $\mathbf{Z}_1^F$ e $\mathbf{Z}_2^F$ como as submatrizes que contêm só aquelas colunas de $\mathbf{Z}_1$ e $\mathbf{Z}_2$ que não estão associadas ao intercepto. Desta forma, as equações (4.2) e (4.3) ficam nas formas

$$\log(\theta_{ij}) = \vartheta_{1ij} = V_j^* + \mathbf{z}_{1ij}^F \beta_1^F \quad \text{e} \quad \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \vartheta_{2ij} = \mathbf{z}_{2i}^F \beta_2^F, \quad (4.9)$$

ou

$$\log(\theta_{ij}) = \vartheta_{1ij} = \mathbf{z}_{1ij}^F \beta_1^F \quad \text{e} \quad \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \vartheta_{2ij} = V_j^* + \mathbf{z}_{2i}^F \beta_2^F, \quad (4.10)$$

em ambos casos, para $1 \leq i \leq n_j, 1 \leq j \leq J$.

Desta forma, assume-se que para $j = 1, \dots, J$,

$$\mathbf{K}_j^* \sim \text{PD}(\alpha \mathcal{G}), \quad \mathcal{G} \mid \alpha, \mathcal{G}_0, \beta_b, \Sigma \sim \text{PD}(\alpha \mathcal{G}_0(\beta_k, \Sigma)), \quad \mathcal{G}_0(\beta_k, \Sigma) = N(\beta_k, \Sigma).$$

Para completar a especificação *a priori* do modelo, usamos as seguintes distribuições *a priori* para os demais parâmetros:

$$\begin{aligned} \gamma &\sim N(\mu_\gamma, \Upsilon_\gamma), \\ \beta_1^F &\sim N(\mu_{\beta_1^F}, \Upsilon_{\beta_1^F}), \\ \beta_2^F &\sim N(\mu_{\beta_2^F}, \Upsilon_{\beta_2^F}), \\ \alpha &\sim \text{Gama}(\nu_1, \nu_2), \\ \beta_k &\sim N(\mu_{\beta_k}, \Upsilon_{\beta_k}), \\ \Sigma &\sim \mathcal{WI}_2(s_0, \mathcal{S}_0). \end{aligned} \quad (4.11)$$

Note que todas estas reparametrizações feitas não modificam a forma da função de verossimilhança dos modelos destrutivos. Assim, a verossimilhança completa do modelo é dada por

$$\mathcal{L} = \prod_{j=1}^J \mathcal{L}_{1j} \times P(\mathbf{K}_j^* \mid \mathbf{K}^{*(j)}, \alpha, \beta_k, \Sigma),$$

em que

$$\mathcal{L}_{1j} = \prod_{i=1}^{n_j} \left\{ S_{pop}(t_{ij} \mid \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij}) \times [h_{pop}(t_{ij} \mid \eta_{ij}, \vartheta_{1ij}, \vartheta_{2ij})]^{\delta_{ij}} \right\}, \quad j = 1, \dots, J, \quad (4.12)$$

com $S_{pop}(\cdot)$ e $h_{pop}(\cdot)$ definidas na Tabela 4.2 e que dependem do modelo destrutivo utilizado e a M.P. associada a $\mathbf{K}_j^* \mid \mathbf{K}^{*(j)}, \alpha, \beta_k, \Sigma$ é dada por

$$dF(\mathbf{K}_j^* \mid \mathbf{K}^{*(j)}, \alpha, \beta_k, \Sigma) \propto \alpha \mathcal{G}_0(\beta_k, \Sigma) + \sum_{\ell \neq j} \delta_{\mathbf{K}_\ell^*}^*,$$

em que $\mathbf{K}^{*(j)}$ representa o vetor $\mathbf{K}^*$ sem o j -ésimo elemento. Portanto, a distribuição *a posteriori* conjunta de $(\gamma, \beta_1^F, \beta_2^F, \mathbf{K}^*, \alpha, \beta_k, \Sigma)$ é dada por

$$\begin{aligned} \pi(\gamma, \beta_1^F, \beta_2^F, \mathbf{K}^*, \alpha, \beta_k, \Sigma) &\propto \left\{ \prod_{j=1}^J \mathcal{L}_{1j} \times P(\mathbf{K}_j^* \mid \mathbf{K}^{*(j)}, \alpha, \beta_k, \Sigma) \right\} \times \\ &\times \pi(\gamma) \pi(\beta_1^F) \pi(\beta_2^F) \pi(\beta_k) \pi(\Sigma) \pi(\alpha), \end{aligned}$$

em que

$$\begin{aligned} \pi(\gamma) &\propto \exp \left\{ -\frac{1}{2}(\gamma - \mu_\gamma)^\top \Upsilon_\gamma^{-1}(\gamma - \mu_\gamma) \right\}, \\ \pi(\beta_1^F) &\propto \exp \left\{ -\frac{1}{2}(\beta_1^F - \mu_{\beta_1^F})^\top \Upsilon_{\beta_1^F}^{-1}(\beta_1^F - \mu_{\beta_1^F}) \right\}, \\ \pi(\beta_2^F) &\propto \exp \left\{ -\frac{1}{2}(\beta_2^F - \mu_{\beta_2^F})^\top \Upsilon_{\beta_2^F}^{-1}(\beta_2^F - \mu_{\beta_2^F}) \right\}, \\ \pi(\alpha) &\propto \alpha^{\nu_1-1} \exp \{ -\nu_2 \alpha \}, \\ \pi(\beta_k) &\propto \exp \left\{ -\frac{1}{2}(\beta_k - \mu_{\beta_k})^\top \Upsilon_{\beta_k}^{-1}(\beta_k - \mu_{\beta_k}) \right\}, \\ \pi(\Sigma) &\propto |\Sigma|^{-(s_0+3)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\mathcal{S}_0 \Sigma^{-1}) \right\}. \end{aligned}$$

Note que a distribuição condicional completa dos vetores γ, β_1^F e β_2^F não tem forma fechada e depende do respectivo modelo destrutivo na Tabela 4.4 na forma

$$\pi(\gamma \mid \beta_1^F, \beta_2^F, \mathbf{K}^*) \propto \left\{ \prod_{j=1}^J \mathcal{L}_{1j} \right\} \pi(\gamma), \quad (4.13)$$

$$\pi(\beta_1^F \mid \gamma, \beta_2^F, \mathbf{K}^*) \propto \left\{ \prod_{j=1}^J \mathcal{L}_{1j} \right\} \pi(\beta_1^F) \quad \text{e} \quad (4.14)$$

$$\pi(\beta_2^F \mid \gamma, \beta_1^F, \mathbf{K}^*) \propto \left\{ \prod_{j=1}^J \mathcal{L}_{1j} \right\} \pi(\beta_2^F), \quad (4.15)$$

em que $\mathcal{L}_{1j}$ é dado em (4.12). Pode-se simular valores destas distribuições utilizando métodos MCMC (Gamerman e Lopes [2006]). Por exemplo, para simular de $\gamma \mid \beta_1^F, \beta_2^F, \mathbf{K}^*$, pode-se usar o seguinte algoritmo baseado no algoritmo de Metrópolis-Hastings e utilizando uma distribuição simétrica para simular os valores propostos:

1. Supor que o valor de γ no passo anterior é dado por γ_ℓ .
2. Simular $\epsilon \sim N(0, \sigma_\epsilon^2)$, em que σ_ϵ^2 é um valor conhecido. Logo, propor $\gamma_{\ell+1} = \gamma_\ell + \epsilon$ como o

próximo valor simulado para γ .

3. Calcular $\pi_{\ell+1}^* = \frac{\pi(\gamma_{\ell+1} | \beta_1^F, \beta_2^F, \mathbf{K}^*)}{\pi(\gamma_\ell | \beta_1^F, \beta_2^F, \mathbf{K}^*)}$ baseado em (4.13).
4. Aceitar $\gamma_{\ell+1}$ como o novo valor proposto para γ se $W \leq \pi_{\ell+1}^*$, em que $W \sim U(0, 1)$; caso contrário, $\gamma_{\ell+1} = \gamma_\ell$.

Valores para β_1^F e β_2^F podem ser simulados de forma similar.

Por outro lado, as distribuições *a posteriori* para os demais parâmetros possuem forma fechada e são apresentadas a seguir. Note que a M.P. *a posteriori* de $\mathbf{K}_j^* | \mathbf{K}^{*(j)}, \gamma, \beta_1^F, \beta_2^F, \alpha, \beta_k, \Sigma$ é dada por

$$dF(\mathbf{K}_j^* | \mathbf{K}^{*(j)}, \gamma, \beta_1^F, \beta_2^F, \alpha, \beta_k, \Sigma) \propto \alpha \mathcal{L}_{1j} \mathcal{G}_0(\beta_k, \Sigma) + \sum_{\ell \neq j} \mathcal{L}_{1\ell} \delta_{\mathbf{K}_\ell^*}^*, \quad (4.16)$$

i.e., uma mistura entre uma distribuição discreta nos pontos do vetor $\mathbf{K}^{*(j)}$ com probabilidades proporcionais a sua contribuição na verossimilhança e a distribuição continua com densidade proporcional a $\alpha \mathcal{L}_{1j} \mathcal{G}_0(\beta_k, \Sigma)$. Para simular valores desta distribuição pode-se utilizar os métodos não conjugados para P.D.'s explicados na seção 2.4.1.

Entretanto, para β_k e Σ , as distribuições *a posteriori* condicionais completas são dadas por

$$\pi(\beta_k | \mathbf{K}^*, \alpha, \Sigma) \propto \exp \left\{ -\frac{1}{2} \left(\beta_k - \mathcal{A}_{\beta_k}^{-1} \mathbf{b}_{\beta_k} \right)^\top \mathcal{A}_{\beta_k} \left(\beta_k - \mathcal{A}_{\beta_k}^{-1} \mathbf{b}_{\beta_k} \right) \right\}, \quad (4.17)$$

$$\pi(\Sigma | \mathbf{K}^*, \alpha, \beta_k) \propto |\Sigma|^{-(J+s_0+3)/2} \exp \left\{ -\frac{1}{2} \text{tr}([\mathcal{S} + \mathcal{S}_0] \Sigma^{-1}) \right\}, \quad (4.18)$$

em que $\mathcal{A}_{\beta_k} = J\Sigma^{-1} + \Upsilon_{\beta_k}^{-1}$, $\mathbf{b}_{\beta_k} = J\Sigma^{-1} \bar{\mathbf{K}}^* + \Upsilon_{\beta_k}^{-1} \mu_{\beta_k}$, $\mathcal{S} = J(\bar{\mathbf{K}}^* - \beta_k)(\bar{\mathbf{K}}^* - \beta_k)^\top$ e $\bar{\mathbf{K}}^* = (1/J) \sum_{j=1}^J \mathbf{K}_j^*$. A partir de (4.17) e (4.18), conclui-se que

$$\beta_k | \mathbf{K}^*, \alpha, \Sigma \sim N(\mathcal{A}_{\beta_k}^{-1} \mathbf{b}_{\beta_k}, \mathcal{A}_{\beta_k}^{-1}), \quad \text{e} \quad (4.19)$$

$$\Sigma | \mathbf{K}^*, \alpha, \beta_k \sim \mathcal{WI}_2(J + s_0, \mathcal{S} + \mathcal{S}_0). \quad (4.20)$$

Finalmente, para amostrar valores de α , pode-se introduzir as variáveis latentes φ e v , com as quais obtém-se, como em (2.49), os seguintes resultados

$$\begin{aligned} \alpha | \varphi, v, \mathbf{K}^* &\sim \text{Gama}(\nu_1 + k^* - v, \nu_2 - \log \varphi), \\ \varphi | \alpha &\sim \text{Be}(\alpha + 1, J) \quad \text{e} \\ v | \alpha &\sim \text{Bern} \left(\frac{J/\alpha}{1 + J/\alpha} \right), \end{aligned} \quad (4.21)$$

em que k^* representa o número de vetores distintos em $\mathbf{K}^*$.

Desta forma, pode-se simular uma amostra das distribuições condicionais completas de forma sequencial usando as expressões (4.13), (4.14), (4.15), (4.16), (4.19), (4.20) e (4.21).

4.4 Estudo de simulação: comparação das abordagens clássica e bayesiana nos MDPPFC com efeitos aleatórios

Será apresentado um estudo de simulação para comparar os MDPPFC com e.a. normais discutido na seção 4.2 com os MDPPFC com e.a. seguindo *a priori* um P.D. como discutido na seção

4.3. Porém, a proposta bayesiana mediante o uso de P.D.'s tem um custo computacional muito elevado. Portanto, é impraticável realizar um estudo fazendo réplicas como nos estudos de simulação apresentados na seção 3. Portanto, cada caso considerado nesta subseção será replicado apenas uma vez. Este procedimento pode parecer inadequado, mas na prática é muito utilizado. Por exemplo, pode ser visto em Jara et al. [2007] e Vieira et al. [2013].

Neste caso, os e.a.'s serão simulados de diferentes distribuições (não só da distribuição Normal), com a finalidade de testar se a metodologia proposta dos P.D.'s consegue capturar ou não a distribuição para os e.a.'s. Foram utilizadas as seguintes distribuições:

- **Caso 1:** $(U_j, V_j)^\top \sim N_2(0, \Sigma)$, $\forall j = 1, \dots, J$, em que $\text{Var}(U_j) = \text{Var}(V_j) = 0.5$ e $\text{Cor}(U_j, V_j) = 0.4$,
- **Caso 2:** $(U_j, V_j)^\top \sim T_2(0, \Sigma, 5)$, $\forall j = 1, \dots, J$, i.e., a distribuição t de Student bivariada central com 5 graus de liberdade e matriz de escala Σ . A matriz Σ é a mesma que a considerada no caso 1,
- **Caso 3:** $(U_j, V_j)^\top \sim SN_2(-\sqrt{\frac{2}{\pi}}\psi, \Sigma, \psi)$, $\forall j = 1, \dots, J$, em que $\psi = (-2, -2)^\top$ i.e., a distribuição normal assimétrica bivariada centrada em zero, com matriz de escala Σ considerada como no caso 1 e vetor de assimetria ψ . (Foi considerada a parametrização desta distribuição proposta em Sahu et al. [2003]),
- **Caso 4:** $U_j \sim U(-1, 1)$ independentes de $V_j \sim U(-1, 1)$, $\forall j = 1, \dots, J$,
- **Caso 5:** $U_j \sim 3 \times \text{Be}(4, 2) - 2$ independentes de $V_j \sim 3 \times \text{Be}(4, 2) - 2$, $\forall j = 1, \dots, J$, em que $B(4, 2)$ representa a distribuição Beta de média $2/3$ e variância $2/63$.
- **Caso 6:** $U_j \sim \frac{1}{2}N(-0.5, 1) + \frac{1}{2}N(0.5, 1)$ independentes de $V_j \sim \frac{1}{2}N(-0.5, 1) + \frac{1}{2}N(0.5, 1)$ $\forall j = 1, \dots, J$.

Em todos os casos considerados, $E(U_j, V_j) = (0, 0)$ e marginalmente a distribuição de U_j coincide com a de V_j para $j = 1, \dots, J$. As funções de densidade marginais destes 6 casos são apresentadas na Figura 4.1. Note que as distribuições seleccionadas envolvem um espectro bem amplo: a distribuição T com 5 graus de liberdade tem caldas bastante mais pesadas que a Normal, a distribuição Skew-Normal representa uma distribuição assimétrica com valores nos reais, a distribuição Uniforme representa uma distribuição simétrica e limitada, a distribuição $3 \times \text{Be}(4, 2) - 2$ é assimétrica e limitada, entanto que a mistura de Normais representa uma distribuição bimodal.

A simulação dos dados é feita de forma similar aos estudos apresentados na seção 3, mas neste caso é necessário simular tanto o número de células iniciais de cada indivíduo (M_{ij}) quanto o número de células que efectivamente podem produzir a falha de cada indivíduo (D_{ij}).

Assuma que o i -ésimo indivíduo da j -ésima clínica possui covariáveis $\mathbf{x}_{ij}$, $\mathbf{z}_{1ij}$ e $\mathbf{z}_{2ij}$ de dimensão r , s_1 e s_2 respectivamente para explicar, nesta mesma ordem, a heterogeneidade na sobrevivência dos indivíduos suscetíveis, o número de células iniciais de cada indivíduo e o número de células que efectivamente podem produzir a falha. Os pares de e.a.'s (U_j, V_j) , $j = 1, \dots, J$ foram gerados conforme cada caso exposto anteriormente. Foi assumido só o caso em que os e.a.'s V_j são introduzidos ao modelo através de p_{ij} . Portanto, os valores de η_{ij} , θ_{ij} e p_{ij} são computados baseados em (4.1) e (4.2), em que γ , β_1 e β_2 são vetores conhecidos com as dimensões apropriadas. Cada valor M_{ij} foi gerado da distribuição Poisson Ponderada de parâmetro θ_{ij} e função de pesos $w(\cdot \mid \phi)$ segundo cada modelo destrutivo da Tabela 4.1. Logo foram simulados os valores de D_{ij} da distribuição $\text{Bin}(M_{ij}, p_{ij})$ caso $M_{ij} > 0$ e definido $D_{ij} = 0$ caso contrário. Se $D_{ij} = 0$, então T_{ij}^* , o tempo de falha do i -ésimo indivíduo da j -ésima clínica, é definido como ∞ . Se $D_{ij} > 0$, então simula-se $Y_{1ij}, \dots, Y_{D_{ij}ij}$ tempos com densidade $h_0(t \mid \lambda)e^{\eta_{ij}}[S_0(t \mid \lambda)]^{\exp \eta_{ij}}$, em que $S_0(t \mid \lambda)$ será assumida provindo da distribuição Weibull neste estudo de simulação, e defina $T_{ij}^* = \min\{Y_{aij}, a = 1, \dots, D_{ij}\}$. Por último,

Figura 4.1: Funções de densidade (marginais) das distribuições seleccionadas para os e.a.'s.

para um tempo de censura C fixado, defina as variáveis $T_{ij} = \min(T_{ij}^*, C)$ e $\delta_{ij} = I(T_{ij}^* \leq C)$, que representam os tempos observados e o indicador de falha para cada indivíduo.

Foi considerado

- $\mathbf{x}_{ij} = \mathbf{z}_{1ij}$,
- em cada clínica, as covariáveis $(\mathbf{z}_{1ij}, \mathbf{z}_{2ij})$ segundo a Tabela de contingência

$\mathbf{z}_{1ij} \backslash \mathbf{z}_{2ij}$	(1,0)	(1,1)
0	$n_j/4$	$n_j/4$
1	$n_j/4$	$n_j/4$

- Para cada modelo destrutivo, foram fixados os parâmetros do modelo conforme a seguinte Tabela

Parâmetro	Modelo		
	DLBP	DEWPM	DBN
γ	0.5	-0.5	-0.5
β_1	0.5	0.5	0.5
β_{20}	0.5	0.5	0.5
β_{21}	-0.5	0.5	-0.5

- Em todas as simulações foi considerado $\{\lambda = 0.01, \nu = 1, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.4, C = 1000\}$.
- Para o número de clínicas e o número de pacientes em cada clínica foram considerados os casos $(J = 5, n_j = 20)$, $(J = 5, n_j = 60)$, $(J = 10, n_j = 20)$ e $(J = 10, n_j = 60)$.

As distribuições *a priori* foram formuladas no sentido de fornecer pouca informação. Pontualmente, foi considerada a distribuição $N(0, 100)$ para $\gamma, \beta_1^F, \beta_2^F$ e β_k . Para o parâmetro α foi considerada a distribuição Gama(1, 1), i.e., uma distribuição que privilegia valores pequenos em relação aos valores elevados. Finalmente, para a matriz Σ foi considerada a distribuição $\mathcal{WI}_2(8, I_2)$, em que I_2 é a matriz identidade de dimensão 2. No caso da abordagem bayesiana, foi simulado uma amostra de tamanho 250.000. Após retirar os 50000 primeiros valores simulados, foi considerado só uma observação a cada 10, obtendo-se uma amostra final de tamanho 10000.

Para os parâmetros associados aos efeitos fixos do modelo, γ, β_1 e β_{21} , são apresentados os estimadores obtidos com a proposta clássica da subseção 4.2 e a média *a posteriori* dos respectivos valores obtidos na proposta bayesiana da seção 4.3. Também são apresentados os intervalos de 95% de confiança usando a distribuição assintótica em (4.6) e os respectivos intervalos de credibilidade (também de 95%) usando as respectivas amostras *a posteriori* obtidas. Estes resultados são apresentados na Tabela 4.3. Foi destacado o melhor estimador pontual (no sentido de ficar mais próximo do verdadeiro valor do parâmetro) para cada parâmetro em cada caso, assim como os intervalos de confiança ou credibilidade que não contém ao verdadeiro valor do parâmetro.

Modelo	Caso	Par.	Verd.	Estimador		Int. conf.	Int. cred.
				Clás.	Bayes.	95%	95%
DLBP ($J = 5$, $n_j = 20$)	1	γ	0.5	0.631	0.648	(0.038,1.225)	(0.362,0.967)
		β_1	0.5	0.532	0.236	(-0.209,1.273)	(-1.605,1.179)
		β_{21}	-0.5	-0.417	-0.322	(-0.969,0.136)	(-1.404,0.766)
	2	γ	0.5	-0.008	-0.005	(-0.59,0.575)	(-0.297,0.292)
		β_1	0.5	0.868	0.624	(0.211,1.526)	(-0.71,1.452)
		β_{21}	-0.5	-0.595	-0.630	(-1.154,-0.037)	(-1.559,0.231)
	3	γ	0.5	1.219	1.238	(0.675,1.763)	(0.959,1.554)
		β_1	0.5	-0.373	-0.926	(-1.698,0.952)	(-2.834,0.591)
		β_{21}	-0.5	-0.301	-0.370	(-0.846,0.244)	(-1.507,0.788)
	4	γ	0.5	0.618	0.638	(0.108,1.129)	(0.36,0.96)
		β_1	0.5	0.616	0.178	(-0.091,1.324)	(-1.933,1.26)
		β_{21}	-0.5	-0.241	-0.457	(-0.803,0.32)	(-1.649,0.611)
	5	γ	0.5	0.690	0.678	(0.165,1.215)	(0.361,0.964)
		β_1	0.5	0.444	0.082	(-0.257,1.146)	(-1.755,1.128)
		β_{21}	-0.5	0.203	0.193	(-0.332,0.739)	(-0.866,1.226)
	6	γ	0.5	1.365	1.339	(0.768,1.962)	(1.016,1.616)
		β_1	0.5	-1.129	-1.316	(-3.614,1.356)	(-2.911,0.383)
		β_{21}	-0.5	-0.800	-0.761	(-1.332,-0.268)	(-1.869,0.342)
DLBP ($J = 5$, $n_j = 60$)	1	γ	0.5	0.755	0.764	(0.444,1.067)	(0.581,0.987)
		β_1	0.5	0.085	-0.143	(-0.446,0.615)	(-1.603,0.579)
		β_{21}	-0.5	-0.769	-0.829	(-1.074,-0.465)	(-1.538,-0.175)
	2	γ	0.5	0.359	0.367	(0.016,0.701)	(0.189,0.587)
		β_1	0.5	0.700	0.645	(0.3,1.1)	(0.015,1.127)
		β_{21}	-0.5	-0.746	-0.748	(-1.072,-0.42)	(-1.294,-0.216)
	3	γ	0.5	0.353	0.359	(0.002,0.704)	(0.174,0.573)
		β_1	0.5	0.690	0.611	(0.278,1.101)	(0.016,1.07)
		β_{21}	-0.5	-0.287	-0.317	(-0.622,0.048)	(-0.9,0.262)
	4	γ	0.5	0.260	0.264	(-0.088,0.608)	(0.074,0.471)
		β_1	0.5	0.835	0.760	(0.436,1.234)	(0.179,1.225)
		β_{21}	-0.5	-0.346	-0.394	(-0.67,-0.022)	(-0.976,0.145)
	5	γ	0.5	0.429	0.435	(0.08,0.778)	(0.242,0.658)
		β_1	0.5	0.349	0.263	(-0.114,0.811)	(-0.582,0.829)
		β_{21}	-0.5	-0.302	-0.343	(-0.617,0.014)	(-0.918,0.216)
	6	γ	0.5	0.464	0.471	(0.149,0.78)	(0.281,0.683)
		β_1	0.5	0.134	-0.100	(-0.389,0.657)	(-1.416,0.587)
		β_{21}	-0.5	-0.473	-0.530	(-0.782,-0.165)	(-1.143,0.034)
DLBP ($J = 10$, $n_j = 20$)	1	γ	0.5	0.313	0.315	(-0.106,0.731)	(0.047,0.583)
		β_1	0.5	0.720	0.697	(0.232,1.207)	(-0.204,1.406)
		β_{21}	-0.5	-0.623	-0.700	(-1.018,-0.227)	(-1.726,0.195)
	2	γ	0.5	0.276	0.298	(-0.187,0.738)	(0.056,0.6)
		β_1	0.5	0.023	-0.254	(-0.714,0.759)	(-2.242,0.823)
		β_{21}	-0.5	-0.630	-0.665	(-1.016,-0.245)	(-1.716,0.269)
	3	γ	0.5	0.404	0.391	(-0.113,0.921)	(0.077,0.677)
		β_1	0.5	0.617	0.571	(0.012,1.221)	(-0.705,1.549)
		β_{21}	-0.5	-0.920	-0.958	(-1.354,-0.485)	(-2.158,0.035)

continua na próxima página.

Tabela 4.3: Comparação das estimativas dos efeitos fixos nos MDPPFC e efeitos aleatórios usando abordagem Clássica e bayesiana.

Modelo	Caso	Par.	Verd.	Estimador		Int. conf.	Int. cred.
				Clás.	Bayes.	95%	95%
	4	γ	0.5	0.341	0.346	(-0.044,0.726)	(0.09,0.617)
		β_1	0.5	0.424	0.222	(-0.124,0.972)	(-1.273,1.047)
		β_{21}	-0.5	-0.755	-0.837	(-1.141,-0.369)	(-1.868,0.107)
	5	γ	0.5	0.534	0.513	(0.187,0.88)	(0.235,0.719)
		β_1	0.5	0.362	0.073	(-0.201,0.926)	(-1.167,0.937)
		β_{21}	-0.5	-0.425	-0.408	(-0.818,-0.032)	(-1.191,0.327)
	6	γ	0.5	0.309	0.315	(-0.148,0.766)	(0.055,0.591)
		β_1	0.5	1.208	1.124	(0.765,1.652)	(0.295,1.857)
		β_{21}	-0.5	-0.298	-0.356	(-0.708,0.112)	(-1.177,0.399)
DLBP ($J = 10$, $n_j = 60$)	1	γ	0.5	0.554	0.558	(0.313,0.794)	(0.375,0.767)
		β_1	0.5	0.531	0.464	(0.229,0.832)	(-0.297,1.056)
		β_{21}	-0.5	-0.533	-0.573	(-0.757,-0.309)	(-1.244,0.035)
	2	γ	0.5	0.482	0.488	(0.23,0.734)	(0.315,0.699)
		β_1	0.5	0.391	0.283	(0.054,0.728)	(-0.629,0.907)
		β_{21}	-0.5	-0.429	-0.422	(-0.659,-0.199)	(-1.004,0.121)
	3	γ	0.5	0.290	0.291	(-0.003,0.582)	(0.057,0.522)
		β_1	0.5	0.821	0.701	(0.498,1.144)	(-0.265,1.487)
		β_{21}	-0.5	-0.209	-0.200	(-0.466,0.049)	(-0.911,0.555)
	4	γ	0.5	0.584	0.582	(0.356,0.812)	(0.373,0.774)
		β_1	0.5	0.515	0.490	(0.217,0.812)	(-0.21,1.09)
		β_{21}	-0.5	-0.553	-0.556	(-0.776,-0.33)	(-1.205,0.068)
	5	γ	0.5	0.492	0.496	(0.264,0.719)	(0.34,0.692)
		β_1	0.5	0.628	0.579	(0.347,0.908)	(-0.016,1.05)
		β_{21}	-0.5	-0.426	-0.426	(-0.65,-0.203)	(-1.082,0.162)
	6	γ	0.5	0.421	0.414	(0.171,0.67)	(0.214,0.574)
		β_1	0.5	0.567	0.538	(0.262,0.872)	(-0.064,1.061)
		β_{21}	-0.5	-0.588	-0.618	(-0.819,-0.357)	(-1.207,-0.07)
DEWPM ($J = 5$, $n_j = 20$)	1	γ	-0.5	-0.4781	-0.48	(-1.101,0.145)	(-0.78,-0.19)
		β_1	0.5	0.716	0.63	(0.206,1.226)	(0.167,1.161)
		β_{21}	0.5	0.1122	1.06	(-1.665,1.889)	(-1.11,2.894)
	2	γ	-0.5	-0.5446	-0.54	(-1.223,0.133)	(-0.867,-0.217)
		β_1	0.5	0.3889	0.211	(-0.187,0.964)	(-0.412,0.954)
		β_{21}	0.5	0.2147	0.306	(-1.165,1.594)	(-1.449,1.853)
	3	γ	-0.5	-0.6362	-0.64	(-1.456,0.184)	(-1.002,-0.294)
		β_1	0.5	0.3116	0.083	(-0.615,1.239)	(-0.899,1.032)
		β_{21}	0.5	0.7033	0.429	(-0.577,1.984)	(-0.921,1.641)
	4	γ	-0.5	-0.6438	-0.62	(-1.235,-0.053)	(-0.912,-0.286)
		β_1	0.5	0.7671	0.646	(0.186,1.348)	(0.038,1.311)
		β_{21}	0.5	0.8083	0.918	(-0.341,1.957)	(-0.596,1.936)
	5	γ	-0.5	-0.3914	-0.41	(-0.895,0.112)	(-0.715,-0.142)
		β_1	0.5	0.6084	0.587	(0.103,1.114)	(0.146,1.092)
		β_{21}	0.5	1.2233	0.811	(-2.172,4.618)	(-0.944,1.948)
	6	γ	-0.5	-0.6946	-0.69	(-1.315,-0.075)	(-0.997,-0.365)
		β_1	0.5	1.1298	0.891	(0.565,1.695)	(0.12,1.63)
		β_{21}	0.5	0.0835	0.224	(-1.14,1.307)	(-1.386,1.765)

continua na próxima página.

Tabela 4.3 - continuação

Modelo	Caso	Par.	Verd.	Estimador		Int. conf.	Int. cred.
				Clás.	Bayes.	95%	95%
DEWPM ($J = 5$, $n_j = 60$)	1	γ	-0.5	-0.498	-0.503	(-0.848,-0.148)	(-0.736,-0.302)
		β_1	0.5	0.519	0.362	(0.207,0.831)	(-0.032,0.814)
		β_{21}	0.5	0.421	0.592	(-0.35,1.191)	(-0.726,1.826)
	2	γ	-0.5	-0.648	-0.645	(-1.045,-0.25)	(-0.845,-0.434)
		β_1	0.5	0.799	0.745	(0.478,1.12)	(0.465,1.052)
		β_{21}	0.5	0.297	0.606	(-0.684,1.279)	(-0.577,1.822)
	3	γ	-0.5	-0.265	-0.272	(-0.699,0.169)	(-0.51,-0.067)
		β_1	0.5	0.904	0.850	(0.515,1.293)	(0.469,1.233)
		β_{21}	0.5	0.279	0.379	(-0.355,0.913)	(-0.269,1.059)
	4	γ	-0.5	-0.522	-0.519	(-0.875,-0.168)	(-0.728,-0.293)
		β_1	0.5	0.496	0.406	(0.187,0.806)	(0.08,0.781)
		β_{21}	0.5	1.383	1.793	(0.023,2.743)	(0.203,2.938)
	5	γ	-0.5	-0.483	-0.466	(-0.873,-0.093)	(-0.672,-0.186)
		β_1	0.5	0.553	0.440	(0.219,0.887)	(-0.012,0.878)
		β_{21}	0.5	0.304	0.340	(-0.401,1.009)	(-0.885,1.53)
	6	γ	-0.5	-0.562	-0.555	(-0.961,-0.163)	(-0.756,-0.321)
		β_1	0.5	0.512	0.431	(0.193,0.832)	(0.096,0.758)
		β_{21}	0.5	0.638	0.857	(-0.145,1.421)	(-0.077,1.837)
DEWPM ($J = 10$, $n_j = 20$)	1	γ	-0.5	-0.801	-0.807	(-1.286,-0.316)	(-1.086,-0.547)
		β_1	0.5	0.784	0.517	(0.369,1.199)	(0.01,1.113)
		β_{21}	0.5	0.354	0.312	(-0.438,1.147)	(-1.091,1.654)
	2	γ	-0.5	-0.276	-0.271	(-0.784,0.232)	(-0.539,0.021)
		β_1	0.5	0.530	0.280	(0.146,0.913)	(-0.165,0.793)
		β_{21}	0.5	0.694	1.013	(-0.321,1.709)	(-0.434,1.946)
	3	γ	-0.5	-0.439	-0.434	(-0.885,0.007)	(-0.679,-0.168)
		β_1	0.5	0.627	0.427	(0.257,0.996)	(0.003,0.945)
		β_{21}	0.5	0.349	0.206	(-0.743,1.442)	(-1.399,1.755)
	4	γ	-0.5	-0.198	-0.177	(-0.646,0.251)	(-0.389,0.114)
		β_1	0.5	0.697	0.617	(0.314,1.081)	(0.186,1.127)
		β_{21}	0.5	-0.357	-0.232	(-1.39,0.675)	(-1.753,1.622)
	5	γ	-0.5	-0.524	-0.511	(-0.963,-0.085)	(-0.738,-0.237)
		β_1	0.5	0.674	0.518	(0.269,1.078)	(0.027,1.029)
		β_{21}	0.5	0.878	1.043	(-0.019,1.774)	(-0.14,1.928)
	6	γ	-0.5	-0.134	-0.124	(-0.625,0.356)	(-0.392,0.178)
		β_1	0.5	0.632	0.552	(0.215,1.049)	(0.007,1.136)
		β_{21}	0.5	0.332	0.508	(-0.562,1.227)	(-0.834,1.798)
DEWPM ($J = 10$, $n_j = 60$)	1	γ	-0.5	-0.426	-0.428	(-0.669,-0.182)	(-0.614,-0.261)
		β_1	0.5	0.495	0.384	(0.292,0.697)	(0.077,0.761)
		β_{21}	0.5	0.436	0.653	(-0.248,1.12)	(-0.622,1.844)
	2	γ	-0.5	-0.453	-0.455	(-0.723,-0.182)	(-0.622,-0.307)
		β_1	0.5	0.671	0.534	(0.45,0.892)	(0.211,0.827)
		β_{21}	0.5	0.760	0.856	(0.207,1.314)	(0.058,1.709)
	3	γ	-0.5	-0.779	-0.752	(-1.074,-0.484)	(-0.921,-0.453)
		β_1	0.5	0.644	0.419	(0.414,0.874)	(-0.066,0.98)
		β_{21}	0.5	0.626	0.356	(-0.055,1.307)	(-1.176,1.741)

continua na próxima página.

Tabela 4.3 - continuação

Modelo	Caso	Par.	Verd.	Estimador		Int. conf. 95%	Int. cred. 95%
				Clás.	Bayes.		
DEWPM ($J = 10$, $n_j = 60$)	4	γ	-0.5	-0.255	-0.259	(-0.501,-0.009)	(-0.473,-0.075)
		β_1	0.5	0.533	0.390	(0.303,0.762)	(-0.028,0.836)
		β_{21}	0.5	0.591	0.707	(0.066,1.117)	(-0.342,1.778)
	5	γ	-0.5	-0.341	-0.334	(-0.62,-0.062)	(-0.517,-0.107)
		β_1	0.5	0.510	0.383	(0.274,0.746)	(-0.022,0.817)
		β_{21}	0.5	0.585	0.709	(0.044,1.126)	(-0.423,1.776)
	6	γ	-0.5	-0.678	-0.676	(-0.99,-0.366)	(-0.866,-0.473)
		β_1	0.5	0.526	0.386	(0.278,0.773)	(-0.002,0.767)
		β_{21}	0.5	0.896	0.942	(0.378,1.414)	(0.223,1.693)
DBN ($J = 5$, $n_j = 20$)	1	γ	-0.5	-0.540	-0.530	(-1.045,-0.035)	(-0.857,-0.178)
		β_1	0.5	0.648	0.544	(-0.129,1.425)	(-0.187,1.371)
		β_{21}	-0.5	-0.986	-0.554	(-2.879,0.907)	(-2.644,2.508)
	2	γ	-0.5	-0.165	-0.189	(-0.795,0.465)	(-0.532,0.11)
		β_1	0.5	0.787	0.547	(0.001,1.574)	(-0.218,1.48)
		β_{21}	-0.5	0.739	1.047	(-0.79,2.267)	(-1.368,2.884)
	3	γ	-0.5	-1.166	-1.141	(-1.945,-0.387)	(-1.482,-0.76)
		β_1	0.5	0.022	-0.168	(-0.881,0.924)	(-1.178,0.864)
		β_{21}	-0.5	1.158	1.357	(-0.42,2.736)	(-0.932,2.897)
	4	γ	-0.5	-1.112	-1.125	(-1.489,-0.734)	(-1.482,-0.794)
		β_1	0.5	1.450	1.347	(0.576,2.324)	(0.302,2.397)
		β_{21}	-0.5	0.005	0.038	(-1.091,1.1)	(-1.459,1.53)
	5	γ	-0.5	-0.940	-0.931	(-1.452,-0.428)	(-1.273,-0.578)
		β_1	0.5	0.928	0.682	(0.15,1.706)	(-0.179,1.708)
		β_{21}	-0.5	0.228	0.738	(-1.098,1.554)	(-1.493,2.807)
	6	γ	-0.5	-0.479	-0.465	(-1.208,0.25)	(-0.8,-0.106)
		β_1	0.5	0.735	0.463	(-0.167,1.637)	(-0.547,1.544)
		β_{21}	-0.5	0.160	0.094	(-1.119,1.438)	(-1.987,2.37)
	1	γ	-0.5	-0.787	-0.792	(-1.152,-0.422)	(-1.066,-0.533)
		β_1	0.5	0.273	0.083	(-0.182,0.727)	(-0.414,0.666)
		β_{21}	-0.5	0.208	0.568	(-0.838,1.255)	(-1.409,2.732)
	2	γ	-0.5	-0.891	-0.898	(-1.434,-0.348)	(-1.176,-0.638)
		β_1	0.5	1.103	0.924	(0.503,1.703)	(0.325,1.516)
		β_{21}	-0.5	-0.336	-0.640	(-1.178,0.505)	(-1.764,0.282)
	3	γ	-0.5	-0.449	-0.448	(-0.891,-0.007)	(-0.705,-0.188)
		β_1	0.5	0.551	0.375	(0.044,1.058)	(-0.119,0.91)
		β_{21}	-0.5	0.260	-0.032	(-0.692,1.213)	(-1.532,1.403)
	4	γ	-0.5	-0.515	-0.514	(-0.834,-0.197)	(-0.772,-0.25)
		β_1	0.5	0.572	0.378	(0.125,1.019)	(-0.182,1.008)
		β_{21}	-0.5	-0.284	-0.289	(-1.198,0.63)	(-1.914,1.805)
	5	γ	-0.5	-0.522	-0.533	(-0.905,-0.139)	(-0.81,-0.293)
		β_1	0.5	0.758	0.611	(0.298,1.217)	(-0.002,1.23)
		β_{21}	-0.5	-0.076	-0.045	(-0.892,0.74)	(-1.347,1.201)
	6	γ	-0.5	0.041	0.045	(-0.403,0.485)	(-0.206,0.31)
		β_1	0.5	0.542	0.494	(0.058,1.026)	(0.031,0.959)
		β_{21}	-0.5	-0.635	-0.793	(-1.468,0.198)	(-1.708,0.073)

continua na próxima página.

Tabela 4.3 - continuação

Modelo	Caso	Par.	Verd.	Estimador		Int. conf.	Int. cred.
				Clás.	Bayes.	95%	95%
DBN ($J = 10$, $n_j = 20$)	1	γ	-0.5	-0.509	-0.493	(-0.931,-0.087)	(-0.769,-0.175)
		β_1	0.5	1.249	1.121	(0.634,1.864)	(0.328,1.904)
		β_{21}	-0.5	0.073	0.081	(-0.714,0.86)	(-0.99,1.146)
	2	γ	-0.5	-0.378	-0.376	(-0.873,0.116)	(-0.689,-0.056)
		β_1	0.5	0.333	0.064	(-0.258,0.924)	(-0.613,0.814)
		β_{21}	-0.5	-0.663	-1.020	(-1.695,0.369)	(-2.727,0.867)
	3	γ	-0.5	-0.720	-0.658	(-1.102,-0.338)	(-0.871,-0.304)
		β_1	0.5	0.471	0.397	(-0.075,1.018)	(-0.282,1.059)
		β_{21}	-0.5	-0.413	0.346	(-2.014,1.187)	(-2.256,2.832)
	4	γ	-0.5	-0.362	-0.364	(-0.739,0.014)	(-0.648,-0.09)
		β_1	0.5	0.467	0.210	(-0.048,0.983)	(-0.272,0.783)
		β_{21}	-0.5	-0.382	0.731	(-1.776,1.013)	(-1.596,2.866)
	5	γ	-0.5	-0.423	-0.400	(-0.806,-0.04)	(-0.652,-0.087)
		β_1	0.5	0.383	0.225	(-0.153,0.919)	(-0.283,0.79)
		β_{21}	-0.5	0.301	0.720	(-1.125,1.727)	(-1.535,2.862)
	6	γ	-0.5	-0.771	-0.780	(-1.216,-0.325)	(-1.089,-0.493)
		β_1	0.5	0.522	0.166	(0.008,1.036)	(-0.368,0.788)
		β_{21}	-0.5	-1.012	-0.413	(-2.361,0.337)	(-2.481,2.559)
DBN ($J = 10$, $n_j = 60$)	1	γ	-0.5	-0.108	-0.119	(-0.394,0.179)	(-0.341,0.045)
		β_1	0.5	0.517	0.497	(0.211,0.823)	(0.081,0.867)
		β_{21}	-0.5	0.102	0.194	(-0.529,0.734)	(-0.686,1.015)
	2	γ	-0.5	-0.613	-0.630	(-0.893,-0.333)	(-0.891,-0.453)
		β_1	0.5	0.541	0.370	(0.234,0.847)	(-0.078,0.883)
		β_{21}	-0.5	0.312	0.673	(-0.403,1.026)	(-0.815,2.54)
	3	γ	-0.5	-0.317	-0.316	(-0.62,-0.014)	(-0.52,-0.112)
		β_1	0.5	0.769	0.668	(0.434,1.105)	(0.171,1.142)
		β_{21}	-0.5	-0.607	-0.745	(-1.157,-0.056)	(-1.76,0.056)
	4	γ	-0.5	-0.569	-0.551	(-0.806,-0.333)	(-0.737,-0.281)
		β_1	0.5	0.384	0.267	(0.069,0.698)	(-0.147,0.716)
		β_{21}	-0.5	-0.345	-0.508	(-1.133,0.442)	(-2.108,1.402)
	5	γ	-0.5	-0.509	-0.491	(-0.753,-0.266)	(-0.675,-0.223)
		β_1	0.5	0.558	0.382	(0.237,0.878)	(-0.09,0.883)
		β_{21}	-0.5	-0.069	-0.068	(-0.709,0.572)	(-1.602,1.673)
	6	γ	-0.5	-0.029	-0.021	(-0.295,0.237)	(-0.231,0.237)
		β_1	0.5	0.417	0.316	(0.095,0.739)	(-0.076,0.753)
		β_{21}	-0.5	-0.774	-0.797	(-1.526,-0.022)	(-2.11,0.39)

Tabela 4.3 - continuação

Percebe-se da Tabela 4.3 que em geral, os intervalos de credibilidade para o parâmetro γ são mais precisos do que os intervalos de confiança, mas os intervalos de credibilidade para os parâmetros β_1 e β_{21} (aqueles associados à cura) são menos precisos do que os respectivos intervalos de confiança. O número de intervalos de confiança que não contêm o verdadeiro valor do parâmetro é similar ao número de intervalos de credibilidade que não contêm esses mesmos valores (16 e 18, respectivamente de um total de 216 intervalos).

As Tabelas 4.4, 4.5 e 4.6 apresentam um breve sumário para as abordagens clássica e bayesiana (separado por distribuição, por parâmetro e por combinação de número de clínicas e pacientes em cada), incluindo o vício médio, a média dos erros quadráticos médios ($\overline{E.Q.M.}$) e a porcentagem de

vezes em que a estimativa clássica ficou mais próxima do verdadeiro valor do parâmetro (% Vezes).

De um ponto descritivo, podemos dizer que ambas análises forneceram resultados muito similares em relação ao vício médio nos casos 1, 4 e 6 (distribuições normal, uniforme e mistura de normais), todas distribuições simétricas em torno de zero, com o detalhe que $\overline{E.Q.M.}$ para a abordagem clássica foi sempre menor. Por outro lado, só no caso 1 (quando os efeitos foram simulados da distribuição normal) a porcentagem de vezes em que os estimadores da abordagem clássica ficaram próximos um número de vezes que destaca-se sobre o resto. Nos casos 2, 3 e 5 (distribuições t de Student com 5 graus de liberdade, normal assimétrica e transformação da beta), a abordagem bayesiana forneceu melhores vícios médios, mas sempre tendo como custo um $\overline{E.Q.M.}$ muito superior àquele fornecido pela abordagem Clássica. Note também que ambas abordagens forneceram praticamente a metade de vezes em que os estimadores ficaram mais próximos do verdadeiro valor. Desta forma, uma hipótese inicial que se pode formular a partir deste estudo de simulação é que a abordagem clássica proposta parece fornecer melhores resultados em relação ao vício médio quando os e.a.'s provêm de uma distribuição simétrica em torno de zero, mas sem caudas muito pesadas, entretanto que a abordagem bayesiana fornece melhores resultados nas outras condições aqui estudadas. Contudo, parece verificar-se também que o $\overline{E.Q.M.}$ dos estimadores bayesianos propostos sempre são maiores que aqueles obtidos a partir da abordagem clássica proposta, o que faz sentido uma vez que o uso dos P.D.'s acrescenta incerteza sobre a distribuição dos e.a.'s.

Por outro lado, uma análise descritiva comparando as duas abordagens em termos de cada parâmetro sugere que os vícios médios de $\hat{\gamma}$ são similares para ambas abordagens, o que também é válido para os $\overline{E.Q.M.}$'s. Porém, a alternativa bayesiana forneceu valores mais próximos de verdadeiro valor de γ em um maior número de vezes. Entretanto, para os estimadores dos parâmetros associados à cura, $\hat{\beta}_1$ e $\hat{\beta}_2$, verifica-se que o vício médio da abordagem clássica é menor que o vício médio fornecido pela alternativa bayesiana, fornecendo também um $\overline{E.Q.M.}$ menor para a alternativa clássica. Destaca-se também o elevado vício para $\hat{\beta}_2$, o qual é ainda maior na abordagem bayesiana.

Finalmente, um aumento do número de pacientes em cada clínica melhorou o desempenho em relação ao vício médio dos estimadores bayesianos para $J = 5$, mas piorou o desempenho deles para $J = 10$. Porém, o $\overline{E.Q.M.}$ da abordagem bayesiana foi maior do que aquele fornecido pela abordagem clássica, sendo que o caso $J = 5$ e $n_j = 20$ o caso em que o $\overline{E.Q.M.}$ foi quase 6 vezes superior ao fornecido pela abordagem clássica. Desta forma, este estudo de simulação permite formular, como hipótese inicial, que a alternativa bayesiana é adequada só se tivermos um número elevado de observações em cada grupo.

Caso	Abordagem	Vício médio	$\overline{E.Q.M.}$	% Vezes
1	Clássica	0.043	0.087	63.9
	bayesiana	0.056	0.114	
2	Clássica	0.053	0.119	52.8
	bayesiana	0.014	0.181	
3	Clássica	0.061	0.181	50.0
	bayesiana	0.002	0.251	
4	Clássica	0.081	0.110	47.2
	bayesiana	0.070	0.174	
5	Clássica	0.144	0.096	52.8
	bayesiana	0.115	0.146	
6	Clássica	0.016	0.185	50.0
	bayesiana	-0.012	0.194	

Tabela 4.4: Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por caso).

Parâmetro	Abordagem	Vício médio	$\overline{E.Q.M.}$	% Vezes
γ	Clássica	-0.005	0.069	40.3
	bayesiana	-0.001	0.068	
β_1	Clássica	0.063	0.122	51.4
	bayesiana	-0.096	0.158	
β_{21}	Clássica	0.140	0.200	66.7
	bayesiana	0.220	0.304	

Tabela 4.5: Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por parâmetro).

J	n_j	Abordagem	Vício médio	$\overline{E.Q.M.}$	% Vezes
5	20	Clássica	0.101	0.060	40.7
		bayesiana	0.060	0.359	
5	60	Clássica	0.064	0.085	50.0
		bayesiana	0.035	0.114	
10	20	Clássica	0.016	0.090	59.3
		bayesiana	0.013	0.164	
10	60	Clássica	0.084	0.049	61.1
		bayesiana	0.056	0.069	

Tabela 4.6: Análise comparativa para os efeitos fixos entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por combinação de número de clínicas e pacientes em cada).

Por outro lado, também é preciso fazer uma análise em relação aos preditores dos efeitos aleatórios. Neste caso, foram comparados as predições dos efeitos aleatórios U_j , $j = 1, \dots, J$ obtidos pela metodologias clássica e bayesiana. Já no caso dos efeitos aleatórios associados à cura, foram comparadas as frações de cura estimadas para cada modelo Destrutivo ($\hat{q}_0$) usando cada abordagem com as verdadeiras frações de cura (q_0). Apresentam-se a média de algumas estatísticas descritivas, como a média, a mediana e o primeiro e terceiro quartis (estes últimos denotados por $P^{(25)}$ e $P^{(75)}$).

respectivamente). Esta análise descritiva foi feita tanto para cada distribuição considerada para os e.a.'s quanto para cada combinação de clínicas e pacientes em cada, que são apresentadas nas Tabelas 4.7 e 4.8.

Em relação às fragilidades U_j 's, percebe-se que em geral, a abordagem bayesiana sobrestima estes e.a.'s entretanto que a alternativa clássica os subestima. Contudo, as médias das médias e medianas de $\hat{U}_j - U_j$ são, em geral, menores para a abordagem clássica. Adicionalmente, as amplitudes interquartis da alternativa clássica são menores, implicando que estas estimativas são mais precisas. Destaca-se também que a amplitude interquartis de ambas abordagens foi excepcionalmente maior para os casos em que os e.a.'s foram simulados da distribuição t de student com 5 graus de liberdade e da distribuição Skew-Normal. Em relação à combinação de número de grupos e observações em cada, observa-se que os casos com $n_j = 60$ pacientes em cada grupo forneceu, em média, melhores predições de U_j quando comparados com os casos em que $n_j = 20$.

No caso dos estimadores da fração de cura, percebe-se que as amplitudes interquartis não distam muito usando ambas abordagens e que os respectivas médias dos vícios médios e medianos da cura estimada não variam muito em ambas abordagens, fornecendo estimativas satisfatórias. Já analisando pelas combinações de clínicas e observações em cada e como esperado, percebe-se que para um número de clínicas fixo, o aumento de pacientes em cada fornece estimativas da cura mais precisas. É importante destacar que, de um ponto de vista descritivo e lembrando que este estudo de simulação é feito sem réplicas, não se encontrou algum cenário no qual o uso dos P.D.'s fornecesse estimativas consideravelmente melhores que a abordagem clássica proposta de um ponto de vista das fragilidades ou das frações de cura estimadas.

Desta forma, sugere-se explorar conjuntamente uma análise clássica e bayesiana dos modelos e depois seleccionar o melhor baseado em algum critério de seleção de modelos. Por exemplo, cAIC ou cBIC desde um ponto de vista clássico e EAIC, EBIC, DIC ou Fator de Bayes desde um ponto de vista bayesiano. Outro critério que pode ser usado é uma análise de resíduos. Nesta linha, os resíduos de Cox-Snell definidos por

$$r_{ij} = H_{pop}(t_{ij} \mid \hat{\eta}_{ij}, \hat{\vartheta}_{1ij}, \hat{\vartheta}_{2ij}), \quad j = 1, \dots, J, i = 1, \dots, n_j, \quad (4.22)$$

em que $H_{pop}(\cdot \mid \hat{\eta}_{ij}, \hat{\vartheta}_{1ij}, \hat{\vartheta}_{2ij}) = -\log S_{pop}(\cdot \mid \hat{\eta}_{ij}, \hat{\vartheta}_{1ij}, \hat{\vartheta}_{2ij})$, $S_{pop}(\cdot \mid \hat{\eta}_{ij}, \hat{\vartheta}_{1ij}, \hat{\vartheta}_{2ij})$ é dado na Tabela 4.2 segundo cada modelo destrutivo, $\hat{\eta}_{ij}$, $\hat{\vartheta}_{1ij}$ e $\hat{\vartheta}_{2ij}$ são computados segundo as equações (4.1), (4.9) e (4.10) dependendo do parâmetro em que são introduzidos os e.a.'s associados à cura. Se o modelo for correto, os resíduos deveriam comportar-se como uma a.a. censurada da distribuição exponencial de média 1.

Caso	Abordagem	$\hat{U}_j - U_j$				$\hat{q}_0 - q_0$			
		$P^{(25)}$	Mediana	Média	$P^{(75)}$	$P^{(25)}$	Mediana	Média	$P^{(75)}$
1	Clássica	-0.185	-0.032	-0.031	0.080	-0.041	0.007	0.006	0.061
	bayesiana	-0.094	0.098	0.064	0.273	-0.078	-0.031	-0.031	0.029
2	Clássica	-0.228	0.005	-0.077	0.181	-0.036	0.019	0.019	0.072
	bayesiana	-0.224	0.057	0.043	0.341	-0.074	-0.011	-0.021	0.045
3	Clássica	-0.521	-0.099	-0.156	0.193	-0.041	0.026	0.021	0.078
	bayesiana	-0.481	0.122	-0.006	0.524	-0.085	0.017	-0.007	0.077
4	Clássica	-0.160	-0.065	-0.044	0.103	-0.039	0.004	0.011	0.059
	bayesiana	-0.152	0.064	0.068	0.294	-0.080	-0.024	-0.024	0.028
5	Clássica	-0.240	-0.087	-0.049	0.099	-0.032	0.011	0.012	0.052
	bayesiana	-0.157	0.055	0.041	0.246	-0.064	-0.020	-0.024	0.021
6	Clássica	-0.189	-0.040	-0.031	0.166	-0.038	0.023	0.017	0.077
	bayesiana	-0.119	0.162	0.118	0.376	-0.096	-0.008	-0.020	0.052

Tabela 4.7: Análise comparativa para os efeitos aleatórios entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por caso).

J	n_j	Abordagem	$\hat{U}_j - U_j$				$\hat{q}_0 - q_0$			
			$P^{(25)}$	Mediana	Média	$P^{(75)}$	$P^{(25)}$	Mediana	Média	$P^{(75)}$
5	20	Clássica	-0.334	-0.040	-0.081	0.125	-0.048	0.021	0.021	0.089
		bayesiana	-0.203	0.133	0.046	0.313	-0.086	-0.023	-0.023	0.048
5	60	Clássica	-0.164	-0.057	-0.041	0.129	-0.032	0.008	0.009	0.051
		bayesiana	-0.095	0.041	0.042	0.195	-0.058	-0.019	-0.022	0.023
10	20	Clássica	-0.334	-0.127	-0.121	0.135	-0.049	0.014	0.012	0.072
		bayesiana	-0.368	0.127	0.055	0.525	-0.108	-0.009	-0.028	0.053
10	60	Clássica	-0.182	0.011	-0.015	0.159	-0.023	0.017	0.016	0.054
		bayesiana	-0.152	0.070	0.075	0.336	-0.065	0.001	-0.012	0.044

Tabela 4.8: Análise comparativa para os efeitos aleatórios entre as abordagens Clássica e bayesiana no MDPPFC e efeitos aleatórios (Detalhe por combinação de número de clínicas e pacientes em cada).

Capítulo 5

Aplicação

Neste capítulo apresenta-se uma aplicação a um banco de dados real dos modelos revisados e propostos nos capítulos 3 e 4.

5.1 Apresentação do banco de dados

O banco de dados que será utilizado nesta seção pode ser encontrado em [Kalbfleisch e Prentice \[2002\]](#). Refere-se a um ensaio clínico com pacientes que tinham câncer de língua. Um total de 195 pacientes foram considerados, com lesões em três zonas diferentes da língua: *faucial arch*, *tonsillar fossa* e *pharyngeal tongue*. Cada paciente foi distribuído a uma das seis clínicas participantes e também foi aleatorizado em um dos dois tratamentos disponíveis: terapia de radiação ou terapia de radiação combinada com quimioterapia. Informação adicional sobre os pacientes também é incluída, como sexo, idade e classificação do estágio do tumor (*T-stage*), entre outras. O evento de interesse foi definido como a morte do paciente, desde que na última visita à clínica tenha sido observado o estágio do tumor mais avançado.

[Kalbfleisch e Prentice \[2002\]](#) discutem a presença de fração de cura para este banco de dados e concluem que há evidência de ter-se indivíduos curados nele. Os pacientes permaneceram no tratamento designado por um período de 90 dias, após o qual eles receberam pós-tratamento. Não haviam restrições para as clínicas em relação ao pós-tratamento, razão pela qual supõe-se que as clínicas podem ter alguma influência tanto nos pacientes suscetíveis quanto nos curados. Como em [Yau e Ng \[2001\]](#), só foram considerados aqueles pacientes com câncer localizado em *pharyngeal tongue* (que corresponde à parte posterior da língua) e que envolve um total de 66 pacientes, em que 47 de eles morreram durante o período do estudo (71%) e 19 ficaram censurados (29%). As curvas de Kaplan-Meier para cada clínica são apresentadas na Figura 5.1.

Segundo [Maller e Zhou \[1996\]](#), uma estimativa consistente para a fração de cura é dada pelo estimador de K-M avaliada no máximo tempo observado (seja falha ou censura). Sendo assim, as estimativas das frações de cura parecem diferir em cada clínica (pelo menos de um ponto de vista descritivo).

Por outro lado, note que as curvas de K-M de cada clínica estabilizam-se nessa estimativa inicial da fração de cura em tempos diferentes. Por exemplo, a curva da clínica 1 estabiliza-se aproximadamente após 400 dias, enquanto que a curva da clínica 5 faz isso ao redor dos 900 dias. Isto pode ser um indício (descritivo) de que os tempos de sobrevivência dos pacientes suscetíveis também são diferentes em cada clínica.

Os dois fatos anteriores sugerem utilizar um modelo que incluía e.a.'s em cada clínica tanto para explicar a sobrevivência dos indivíduos suscetíveis quanto para explicar a sobrevivência dos curados como o contexto explicado nas seções 3 e 4.

Figura 5.1: Curva de K-M para as seis clínicas que participavam no estudo.

Usando a análise de Kalbfleisch e Prentice [2002], sugere-se utilizar só a covariável T -stage para explicar ambas componentes do modelo (suscetíveis e curados), mas recategorizando essa covariável como segue

$$T_{stage_{ij}} = \begin{cases} 1 & , \text{ se o } i\text{-ésimo indivíduo da } j\text{-ésima clínica possui câncer no estágio mais avançado,} \\ 0 & , \text{ caso contrário.} \end{cases}$$

A Tabela 5.1 apresenta os indivíduos que possuem esta variável em cada clínica.

Clínica	1	2	3	4	5	6	Total
T -stage=0	6	7	16	6	4	8	47
T -stage=1	3	7	0	1	7	1	19
Total	9	14	16	7	11	9	66

Tabela 5.1: Distribuição de clínica e da variável T -stage categorizada.

O estimador de K-M para cada perfil recategorizado é dado na figura 5.2. Como esperado, a curva dos indivíduos com câncer no estágio mais avançado está muito abaixo da curva dos indivíduos com câncer nos demais estágios. De fato, os testes de *Log-rank* e *Peto-Peto* (Colosimo e Giolo [2006]) concluem que com 1% de significância que as duas curvas são diferentes.

Figura 5.2: Curva de K-M para a covariável recategorizada *T-stage*.

Por outro lado, como discutido na subseção 2.2.4, por uma questão de identificabilidade os modelos destrutivos requerem introduzir pelo menos duas covariáveis diferentes. Kalbfleisch e Prentice [2002] concluíram que as demais covariáveis não são significativas no modelo, mas foi seleccionada a covariável Idade por ser a única variável contínua e assim exemplificar como pode modificar as predições da cura uma ligação assimétrica como no modelo exposto na seção 3.3. A figura 5.3 mostra para idade o estimador de K-M, separando esta variável pela mediana (62.5 anos). Note que para este tipo de câncer, ser mais velho parece ser um fator de proteção, pois a sobrevivência estimada via K-M para os indivíduos mais velhos é levemente maior que aquela estimada para os mais jovens. Contudo, quando aplicado os testes de *Log-rank* e *Peto-Peto* para comparar as duas curvas de sobrevivência, ambos têm um valor p associado próximo de 20%, indicando que há evidência estatística para dizer que as duas curvas são iguais, e portanto a idade não é uma variável significativa no modelo.

Figura 5.3: Curva de K-M para a variável Idade (categorizada).

Finalmente, a Figura 5.4 apresenta a função de risco acumulada (que pode ser estimada em base ao estimador de K-M) para os dois perfis segundo seu estágio do câncer. Descritivamente, note que a função parece comportar-se com um crescimento constante para os indivíduos suscetíveis. Desta forma, propõe-se a distribuição Exponencial, Weibull e Exponencial Segmentada com dois Intervalos, com partição do eixo do tempo em $a = 250$ (sinalizado no gráfico), que pode sugerir uma mudança no crescimento do risco acumulado. Desta forma, os análises serão feitos baseados nessas três distribuições basais mencionadas.

Figura 5.4: Função de risco acumulada estimada para os perfis $X = 0$ e $X = 1$.

5.2 Ajuste clássico dos modelos

Segundo o discutido na subseção anterior e para fazer uma comparação “justa” entre os modelos Destrutivos e os modelos de mistura e de [Chen et al. \[1999a\]](#), será considerado o vetor que explica a componente suscetível do modelo como $\mathbf{x}_{ij} = (\text{Tstage}_{ij})$ e o vetor que explica a fração de cura como $\mathbf{z}_{ij} = (1, z_{1ij}, z_{2ij}) = (1, \text{Tstage}_{ij}, \text{Idade}_{ij}^*)$, em que Idade_{ij}^* representa a covariável Idade menos a sua respectiva mediana.

Em uma primeira análise, será ajustado o modelo de mistura com diferentes ligações proposto no capítulo 3.3. Os critérios de seleção de modelos para combinação de ligação usada na cura (logito, probito, complementar log-log) e função de sobrevivência basal (Exponencial, Weibull e Exponencial Segmentada) encontra-se na Tabela 5.2.

Note que em geral, os critérios cAIC e cBIC são menores para os modelos em que é utilizada a distribuição Weibull para $S_0(\cdot | \boldsymbol{\lambda})$. Dentre os modelos ajustados e segundo estes critérios, o modelo mais adequado é o modelo com efeitos aleatórios dependentes com ligação complementar log-log na cura e usando distribuição Weibull para a função de sobrevivência basal. A Tabela 5.3 contém as estimativas para o modelo escolhido. Note que o estágio do câncer é significativo na parte suscetível do modelo, mas não na parte associada à cura. Entretanto, e como esperado, a Idade também não é significativa na cura do modelo.

Para exemplificar como pode variar a estimação na cura usando diferentes ligações, calculou-se as estimativas da fração de cura para os perfis $X = 0$ e $X = 1$ usando as três ligações no modelo com e.a.’s dependentes e sobrevivência basal Weibull (linhas contínuas). Também pode-se construir

$S_0(\cdot \lambda)$ proveniente da distribuição Exponencial						
Ligação	Logito	Probit	Cloglog	Logito	Probit	Cloglog
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	717.13	716.94	716.74	716.88	716.70	716.37
cBIC	727.34	727.39	726.96	726.95	727.05	726.96
$S_0(\cdot \lambda)$ proveniente da distribuição Weibull						
Ligação	Logito	Probit	Cloglog	Logito	Probit	Cloglog
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	716.20	716.00	716.32	716.04	715.85	715.48
cBIC	726.73	726.86	727.37	726.55	726.48	726.44
$S_0(\cdot \lambda)$ proveniente da distribuição ES ($L = 2, a = 250$)						
Ligação	Logito	Probit	Cloglog	Logito	Probit	Cloglog
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	720.08	719.97	719.80	719.76	719.60	719.37
cBIC	729.46	729.56	729.34	729.21	729.16	729.00

Tabela 5.2: Critérios de seleção para diferentes ligações no modelo de mistura com e.a.'s

Parâmetro	Estimativa	Desvio padrão	E.a.	Predito	E.a.	Predito
γ	0.591	0.255	U_1	-0.0227	V_1	-0.0819
β_0	0.259	0.209	U_2	0.0483	V_2	0.1547
β_1	0.425	0.368	U_3	-0.0415	V_3	-0.1058
β_2	-0.013	0.013	U_4	-0.0070	V_4	-0.0319
σ_u^2	0.0064	0.0774	U_5	0.0236	V_5	0.0743
σ_v^2	0.0430	0.1315	U_6	-0.0008	V_6	-0.0094
ρ	0.7900	6.2616				
λ	0.00024	0.00031				
ν	1.346	0.292				

Tabela 5.3: Estimativas para o modelo de mistura com e.a.'s dependentes usando ligação cloglog na cura e função de sobrevivência basal Weibull

os respectivos intervalos de confiança ao 95% (linhas tracejadas) usando o método delta (Sen et al. [2010]). Porém, são omitidos estes intervalos devido a que são demasiado extensos dado que o tamanho amostral em cada clínica é demasiado pequeno. As mencionadas estimativas encontram-se na Figura 5.5. Note que as estimativas pontuais podem variar até 5% nos casos mais extremos da covariável idade.

Figura 5.5: *Fração de cura estimada para pacientes da clínica 2, diferentes perfis de Idade e (a) $X = 0$, (b) $X = 1$.*

Por outro lado, a Figura 5.6 apresenta as curvas de K-M e as curvas estimadas para o modelo escolhido. Note que a curva estimada para os dois perfis é próxima daquela estimada via K-M, mesmo levando em conta que há um número pequeno de observações.

Figura 5.6: *Função de sobrevivência estimada via K-M e pelo modelo escolhido para a clínica 2 e os dois perfis da variável recategorizada estágio do tumor: (a) $X = 0$, (b) $X = 1$.*

Finalmente, na Figura 5.8 apresenta-se a função de sobrevivência estimada para o modelo escolhido em cada clínica segundo os perfis $X = 0$ e $X = 1$.

O ajuste do modelo de mistura parece satisfatório. De fato, foram calculados os resíduos de Cox-Snell para o modelo escolhido. Se o modelo for adequado, então eles deveriam comportar-se como uma amostra censurada da distribuição exponencial de média 1. A Figura 5.7 apresenta o estimador de K-M para esses resíduos (com seus respectivos intervalos de confiança) e a curva teórica da distribuição exponencial de média 1. Baseados nessa análise, pode-se concluir que o modelo escolhido é adequado.

Porém, a desvantagem do uso deste modelo é que todas as interpretações são feitas apenas para a fração de cura e a função de sobrevivência dos indivíduos suscetíveis. Por este motivo, será

Figura 5.7: Curva de K-M para os resíduos de Cox-Snell do modelo escolhido.

Figura 5.8: Função de sobrevivência estimada em cada clínica pelo modelo escolhido para os dois perfis da variável recategorizada estágio do tumor: (a) $X = 0$, (b) $X = 1$.

utilizado este modelo como simples referência para os demais modelos que serão ajustados.

Do ponto de vista clássico, também será feito o ajuste do modelo de Chen com e.a.'s discutido na seção 3.2 e os modelos destrutivos de Poisson ponderado com e.a.'s discutidos na seção 4.2. No caso do modelo do Chen, foi utilizado o vetor $\mathbf{x}_{ij} = (\text{Tstage}_{ij})$ para explicar a sobrevivência dos suscetíveis e o vetor que explica a fração de cura foi considerado como $\mathbf{z}_{ij} = (1, z_{1ij}, z_{2ij}) = (1, \text{Tstage}_{ij}, \text{Idade}_{ij}^*)$, em que Idade_{ij}^* representa a Idade do paciente menos a mediana das idades observadas. Já no caso dos modelos Destrutivos, $\mathbf{x}_{ij}$ foi considerado como no caso anterior para todos os modelos.

Lamentavelmente, não há interação com os pesquisadores associados a estes dados. Eles poderiam ter uma ideia de qual variável está associada ao número inicial de células e qual à probabilidade de ativação. Ainda mais, também está a possibilidade de incorporar os efeitos aleatórios no parâmetro associado ao número inicial de células ou no parâmetro associado à probabilidade de ativação.

Desta forma, para os modelos destrutivos tiveram que ser consideradas as duas situações e escolher o melhor modelo usando algum critério de seleção de modelos. A combinação de covariáveis e efeitos aleatórios na cura considerada para cada modelo destrutivo pode ser vista na Tabela 5.4.

Caso	z_{1ij}	z_{2ij}	Parâmetro em que incluem-se os V_j 's
1	(1, Tstage $_{ij}$)	(Idade $^*_{ij}$)	θ
2	(Tstage $_{ij}$)	(1, Idade $^*_{ij}$)	p
3	(1, Idade $^*_{ij}$)	(Tstage $_{ij}$)	θ
4	(Idade $^*_{ij}$)	(1, Tstage $_{ij}$)	p

Tabela 5.4: Combinações de covariáveis e e.a.'s na cura consideradas para os modelos Destrutivos.

Em relação à distribuição basal $S_0(t | \lambda)$ foi considerada a distribuição exponencial segmentada com duas partições do eixo (em $a = 250$ como discutido no ajuste do modelo de mistura com e.a.'s), a distribuição Weibull e a distribuição exponencial. Por último, também foram considerados de forma separada os casos em que os e.a.'s são dependentes e independentes.

Como anteriormente, calculou-se os critérios cAIC para cada modelo ajustado. No caso dos modelos destrutivos e por uma questão de espaço, apresenta-se só estes valores para os modelos em que as covariáveis e os e.a.'s foram introduzidos no modelo como no caso 1 da Tabela 5.4. A Tabela 5.5 apresenta os critérios cAIC para os modelos considerados. Como no caso do modelo de mistura, os modelos em que foi considerada a distribuição Weibull para a distribuição basal $S_0(t | \lambda)$ apresentaram critérios cAIC mais baixos, mas neste caso privilegia-se os modelos em que os e.a.'s são independentes. Por outro lado, note que os valores de cAIC são muito próximos para os modelos de Chen, DLBP e DEWPM com e.a.'s independentes e distribuição basal Weibull.

Outro ponto importante é relacionado ao modelo DBN com e.a.'s. Para cada caso da Tabela 5.4, foi construída uma grade de valores para o parâmetro ϕ . Para cada valor de ϕ foram calculados os critérios cAIC e cBIC. A Figura 5.9 apresenta esses valores para o caso 1. Em todos os casos pode-se verificar um padrão similar. Note que a figura sugere que $\phi = 0$ minimiza ambos critérios, o que equivale a escolher o modelo DEWPM ao invés do modelo DBN. Por este motivo, dos modelos destrutivos discutidos na seção 2.2.4, só serão discutidos em detalhe os modelos DLBP e DEWPM.

Figura 5.9: Critérios cAIC e cBIC para o modelo DBN (caso 1).

As estimativas obtidas para os modelos de Chen, DLBP e DEWPM com e.a.'s são apresentadas

Modelo de Chen com e.a.'s						
$S_0(\cdot \lambda)$	ES	Weibull	Exponencial	ES	Weibull	Exponencial
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	717.30	716.30	721.26	719.90	717.99	728.33
cBIC	729.04	728.15	732.14	739.51	734.68	749.35
Modelo DLBP com e.a.'s (caso 1)						
$S_0(\cdot \lambda)$	ES	Weibull	Exponencial	ES	Weibull	Exponencial
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	717.31	716.25	721.30	726.40	716.30	731.64
cBIC	729.03	727.92	732.52	758.63	737.55	763.49
Modelo DEWPM com e.a.'s (caso 1)						
$S_0(\cdot \lambda)$	ES	Weibull	Exponencial	ES	Weibull	Exponencial
E.a. dependentes?	Não	Não	Não	Sim	Sim	Sim
cAIC	717.38	716.37	721.34	719.88	718.19	728.25
cBIC	729.05	728.09	732.15	739.41	735.11	749.13

Tabela 5.5: Critérios cAIC e cBIC para os modelos de Chen e Destrutivos ajustados.

na Tabela 5.6.

Note que, como esperado, as estimativas do modelo de Chen são muito próximas do modelo DEWPM nos parâmetros e efeitos que são comparáveis (como são $\gamma, \sigma_u^2, \lambda, \nu$ e os efeitos aleatórios $U_j, j = 1, \dots, J$). A seguir, apresentamos algumas formas em que se pode comparar de forma mais direta estes três modelos. Nestas comparações, foi incluído o modelo de mistura com ligação complementar log-log na cura e e.a.'s normais bivariados escolhido anteriormente.

A primeira forma de compará-los é através da função de sobrevivência populacional estimada, $S_{pop}(\cdot)$, para cada modelo. A Figura 5.10 apresenta essa função para a clínica 2 e diferentes combinações dos perfis do estágio do tumor e da variável idade centralizada $(-30, 0, 30)$. Note que em todos os casos as funções estimadas são próximas, mas elas ficam ainda mais próximas quando o perfil comparado na idade é próximo de zero (Figuras (b) e (e)). A diferença com o modelo de mistura parece ser basicamente por conta da diferença na estimativa pontual da cura em relação aos outros três modelos. De fato, a estimativa da fração de cura é outro ponto com o qual podemos comparar de forma mais direta os três modelos e que será abordada em breve.

Parâmetro	Modelo					
	Chen		DLBP		DEWPM	
	com e.a.'s		com e.a.'s (caso 1)		com e.a.'s (caso 1)	
	Est.	d.p.	Est.	d.p.	Est.	d.p.
γ	0.484	0.224	0.378	0.241	0.484	0.223
β_0	0.276	0.042	0.211	0.170	0.970	0.042
β_1	0.411	0.161	0.778	0.446	0.419	0.160
β_2	-0.008	0.0002	-0.012	0.0003	-0.012	0.0005
σ_u^2	0.0276	0.0367	0.0276	0.0359	0.0274	0.0364
σ_v^2	0.0389	0.0163	0.1097	0.1624	0.0371	0.0158
λ	0.0000602	0.0000708	0.0000796	0.0000809	0.0000605	0.0000706
ν	1.496	0.301	1.474	0.312	1.495	0.301
Efeito	Pred.		Pred.		Pred.	
U_1	0.0187	-	0.0197	-	0.0191	-
U_2	0.0071	-	0.0067	-	0.0070	-
U_3	-0.0883	-	-0.0895	-	-0.0878	-
U_4	0.0273	-	0.0285	-	0.0267	-
U_5	0.0147	-	0.0146	-	0.0148	-
U_6	0.0208	-	0.0198	-	0.0205	-
V_1	-0.0622	-	-0.0817	-	-0.0585	-
V_2	0.1285	-	0.2053	-	0.1249	-
V_3	-0.1081	-	-0.1784	-	-0.1031	-
V_4	-0.0318	-	-0.0649	-	-0.0322	-
V_5	0.0688	-	0.0928	-	0.0649	-
V_6	0.0048	-	0.0269	-	0.0039	-

Tabela 5.6: Estimativas obtidas para os modelos escolhidos usando o critério *cAIC* e *cBIC*.

(a)

(b)

(c)

(d)

(e)

(f)

Figura 5.10: Função de sobrevivência populacional estimada na clínica 2 para diferentes modelos de cura e diferentes perfis de pacientes: (a) $X = 0, Idade^* = -30$, (b) $X = 0, Idade^* = 0$, (c) $X = 0, Idade^* = 30$, (d) $X = 1, Idade^* = -30$, (e) $X = 1, Idade^* = 0$, (f) $X = 1, Idade^* = 30$.

Como segundo ponto de comparação, será utilizada a função de sobrevivência dos indivíduos suscetíveis, $S_{sus}(\cdot)$. No modelo de Chen, esta função é dada na equação (2.15), enquanto que nos modelos DLBP e DEWPM pode ser vista em (2.25) e (2.30) respectivamente. A função $S_{sus}(\cdot)$ estimada para cada modelo para vários perfis de pacientes pode ser vista na Figura 5.11. Note que todas as curvas estimadas são praticamente as mesmas, com exceção do caso quando a idade

centrada do paciente é um valor muito baixo, em que a curva do modelo de mistura difere um pouco das outras.

(a)

(b)

(c)

(d)

(e)

(f)

Figura 5.11: Função de sobrevivência estimada para os indivíduos suscetíveis da clínica 2 para diferentes modelos de cura e diferentes perfis de pacientes: (a) $X = 0, Idade^* = -30$, (b) $X = 0, Idade^* = 0$, (c) $X = 0, Idade^* = 30$, (d) $X = 1, Idade^* = -30$, (e) $X = 1, Idade^* = 0$, (f) $X = 1, Idade^* = 30$.

Uma outra forma direta de comparar estes modelos é através da fração de cura estimada. A Figura 5.12 apresenta as estimativas pontuais para a fração de cura de pacientes da clínica 2 nos modelos considerados.

Note que em torno da idade mediana as estimativas pontuais da fração de cura de todos os modelos são próximas, entanto que essas estimativas vão distanciando-se enquanto a idade aproxima-se dos valores extremos observados nessa variável. Contudo, as frações de cura estimadas pelos modelos de Chen, DLBP e DEWPM são muito próximas.

Figura 5.12: Fração de cura estimada na clínica 2 para os diferentes modelos ajustados: (a) $X = 0$, (b) $X = 1$.

Finalmente, uma última forma de comparar estes modelos é através dos resíduos de cada modelo. Os resíduos de Cox-Snell (equação (4.22)) foram calculados para cada modelo. Caso o ajuste seja adequado, esses resíduos deveriam comportar-se como uma a.a. censurada da distribuição exponencial de média 1. Uma forma de verificar este fato é plotar o estimador de K-M destes resíduos junto à função de sobrevivência teórica da distribuição exponencial de média 1. Caso o ajuste seja adequado, ambas curvas deveriam ser próximas. A Figura 5.13 apresenta essas curvas para cada modelo. As análises anteriores tem mostrado que as estimativas que fornecem os diferentes modelos são próximas, tanto para a função de sobrevivência populacional, dos indivíduos suscetíveis quanto para a fração de cura. Portanto, era esperado que todas as curvas de K-M na Figura 5.13 fossem próximas da função de sobrevivência teórica da Exponencial de média 1, embora descritivamente a curva do modelo DEWPM (curva (d)) parece apresentar um melhor ajuste.

Figura 5.13: Estimador de K-M para os resíduos de Cox-Snell para os modelos com e.a.'s ajustados: (a) Modelo de mistura, (b) Modelo de Chen, (c) Modelo DLBP, (d) DEWPM.

Todas estas análises concordam em que a qualidade de ajuste de todos os modelos discutidos são aceitáveis e concordantes entre eles. Por outro lado, um fato importante é que todos estes modelos contém exatamente o mesmo número de parâmetros estimados e e.a.'s preditos (com exceção do modelo de mistura em que foi escolhido o modelo com e.a.'s dependentes e portanto, devia ser estimado um parâmetro a mais, ρ).

Porém, como discutido na seção 2.2, os modelos destrutivos permitem obter informação adicional sobre aspectos que não podem fornecer informação os modelos de mistura e os modelos de Chen. No caso do modelo de mistura, não se pode obter informação nem do número inicial de células de cada indivíduo (D_{ij}), nem do número de células ativadas de cada indivíduo (M_{ij}), nem da probabilidade de ativação das células iniciais de cada indivíduo (p_{ij}). No modelo de Chen, assume-se que a probabilidade de ativação é 1 para todas as células, implicando que a estimativa do número de células iniciais é a mesma que a estimativa do número de células ativadas. Finalmente, nos modelos destrutivos pode-se obter informação sobre as três componentes mencionadas. A Figura 5.14 apresenta as estimativas para D_{ij} , M_{ij} e p_{ij} nos modelos de Chen, DLBP e DEWPM para alguns pacientes escolhidos ao acaso. Uma vez que foram estimados θ_{ij} e p_{ij} para cada indivíduo, pode-se usar os resultados em (2.26) e (2.31) para construir intervalos de predição para M_{ij} e D_{ij} .

Em relação ao gráfico (a) da Figura 5.14, note que as estimativas pontuais de M_{ij} dos modelos DLBP e DEWPM são muito próximas, entanto que o modelo de Chen tende a subestimar esta

quantidade em relação aos outros dois modelos. Note que a estimativa pontual do número inicial de células para o modelo de Chen no paciente 7 é muito próxima de zero, sendo conceitualmente inconsistente com o estudo pois os pacientes já tinham algum tipo de indício de possuir câncer de língua, portanto prever $\widehat{M}_{ij} \approx 0$ é equivalente a dizer que esse paciente não tinha este tipo de câncer. Por outro lado, os intervalos de confiança do modelo DLBP são muito mais amplos em relação aos intervalos do modelo DEWPM.

Do gráfico (b) da Figura 5.14, observa-se que as estimativas para o número de células ativadas são próximas para todos os modelos considerados, mas em geral os intervalos de confiança para D_{ij} são mais acurados para o modelo DLBP (no máximo iguais).

Por outro lado, os gráficos (c) e (d) da Figura 5.14 apresentam as probabilidades de ativação das células de cada indivíduo usando cada modelo. Note que os intervalos para p_{ij} são muito próximos nos modelos DLBP e DEWPM, e em contrapartida, muito distantes do valor assumido no modelo de Chen ($p_{ij} = 1$).

Figura 5.14: Diferentes componentes estimadas para os modelos de Chen, DLBP e DEWPM com e.a.'s. A bolinha indica a estimativa pontual e a linha tracejada a respectiva estimativa intervalar ao 95% de confiança. (a) Número de células iniciais, (b) Número de células ativadas, (c) Probabilidade de ativação de cada célula inicial, (d) Probabilidade de ativação de cada célula inicial (aproximação do gráfico).

(a)

(b)

(c)

Figura 5.15: Diferentes componentes estimadas para o modelos DLBP e DEWPM com e.a.'s em todos os pacientes do estudo. A bolinha indica a estimativa pontual e a linha tracejada a respectiva estimativa intervalar ao 95% de confiança. (a) Número de células iniciais, (b) Número de células ativadas, (c) Probabilidade de ativação de cada célula inicial.

Desta forma, os modelos DBLP e DEWPM fornecem praticamente os mesmos resultados e do ponto de vista clássico, pode-se usar indistintamente qualquer um. Porém, destaca-se um detalhe

em relação ao modelo DLBP. A Figura 5.16 apresenta as funções de verossimilhança perfilada para β_{11} e para V_1 nos modelos DLBP e DEWPM com e.a.'s, uma vez fixados os demais parâmetros e e.a.'s nos valores obtidos como estimativas (ver Tabela 5.6). Note que a função de verossimilhança é muito mais suave no modelo DLBP tanto para β_{11} quanto para V_1 . Para os demais efeitos aleatórios associados à cura ($V_2, \dots, V_6$), a forma desta função é similar. Do ponto de vista clássico, o único problema que isto constitui tem a ver com o fato de que o método de estimação pode cair em máximos locais ao invés do máximo global. Para contornar este problema, iniciou-se o algoritmo com diferentes valores iniciais e logo seleccionou-se aquele valor que produzia o valor máximo da função de verossimilhança. Mas, do ponto de vista bayesiano, a forma suave da verossimilhança pode produzir problemas de estimação mais graves como será visto na próxima seção. Basicamente, o problema vem do fato que desde um ponto de vista Bayesiano, a distribuição *a posteriori* calcula-se como a respectiva função de verossimilhança multiplicada pela distribuição *a priori*. Dessa forma, por exemplo para V_1 um valor de -4 terá uma probabilidade *a posteriori* muito similar a um valor próximo de 0, implicando uma grande variabilidade nas estimativas desse e.a. (na verdade de todos os e.a.'s associados à cura). Um efeito similar produz-se para β_{11} , embora mais leve. Note que nas funções de verossimilhança perfilada para β_{11} e V_1 esse efeito mantém-se só para os V_j 's, mas em um grau muito menor do que no modelo DLBP.

Figura 5.16: Função de log-verossimilhança perfilada para: (a) β_{11} no modelo DLBP com e.a.'s, (b) β_{11} no modelo DEWPM com e.a.'s, (c) V_1 no modelo DLBP com e.a.'s (d) e no modelo DEWPM com e.a.'s.

5.3 Ajuste bayesiano dos modelos

Do ponto de vista bayesiano, o custo computacional aumenta consideravelmente, tornando impraticável ajustar uma grande quantidade de modelos como na seção anterior. Desta forma, só serão ajustados os modelos DLBP e DEWPM com e.a.'s que obtiveram menores critérios cAIC e cBIC e discutidos com grande detalhe anteriormente.

As distribuições *a priori* para os parâmetros envolvidos utilizadas (como em 4.11) são as seguintes

$$\begin{aligned}\gamma &\sim N(0, 100), \\ \beta_1^F &\sim N(0, 100), \\ \beta_2^F &\sim N(0, 100), \\ \beta_k &\sim N(0, 10I_2), \\ \Sigma &\sim \mathcal{WI}_2(7, 0.2I_2),\end{aligned}$$

em que I_n denota a matriz identidade de dimensão n . Para α , a distribuição a priori usada é dada por

$$\alpha \sim \text{Gama}(\nu_1, \nu_2).$$

Foram utilizados três possíveis casos *a priori* para α , que são detalhados na Tabela 5.7, na qual são especificados os valores de ν_1 e ν_2 considerados, além da respectiva esperança e desvio padrão *a priori* para o parâmetro de precisão α e para o número de grupos k^* . Estes casos foram considerados em conjunto com o fato que k^* só pode assumir valores entre 1 e 6, pois este é o número de clínicas do estudo.

	ν_1	ν_2	$E(\alpha)$	$dp(\alpha)$	$E(k^*)$	$dp(k^*)$
Distribuição <i>a priori</i> 1	1.44	2.40	0.6	0.25	2.01	1.08
Distribuição <i>a priori</i> 2	1.78	4.00	4.05	9.19	4.00	1.35
Distribuição <i>a priori</i> 3	-	-	∞	-	-	-

Tabela 5.7: Distribuições a priori consideradas para α .

Como discutido na seção 2.4, o caso 3, em que é assumido $\alpha \rightarrow \infty$, equivale a assumir simplesmente

$$K_j^* | \beta_k, \Sigma \sim N(\beta_k, \Sigma), \quad j = 1, \dots, J.$$

Para cada caso, foram simuladas de forma independente três cadeias de tamanho final 10000 cada uma com salto igual a 10, após desconsiderar um período de *burn-in* de tamanho 50000. Para verificar a convergência das cadeias, calculou-se a estatística R de Gelman e Rubin [1992] para cada parâmetro e efeito aleatório do modelo através do pacote *coda* disponível na linguagem R Core Team [2012], os quais são apresentados na Tabela 5.8. Note que todos os valores da tabela são próximos de 1, indicando que as cadeias de cada parâmetro, para cada modelo e para cada distribuição *a priori* considerada para α convergiram.

Por outro lado, a Tabela 5.9 apresenta os fatores de Bayes para os modelos DLBP e DEWPM usando as diferentes distribuições *a priori* para α discutidas na Tabela 5.7. (Vide Basu e Chib [2003] para o cálculo do Fator de Bayes no contexto dos P.D.'s). Note que os resultados fornecem evidência substancial em favor dos modelos DLBP e DEWPM usando a distribuição *a priori* 3 para

Modelo Priori para α	DLBP			DEWPM		
	1	2	3	1	2	3
γ	1.0007	1.0009	1.0045	1.0011	1.0057	1.0009
β_{11}	1.0009	1.0010	1.0047	1.0034	1.0038	1.0009
β_{21}	1.0007	1.0019	1.0062	1.0019	1.0029	1.0008
U_1	1.0007	1.0046	1.0084	1.0030	1.0069	1.0008
U_2	1.0009	1.0019	1.0052	1.0023	1.0037	1.0004
U_3	1.0009	1.0033	1.0048	1.0039	1.0058	1.0009
U_4	1.0038	1.0064	1.0095	1.0038	1.0083	1.0005
U_5	1.0037	1.0075	1.0084	1.0032	1.0040	1.0006
U_6	1.0037	1.0087	1.0091	1.0027	1.0036	1.0011
V_1	1.0037	1.0060	1.0080	1.0019	1.0019	1.0010
V_2	1.0035	1.0048	1.0088	1.0034	1.0041	1.0003
V_3	1.0033	1.0034	1.0041	1.0026	1.0034	1.0008
V_4	1.0006	1.0035	1.0072	1.0033	1.0078	1.0010
V_5	1.0032	1.0070	1.0099	1.0036	1.0038	1.0008
V_6	1.0022	1.0051	1.0078	1.0073	1.0083	1.0030
$\mu_{\beta_{k1}}$	1.0012	1.0032	1.0080	1.0034	1.0066	1.0000
$\mu_{\beta_{k2}}$	1.0009	1.0042	1.0083	1.0030	1.0051	1.0000
Σ_{11}	1.0000	1.0027	1.0058	1.0022	1.0055	1.0000
Σ_{12}	1.0001	1.0040	1.0070	1.0040	1.0083	1.0000
Σ_{22}	1.0017	1.0054	1.0072	1.0007	1.0055	1.0001
α	1.0000	1.0000	1.0036	1.0043	1.0052	1.0000

Tabela 5.8: Valores de $\sqrt{\hat{R}}$ nos modelos ajustados.

α , i.e., assumindo a distribuição normal para os e.a.'s.

Modelo	Modelo Priori	DLBP	DEWPM	DLBP	DEWPM	DLBP	DEWPM
		1	1	2	2	3	3
DLBP	1	-29.30					
DEWPM	1	1.45	-27.85				
DLBP	2	3.46	2.01	-25.84			
DEWPM	2	3.57	2.12	0.11	-25.73		
DLBP	3	6.51	5.06	3.05	2.94	-22.79	
DEWPM	3	7.00	3.55	3.54	3.43	0.49	-22.30

Tabela 5.9: Fator de Bayes para os modelos ajustados. Na diagonal, os valores da função de log-verossimilhança marginais para cada modelo. Na parte abaixo da diagonal, o logaritmo do fator de Bayes do modelo da respectiva linha versus o modelo da respectiva coluna.

As estimativas dos parâmetros dos modelos DLBP e DEWPM usando a terceira distribuição *a priori* para α da Tabela 5.7 são dadas na Tabela 5.10.

	Modelo DLBP com e.a.'s					Modelo DEWPM com e.a.'s				
	média	mediana	d.p.	2.5%	97.5%	média	mediana	d.p.	2.5%	97.5%
γ	0.371	0.370	0.139	0.046	0.694	0.480	0.480	0.136	0.157	0.808
β_{11}	1.078	0.894	1.616	-2.377	4.431	0.271	0.267	0.453	-0.612	1.165
β_{21}	-0.014	-0.014	0.023	-0.061	0.031	-0.015	-0.015	0.025	-0.066	0.033
U_1	0.089	0.346	1.203	-3.520	1.665	-0.137	0.005	1.089	-2.517	1.636
U_2	-0.164	0.093	0.974	-3.112	0.989	-0.509	-0.427	0.775	-2.133	0.824
U_3	-0.502	-0.256	1.021	-3.625	0.789	-0.774	-0.658	0.890	-2.684	0.742
U_4	0.145	0.395	1.195	-3.462	1.677	-0.113	0.031	1.061	-2.451	1.591
U_5	-0.122	0.115	0.961	-3.222	0.989	-0.363	-0.240	0.774	-2.196	0.870
U_6	-0.074	0.199	1.065	-3.430	1.159	-0.282	-0.117	0.923	-2.422	1.135
V_1	-1.189	-1.140	2.020	-4.675	3.278	0.878	0.827	0.769	-0.542	2.563
V_2	-0.607	-0.464	2.195	-4.577	3.824	1.609	1.602	0.722	0.183	2.889
V_3	-1.033	-0.924	2.018	-4.665	3.305	1.087	1.005	0.664	-0.069	2.572
V_4	-1.206	-1.156	2.069	-4.726	3.303	0.949	0.901	0.743	-0.425	2.569
V_5	-0.763	-0.612	2.108	-4.579	3.523	1.401	1.378	0.674	0.096	2.752
V_6	-0.834	-0.629	2.012	-4.623	3.396	1.134	1.070	0.669	-0.076	2.639
$\mu_{\beta_{k1}}$	-0.108	0.111	0.991	-3.220	1.149	-0.366	-0.277	0.746	-2.056	0.886
$\mu_{\beta_{k2}}$	-0.943	-0.840	1.909	-4.285	3.293	1.177	1.147	0.600	0.052	2.431
Σ_{11}	0.026	0.022	0.017	0.010	0.067	0.025	0.021	0.016	0.010	0.064
Σ_{12}	0.000	0.000	0.011	-0.021	0.020	0.000	0.000	0.010	-0.020	0.020
Σ_{22}	0.026	0.022	0.017	0.010	0.068	0.025	0.021	0.016	0.010	0.065

Tabela 5.10: Estimativas obtidas para o modelo DLBP e DEWPM com e.a.'s normais usando a abordagem bayesiana.

Note que as estimativas de todas as componentes do modelo são muito próximas no modelo DEWPM com as obtidas no contexto clássico (Tabela 5.15). No modelo DLBP acontece algo similar, mas variando um pouco as estimativas de β_{11} e os V_j 's como esperado conforme à discussão em relação à função de verossimilhança perfilada de ambas componentes do modelo. Destaca-se a grande dispersão de ambas componentes nesse modelo.

Dado que as estimativas pontuais da abordagem bayesiana são próximas das obtidas na abordagem clássica para o modelo DEWPM, então as funções de sobrevivência populacional, de sobrevivência dos indivíduos suscetíveis e as frações de cura apresentadas nas Figuras 5.10, 5.11 e 5.12 são muito próximas das obtidas desde um ponto de vista Bayesiano, implicando que vale novamente o fato que ambos modelos fornecem estimativas similares para cada uma dessas componentes. Já no modelo DLBP obtém-se variações em relação à abordagem clássica.

Para as estimativas obtidas nos dois modelos ajustados, apresenta-se as funções de sobrevivência populacional estimadas para cada modelo. Note que, diferentemente da abordagem clássica, as duas funções variam para os diferentes perfis de pacientes aqui considerados.

Figura 5.17: Função de sobrevivência populacional da clínica 2 para os modelos DLBP e DEWPM com e.a.'s e diferentes perfis: (a) $X = 0, Idade^* = -30$, (b) $X = 0, Idade^* = 0$, (c) $X = 0, Idade^* = 30$, (d) $X = 1, Idade^* = -30$, (e) $X = 1, Idade^* = 0$, (f) $X = 1, Idade^* = 30$.

Por outro lado, a Figura 5.18 apresenta as estimativas para a fração de cura de pacientes na clínica 2 e diferentes valores para a idade. Note que neste caso as estimativas diferem um pouco para os dois perfis apresentados.

Figura 5.18: *Fração de cura estimada na clínica 2: (a) $X = 0$, (b) $X = 1$.*

Na Figura 5.19, são apresentados os intervalos de credibilidade de 95% para alguns pacientes escolhidos ao acaso para o número inicial de células (M_{ij}), número de células ativadas (D_{ij}) e probabilidade de ativação das células (p_{ij}). O efeito de ter grande variabilidade nas cadeias de β_{11} e os V_j 's são herdados pelas componentes latentes de M_{ij} e D_{ij} . Por este motivo, os intervalos para M_{ij} e D_{ij} são mais amplos no modelo DLBP com e.a.'s do que no modelo DEWPM com e.a.'s. No caso do parâmetro p_{ij} , este não é influenciado por esse efeito, pois não depende nem de β_{11} nem dos V_j 's.

Figura 5.19: Intervalos de credibilidade ao 95% para o número inicial de células, número de células ativadas e probabilidade de ativação das células para alguns pacientes.

Por outro lado, na Figura 5.20 apresenta-se a função de densidade estimada para cada parâmetro baseada na amostra *a posteriori* obtida. Note que a função de densidade estimada do parâmetro γ está muito concentrada ao redor de um valor para os dois modelos. Como já mencionado, a principal diferença produz-se no parâmetro β_{11} e nos e.a.'s V_j 's. Por todos os argumentos expostos, escolhe-se o modelo DEWPM com e.a.'s desde um ponto de vista Bayesiano.

Finalmente, a Tabela 5.11 apresenta os valores do critério DIC para os modelos DLBP e DEWPM com e.a.'s. Note que este critério apoia a escolha feita.

	DLBP com e.a.'s	DEWPM com e.a.'s
DIC	3719.89	3671.91

Tabela 5.11: Critério DIC para o modelos DLBP e DEWPM com e.a.'s.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figura 5.20: Função de densidade estimada baseada na amostra a posteriori. A linha contínua é para o modelo DLBP com e.a.'s e a linha tracejada é para o modelo DEWPM com e.a.'s.

Capítulo 6

Conclusões

6.1 Considerações finais

Neste trabalho foram estudados alguns modelos com fração de cura e efeitos aleatórios. No capítulo 2 foram apresentados alguns elementos necessários para o desenvolvimento desta tese.

A partir do capítulo 3 trabalha-se no contexto em que as observações são agrupadas. Estende-se o modelo de [Lopes e Bolfarine \[2012\]](#) incluindo efeitos aleatórios além da probabilidade de cura, na sobrevivência dos indivíduos suscetíveis. Estudou-se a estimação dos parâmetros do modelo via máxima verossimilhança restrita e exemplificou-se através de estudos de simulação que este processo de estimação fornece resultados satisfatórios. Além disso, estuda-se o impacto nas estimativas quando considerado erroneamente efeitos aleatórios independentes ao invés de efeitos aleatórios bivariados e vice-versa. Neste capítulo também apresentamos o modelo de mistura usando outras ligações na cura (além da logito) com a finalidade de dar uma maior flexibilidade ao modelo.

No capítulo 4 estendemos os modelos destrutivos ([Rodrigues et al. \[2011\]](#)) incluindo efeitos aleatórios nas mesmas duas componentes do modelo. Discute-se o fato de que a inclusão de efeitos aleatórios pode ser feita tanto no parâmetro associado ao número inicial de células quanto no parâmetro associado à probabilidade de ativação de cada célula, mas não em ambos simultaneamente. Além de uma abordagem clássica através de máxima verossimilhança restrita, discute-se uma abordagem bayesiana baseada em processos de Dirichlet ([Antoniak \[1974\]](#)), o que permite flexibilizar a distribuição dos efeitos aleatórios. Discute-se através de estudos de simulação o ganho do uso dos processos de Dirichlet neste contexto.

No capítulo 5 ilustramos os modelos discutidos nos capítulos 3 e 4, comparando os resultados obtidos a partir de cada modelo e destacando a vantagem do uso dos modelos destrutivos neste contexto. Discutimos separadamente as abordagens clássica e bayesiana.

6.2 Sugestões para pesquisas futuras

- Estudar e implementar o algoritmo EM nos modelos destrutivos com efeitos aleatórios discutidos no capítulo 4.
- Estudar o uso de outras medidas basais $\mathcal{G}_0$ mais gerais do que a distribuição normal. Por exemplo, pode-se utilizar a distribuição Skew-Normal que tem a vantagem de produzir distribuições *a posteriori* marginais fáceis de simular para todos seus parâmetros quando utilizadas certas distribuições *a priori*.
- Estender o modelo destrutivo com diferentes esquemas de ativação ([Cancho et. al \[2013a\]](#)) incluindo efeitos aleatórios no contexto em que as observações estejam agrupadas.

Apêndice A

Apêndice

A.1 Alguns aspectos dos modelos destrutivos

Nesta seção explicita-se alguns resultados dos modelos destrutivos.

A.1.1 Cálculo de $P(D = d \mid \theta, p, \phi)$

Apresenta-se os detalhes do cálculo de $P(D = d \mid \theta, p, \phi)$ em (2.22). Note que para $d = 0$, tem-se que

$$\begin{aligned} P(D = 0 \mid \theta, p, \phi) &= \sum_{j=0}^{\infty} P(D = 0 \mid M = j, p) P(M = j \mid \theta, \phi), \\ &= P(D = 0 \mid M = 0, p) P(M = 0 \mid \theta, \phi), \\ &+ \sum_{j=1}^{\infty} P(D = 0 \mid M = j, p) P(M = j \mid \theta, \phi) \\ &= \frac{w(0 \mid \phi) e^{-\theta}}{E_{\theta}[w(M \mid \phi)]} + \sum_{j=1}^{\infty} \binom{j}{0} p^0 (1-p)^{j-0} \frac{w(j \mid \phi) \theta^j e^{-\theta}}{j! E_{\theta}[w(M \mid \phi)]}, \\ &= \sum_{j=0}^{\infty} \frac{w(j \mid \phi) [\theta(1-p)]^j e^{-\theta}}{j! E_{\theta}[w(M \mid \phi)]}, \\ &= \frac{e^{-\theta p}}{E_{\theta}[w(M \mid \phi)]} \sum_{j=0}^{\infty} \frac{w(j \mid \phi) [\theta(1-p)]^j e^{-\theta(1-p)}}{j!}, \\ &= \frac{e^{-\theta p} E_{\theta(1-p)}[w(M \mid \phi)]}{E_{\theta}[w(M \mid \phi)]}. \end{aligned}$$

Analogamente, se $d > 0$

$$\begin{aligned} P(D = d \mid \theta, p, \phi) &= \sum_{j=d}^{\infty} P(D = d \mid M = j, p) P(M = j \mid \theta, \phi), \\ &= \sum_{j=d}^{\infty} \binom{j}{d} p^d (1-p)^{j-d} \frac{w(j \mid \phi) \theta^j e^{-\theta}}{E_{\theta}[w(M \mid \phi)] j!}, \\ &= \frac{(\theta p)^d e^{-\theta p}}{d! E_{\theta}[w(M \mid \phi)]} \sum_{j=d}^{\infty} \frac{w(j \mid \phi) [\theta(1-p)]^{j-d} e^{-\theta(1-p)}}{(j-d)!}, \\ &= \frac{(\theta p)^d e^{-\theta p}}{d! E_{\theta}[w(M \mid \phi)]} \sum_{i=0}^{\infty} \frac{w(i+d \mid \phi) [\theta(1-p)]^i e^{-\theta(1-p)}}{i!}, \\ &= \frac{(\theta p)^d e^{-\theta p} E_{\theta(1-p)}[w(M+d \mid \phi)]}{d! E_{\theta}[w(M \mid \phi)]}. \end{aligned}$$

Dessa forma,

$$P(D = d \mid \theta, p, \phi) = \frac{(\theta p)^d e^{-\theta p} E_{\theta(1-p)}[w(M + d \mid \phi)]}{d! E_{\theta}[w(M \mid \phi)]}, \quad d = 0, 1, 2, \dots$$

A.1.2 Calculando a distribuição de D, M e $D \mid M = m$ dada a informação observada

Modelo DLBP

Para simplificar a notação, defina $f^* = f(t \mid \boldsymbol{\lambda})$, $F^* = F(t \mid \boldsymbol{\lambda})$ e $S^* = S(t \mid \boldsymbol{\lambda})$. Note que

$$\begin{aligned} f(t, \delta \mid \boldsymbol{\lambda}, \theta, p) &= \sum_{d=0}^{\infty} f(t, \delta, d \mid \boldsymbol{\lambda}, \theta, p), \\ &= \sum_{d=0}^{\infty} [S^*]^{d-\delta} [df^*]^{\delta} \frac{(\theta p)^d e^{-\theta p}}{d!} \left\{ 1 - p + \frac{d}{\theta} \right\}, \\ &= \left[\frac{f^*}{S^*} \right]^{\delta} \left\{ (1-p)[\theta p S^*]^{\delta} e^{-\theta p} \sum_{d=\delta}^{\infty} \frac{[\theta p S^*]^{d-\delta}}{(d-\delta)!} + p S^* e^{-\theta p} \sum_{d=\delta}^{\infty} \frac{[\theta p S^*]^{d-1} d}{(d-\delta)!} \right\}, \\ &= \left[\frac{f^*}{S^*} \right]^{\delta} \left\{ (1-p)[\theta p S^*]^{\delta} e^{-\theta p} e^{\theta p S^*} + p S^* e^{-\theta p} \sum_{h=0}^{\infty} \frac{[\theta p S^*]^{h+\delta-1} (h+\delta)!}{h!} \right\}, \\ &= \left[\frac{f^*}{S^*} \right]^{\delta} \left\{ (1-p)[\theta p S^*]^{\delta} e^{\theta p F^*} \right. \\ &\quad \left. + [\theta p S^*]^{\delta} p S^* e^{-\theta p} \left(\sum_{h=1}^{\infty} \frac{[\theta p S^*]^{h-1}}{(h-1)!} + \delta [\theta p S^*]^{-1} \sum_{h=0}^{\infty} \frac{[\theta p S^*]^h}{h!} \right) \right\}, \\ &= [\theta p f^*]^{\delta} \left\{ (1-p) e^{-\theta p F^*} + p S^* e^{-\theta p F^*} + \frac{\delta e^{-\theta p F^*}}{\theta} \right\}, \\ &= [\theta p f^*]^{\delta} e^{-\theta p F^*} \left\{ (1-p) + p S^* + \frac{\delta}{\theta} \right\}. \end{aligned}$$

Desse jeito,

$$\begin{aligned} P(D = d, M = m \mid t, \delta) &= \frac{f(t, \delta, d, m)}{f(t, \delta)}, \\ &= \frac{[S^*]^{d-\delta} [df^*]^{\delta} \binom{m}{d} p^d (1-p)^{m-d} \frac{\theta^{m-1} e^{-\theta}}{(m-1)!}}{[\theta p f^*]^{\delta} e^{-\theta p F^*} \left\{ (1-p) + p S^* + \frac{\delta}{\theta} \right\}}, \\ &= \binom{m-\delta}{d-\delta} \left\{ \frac{p S^*}{1-p F^*} \right\}^{d-\delta} \left\{ \frac{1-p}{1-p F^*} \right\}^{m-d} \times \\ &\quad \times \left\{ \theta (1-p F^*) \right\}^{m-\delta} \frac{e^{-\theta(1-p F^*)}}{(m-\delta)!} \left\{ \frac{\delta/\theta}{1-p F^* + \delta/\theta} + \frac{m-\delta}{\theta(1-p F^* + \delta/\theta)} \right\}. \end{aligned}$$

Assim,

$$\begin{aligned} M - \delta \mid t, \delta &\sim DLBP \left(\theta [1 - p F(t \mid \boldsymbol{\lambda}) + \delta/\theta], \frac{1 - p F(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda}) + \delta/\theta} \right), \\ D - \delta \mid M = m, t, \delta &\sim Bin \left(m - \delta, \frac{p S(t \mid \boldsymbol{\lambda})}{1 - p F(t \mid \boldsymbol{\lambda})} \right). \end{aligned}$$

Note que $P(D = d \mid t, \delta)$ pode ser calculado da seguinte forma

$$\begin{aligned}
 P(D = d \mid t, \delta) &= \sum_{m=d}^{\infty} P(D = d, M = m \mid t, \delta), \\
 &= \sum_{m=d}^{\infty} \frac{1}{(d-\delta)!(m-d)!} (pS^*)^{d-\delta} (1-p)^{m-d} \theta^{m-\delta} e^{-\theta(1-pF^*)} \left\{ \frac{m}{1-pF^* + \delta/\theta} \right\}, \\
 &= \frac{(\theta pS^*)^{d-\delta} e^{-\theta(1-pF^*)}}{(d-\delta)!(1-pF^* + \delta/\theta)} \sum_{m=0}^{\infty} \frac{[\theta(1-p)]^m}{m!} (m+d), \\
 &= \frac{(\theta pS^*)^{d-\delta} e^{-\theta(1-pF^*)}}{(d-\delta)!(1-pF^* + \delta/\theta)} \left\{ \theta(1-p)e^{\theta(1-p)} + de^{\theta(1-p)} \right\}, \\
 &= \frac{(\theta pS^*)^{d-\delta} e^{-\theta pS^*}}{(d-\delta)!} \left\{ \frac{\theta pS^*}{\theta(1-pF^* + \delta/\theta)} + \frac{d-\delta}{\theta(1-pF^* + \delta/\theta)} \right\}.
 \end{aligned}$$

Portanto,

$$D - \delta \mid t, \delta \sim DLBP \left(\theta[1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta], \frac{pS(t \mid \boldsymbol{\lambda})}{1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta} \right).$$

Desse jeito, dada a informação observada, tem-se que a esperança condicional de $D \mid M = m$, M e D é

$$\begin{aligned}
 E(M \mid \mathbf{D}_{obs}) &= \delta + \frac{\{\theta[1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta] + 1\}(1 - pF(t \mid \boldsymbol{\lambda}))}{1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta}, \\
 E(D \mid \mathbf{D}_{obs}) &= \delta + \frac{\{\theta[1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta] + 1\}pS(t \mid \boldsymbol{\lambda})}{1 - pF(t \mid \boldsymbol{\lambda}) + \delta/\theta}, \\
 E(D \mid M = m, \mathbf{D}_{obs}) &= (m - \delta) \frac{p(1-p)S(t \mid \boldsymbol{\lambda})}{\{1 - pF(t \mid \boldsymbol{\lambda})\}^2}.
 \end{aligned}$$

Modelo DBN

Como anteriormente, defina $f^* = f(t \mid \boldsymbol{\lambda})$, $F^* = F(t \mid \boldsymbol{\lambda})$ e $S^* = S(t \mid \boldsymbol{\lambda})$. Assim,

$$\begin{aligned}
 f(t, \delta \mid \boldsymbol{\lambda}, \theta, p, \phi) &= \sum_{d=0}^{\infty} f(t, \delta \mid D = d) P(D = d), \\
 &= \sum_{d=0}^{\infty} S(t \mid \boldsymbol{\lambda})^{d-\delta} [df(t \mid \boldsymbol{\lambda})]^\delta \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})d!} \left(\frac{\phi\theta p}{1 + \phi\theta p} \right)^d (1 + \phi\theta p)^{-\phi^{-1}}, \\
 &= S(t \mid \boldsymbol{\lambda})^{-\delta} f(t \mid \boldsymbol{\lambda})^\delta (1 + \phi\theta p)^{-\phi^{-1}} \sum_{d=\delta}^{\infty} \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})(d-\delta)!} \left(\frac{\phi\theta pS(t \mid \boldsymbol{\lambda})}{1 + \phi\theta p} \right)^d.
 \end{aligned}$$

Definindo $R = \left(\frac{\phi\theta pS(t \mid \boldsymbol{\lambda})}{1 + \phi\theta p} \right)$, usando propriedades da função gama e da distribuição Binomial Negativa, tem-se que

- Se $\delta = 0$,

$$\sum_{d=0}^{\infty} \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})d!} R^d = (1 - R)^{-\phi^{-1}}.$$

- Se $\delta = 1$,

$$\begin{aligned}
\sum_{d=1}^{\infty} \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})(d-1)!} R^d &= \frac{R}{(1-R)^{\phi^{-1}}} \sum_{d=1}^{\infty} \frac{(\phi^{-1} + d - 1)\Gamma(\phi^{-1} + d - 1)}{\Gamma(\phi^{-1})(d-1)!} R^{d-1} (1-R)^{\phi^{-1}}, \\
&= \frac{R}{(1-R)^{\phi^{-1}}} \sum_{w=0}^{\infty} \frac{(\phi^{-1} + w)\Gamma(\phi^{-1} + w)}{\Gamma(\phi^{-1})w!} R^w (1-R)^{\phi^{-1}}, \\
&= \frac{R}{(1-R)^{\phi^{-1}}} (\phi^{-1} + E(W)) \quad (\text{em que } W \sim \text{BN}(\phi, R)), \\
&= \frac{R}{(1-R)^{\phi^{-1}}} \left(\phi^{-1} + \phi^{-1} \frac{R}{1-R} \right), \\
&= \frac{\phi^{-1} R}{(1-R)^{\phi^{-1}+1}}.
\end{aligned}$$

Portanto,

$$\begin{aligned}
f(t, \delta \mid \boldsymbol{\lambda}, \theta, p, \phi) &= S(t \mid \boldsymbol{\lambda})^{-\delta} f(t \mid \boldsymbol{\lambda})^{\delta} ((1 + \phi\theta p)(1-R))^{-\phi^{-1}} \left(\frac{\phi^{-1} R}{1-R} \right)^{\delta}, \\
&= (\theta p f(t \mid \boldsymbol{\lambda}))^{\delta} (1 + \phi\theta p F(t \mid \boldsymbol{\lambda}))^{-(\phi^{-1} + \delta)}.
\end{aligned}$$

Desse jeito,

$$\begin{aligned}
P(D = d, M = m \mid t, \delta, \boldsymbol{\lambda}, \theta, p, \phi) &= \frac{f(t, \delta, d, m \mid \boldsymbol{\lambda}, \theta, p, \phi)}{f(t, \delta \mid \boldsymbol{\lambda}, \theta, p, \phi)}, \\
&= \frac{[S^*]^{d-\delta} [df^*]^{\delta} \binom{m}{d} p^d (1-p)^{m-d} \frac{\Gamma(\phi^{-1} + d)}{\Gamma(\phi^{-1})d!} \left(\frac{\phi\theta}{1+\phi\theta} \right)^d (1 + \phi\theta)^{-\phi^{-1}}}{(\theta p f(t \mid \boldsymbol{\lambda}))^{\delta} (1 + \phi\theta p F(t \mid \boldsymbol{\lambda}))^{-(\phi^{-1} + \delta)}}, \\
&= \binom{m-\delta}{d-\delta} \left\{ \frac{pS^*}{1-pF^*} \right\}^{d-\alpha} \left\{ \frac{1-p}{1-pF^*} \right\}^{m-d} \times \\
&\quad \frac{\Gamma(\phi^{-1} + m)}{\Gamma(\phi^{-1} + \delta)(m-\delta)!} \left(\frac{\phi\theta(1-pF^*)}{1+\phi\theta} \right)^{m-\delta} \left(\frac{1+\phi\theta pF^*}{1+\phi\theta} \right)^{(\phi^{-1} + \delta)}.
\end{aligned} \tag{A.1}$$

Assim, conclui-se que

$$\begin{aligned}
M - \delta \mid t, \delta &\sim \text{BN} \left((\phi^{-1} + \delta)^{-1}, \left(\frac{\phi\theta(1-pF(t \mid \boldsymbol{\lambda}))}{1+\phi\theta} \right) \right), \\
D - \delta \mid M = m, t, \delta &\sim \text{Bin} \left(m - \delta, \frac{pS(t \mid \boldsymbol{\lambda})}{1-pF(t \mid \boldsymbol{\lambda})} \right).
\end{aligned}$$

Somando sobre $M \in \{d, d+1, d+2, \dots\}$, pode-se verificar também que

$$D - \delta \mid t, \delta \sim \text{BN} \left((\phi^{-1} + \delta)^{-1}, \left(\frac{\phi\theta p S(t \mid \boldsymbol{\lambda})}{1+\phi\theta} \right) \right).$$

Portanto, dada a informação observada, tem-se que

$$\begin{aligned} E(M \mid \mathbf{D}_{obs}) &= \frac{\delta(1 + \phi\theta) + \theta\{1 - pF(t \mid \boldsymbol{\lambda})\}}{1 + \phi\theta pF(t \mid \boldsymbol{\lambda})}, \\ E(D \mid \mathbf{D}_{obs}) &= \frac{\delta(1 + \phi\theta) + \theta pS(t \mid \boldsymbol{\lambda})}{1 + \phi\theta\{1 - pS(t \mid \boldsymbol{\lambda})\}}, \\ E(D \mid M = m, \mathbf{D}_{obs}) &= (m - \delta) \frac{p(1 - p)S(t \mid \boldsymbol{\lambda})}{\{1 - pF(t \mid \boldsymbol{\lambda})\}^2}. \end{aligned}$$

A.1.3 Provando que a fração de cura do modelo DBLP não é identificável quando z_1 e z_2 incluem intercepto

Sob as condições,

- $\mathbf{z}_{1i}^\top = (1, z_{1i})$;
- $\mathbf{z}_{2i}^\top = (1, z_{2i})$;
- z_{1i} e z_{2i} são variáveis binárias. Dessa forma, só existem 4 possíveis combinações para o par $(z_{1i}, z_{2i}) : (0, 0), (0, 1), (1, 0)$ e $(1, 1)$.

e seguindo (2.35), pode-se definir

$$\begin{aligned} \log \theta_0 &= \beta_{10}, \quad \log \theta_1 = \beta_{10} + \beta_{11}, \\ \log \left(\frac{p_0}{1 - p_0} \right) &= \beta_{20} \quad \text{e} \quad \log \left(\frac{p_1}{1 - p_1} \right) = \beta_{20} + \beta_{21}. \end{aligned}$$

A prova consistirá em mostrar $\theta_0^*, \theta_1^*, p_0^*$ e p_1^* que produzam a mesma fração de cura que θ_0, θ_1, p_0 e p_1 . Sem perda de generalidade, suponha que $p_1 \leq p_0$ e defina ξ que satisfaça

$$\exp\{-\min(\theta_0, \theta_1)p_0\}(1 - p_0) < \frac{1}{\xi} < 1. \quad (\text{A.2})$$

Note que θ_i , $i = 0, 1$ é a forma de simplificar $\theta(z_{1i})$, dado que só assume valores 0 e 1. O mesmo acontece com p_i . Assim, a fração de cura do modelo DBLP é dada por

$$\exp(-\theta_i p_j)(1 - p_j), \quad i = 0, 1; j = 0, 1.$$

Desta forma, requer-se que $\theta_0^*, \theta_1^*, p_0^*$ e p_1^* satisfaçam

$$\begin{aligned} \exp(-\theta_0^* p_0^*)(1 - p_0^*) &= \exp(-\theta_0 p_0)(1 - p_0), \\ \exp(-\theta_0^* p_1^*)(1 - p_1^*) &= \exp(-\theta_0 p_1)(1 - p_1), \\ \exp(-\theta_1^* p_0^*)(1 - p_0^*) &= \exp(-\theta_1 p_0)(1 - p_0), \\ \exp(-\theta_1^* p_1^*)(1 - p_1^*) &= \exp(-\theta_1 p_1)(1 - p_1). \end{aligned}$$

Fixe $p_0^* = 1/\xi$. Desta forma, o sistema de equações anterior fica

$$\begin{aligned} \exp(-\theta_0^*/\xi)(1 - 1/\xi) &= \exp(-\theta_0 p_0)(1 - p_0), \\ \exp(-\theta_0^* p_1^*)(1 - p_1^*) &= \exp(-\theta_0 p_1)(1 - p_1), \\ \exp(-\theta_1^*/\xi)(1 - 1/\xi) &= \exp(-\theta_1 p_0)(1 - p_0), \\ \exp(-\theta_1^* p_1^*)(1 - p_1^*) &= \exp(-\theta_1 p_1)(1 - p_1). \end{aligned}$$

Pode-se verificar que a solução do sistema anterior é

$$\begin{aligned}\theta_0^* &= -\xi[\log \xi + \log(1 - p_0) - \theta_0 p_0], \\ \theta_1^* &= -\xi[\log \xi + \log(1 - p_0) - \theta_1 p_0], \\ p_0^* &= \frac{1}{\xi}, \\ p_1^* &= \frac{p_1}{\xi p_0}.\end{aligned}\tag{A.3}$$

A condição em (A.2) garanta que θ_0^* e θ_1^* pertencem ao intervalo $(0, \infty)$, mesma condição que junto à suposição $p_1 \leq p_0$ garantam que p_0^* e p_1^* pertencem ao intervalo $(0, 1]$. Assim, pode-se isolar $\beta_{10}, \beta_{11}, \beta_{20}$ e β_{21} de (A.3) e portanto, para cada ξ que satisfaça (A.2) permite gerar um novo conjunto de parâmetros que gera as mesmas frações de cura que os betas originais.

A.2 Derivadas necessárias na implementação do algoritmo de Newton-Raphson

Nesta seção, apresenta-se as derivadas de primer e segundo ordem necessárias para implementar o algoritmo de Newton-Raphson nos diferentes modelos abordados neste trabalho.

A.2.1 Modelo de Chen com efeitos aleatórios

No modelo de Chen com efeitos aleatórios normais Bivariados, o vetor escore e a matriz $\mathbf{B}$, menos a segunda derivada de l em relação a $\boldsymbol{\zeta}$, são dados por

$$\begin{aligned}\frac{\partial l}{\partial \boldsymbol{\zeta}} &= \begin{bmatrix} \mathbf{X}^\top & 0 \\ 0 & \mathbf{Z}^\top \\ \mathbf{G}^\top & 0 \\ 0 & \mathbf{G}^\top \end{bmatrix} \begin{bmatrix} \partial l_1 / \partial \boldsymbol{\eta} \\ \partial l_1 / \partial \boldsymbol{\vartheta} \end{bmatrix} - \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 \\ 0 \\ \sigma_v^2 \mathbf{U} - \rho \sigma_u \sigma_v \mathbf{V} \\ \sigma_u^2 \mathbf{V} - \rho \sigma_u \sigma_v \mathbf{U} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{X}^\top & 0 \\ 0 & \mathbf{Z}^\top \\ \mathbf{G}^\top & 0 \\ 0 & \mathbf{G}^\top \end{bmatrix} \begin{bmatrix} -\partial^2 l_1 / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top & -\partial^2 l_1 / \partial \boldsymbol{\eta} \partial \boldsymbol{\vartheta}^\top \\ -\partial^2 l_1 / \partial \boldsymbol{\vartheta} \partial \boldsymbol{\eta}^\top & -\partial^2 l_1 / \partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top \end{bmatrix} \begin{bmatrix} \mathbf{X} & 0 & \mathbf{G} & 0 \\ 0 & \mathbf{Z} & 0 & \mathbf{G} \end{bmatrix} \\ &\quad + \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_v^2 I_J & -\rho \sigma_u \sigma_v I_J \\ 0 & 0 & -\rho \sigma_u \sigma_v I_J & \sigma_u^2 I_J \end{bmatrix},\end{aligned}$$

com $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{U}$ e $\boldsymbol{\vartheta} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{G}\mathbf{V}$. I_J é a matriz identidade de dimensão J e $\mathbf{X}, \mathbf{Z}$ e $\mathbf{G}$ são as matrizes de desenho de $\boldsymbol{\gamma}, \boldsymbol{\beta}$ e dos efeitos aleatórios $\mathbf{U}$ and $\mathbf{V}$ respectivamente. Seja $H_0(t | \boldsymbol{\lambda})$ o

risco basal acumulado. A primeira e segunda derivada de l_1 em relação a $\boldsymbol{\eta}$ and $\boldsymbol{\vartheta}$ são dados por:

$$\begin{aligned} \frac{\partial l_1}{\partial \eta_i} &= \delta_i(1 - H_0(t_i | \boldsymbol{\lambda})e^{\eta_i}) - e^{\vartheta_i + \eta_i} H_0(t_i | \boldsymbol{\lambda})[S_0(t_i | \boldsymbol{\lambda})]^{\eta_i}, \\ -\frac{\partial^2 l_1}{\partial \eta_i^2} &= H_0(t_i | \boldsymbol{\lambda}) \left[\delta_i e^{\eta_i} + e^{\vartheta_i + \eta_i} [S_0(t_i | \boldsymbol{\lambda})]^{\eta_i} (1 - H_0(t_i | \boldsymbol{\lambda})e^{\eta_i}) \right], \\ \frac{\partial l_1}{\partial \vartheta_i} &= \delta_i - e^{\vartheta_i} (1 - [S_0(t_i | \boldsymbol{\lambda})]^{\eta_i}), \\ -\frac{\partial^2 l_1}{\partial \vartheta_i^2} &= e^{\vartheta_i} (1 - [S_0(t_i | \boldsymbol{\lambda})]^{\eta_i}), \\ -\frac{\partial^2 l_1}{\partial \eta_i \vartheta_i} &= e^{\vartheta_i + \eta_i} H_0(t_i) [S_0(t_i | \boldsymbol{\lambda})]^{\eta_i}, \\ -\frac{\partial^2 l_1}{\partial \eta_i \eta_j} &= 0, \quad -\frac{\partial^2 l_1}{\partial \vartheta_i \vartheta_j} = 0, \quad -\frac{\partial^2 l_1}{\partial \eta_i \vartheta_j} = 0 \quad \text{se } i \neq j. \end{aligned}$$

Para estimar $S_0(t | \boldsymbol{\lambda})$, considere em um primeiro caso que $S_0(t | \boldsymbol{\lambda}) = \exp\{-\lambda t^\nu\}$. Defina $\theta_{ij} = \exp\{\vartheta_{ij}\}$ e $S_{ij} = [S_0(t_{ij} | \boldsymbol{\lambda})]^{\exp \eta_{ij}}$. Desta forma, a log-verossimilhança perfilada para o vetor $\boldsymbol{\lambda}$ é dada por

$$l_p = \sum_{j=1}^J \sum_{i=1}^{n_j} \left[\delta_{ij} \{ \log \lambda + \log \nu + \nu \log t_{ij} - \lambda e^{\eta_{ij}} t_{ij}^\nu \} + \theta_{ij} \exp\{-\lambda e^{\eta_{ij}} t_{ij}^\nu\} \right].$$

Os elementos do vetor $\partial l_p / \partial \boldsymbol{\lambda}$ são dados por

$$\begin{aligned} \frac{\partial l_p}{\partial \lambda} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left[\delta_{ij} \{ \lambda^{-1} - e^{\eta_{ij}} t_{ij}^\nu \} - \theta_{ij} e^{\eta_{ij}} t_{ij}^\nu S_{ij} \right], \\ \frac{\partial l_p}{\partial \nu} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left[\delta_{ij} \{ \nu^{-1} + \log t_{ij} - \lambda e^{\eta_{ij}} t_{ij}^\nu \log t_{ij} \} - \lambda \theta_{ij} e^{\eta_{ij}} t_{ij}^\nu S_{ij} \log t_{ij} \right], \end{aligned}$$

e os elementos da matriz $\mathbf{C}$ em (3.25) são dados por

$$\begin{aligned} C_{11} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left[-\frac{\delta_{ij}}{\lambda^2} + \theta_{ij} (e^{\eta_{ij}} t_{ij}^\nu)^2 S_{ij} \right], \\ C_{12} &= -\sum_{j=1}^J \sum_{i=1}^{n_j} e^{\eta_{ij}} t_{ij}^\nu \log t_{ij} \left[\delta_{ij} + \theta_{ij} S_{ij} (1 - \lambda t_{ij}^\nu e^{\eta_{ij}}) \right], \\ C_{22} &= -\sum_{j=1}^J \sum_{i=1}^{n_j} \left[\delta_{ij} (\nu^{-2} + \lambda e^{\eta_{ij}} t_{ij}^\nu \log^2 t_{ij}) + \lambda \theta_{ij} e^{\eta_{ij}} t_{ij}^\nu S_{ij} \log^2 t_{ij} (1 - \lambda t_{ij}^\nu e^{\eta_{ij}}) \right]. \end{aligned}$$

Se for considerado que $S_0(t | \boldsymbol{\lambda}) = \exp\left\{-\sum_{l=1}^L \lambda_l \nabla_l(t)\right\}$, em que $\nabla_l(t)$ é definido na equação (2.5), então a log-verossimilhança perfilada para o vetor $\boldsymbol{\lambda}$ é dada por

$$l_p = \sum_{j=1}^J \sum_{i=1}^{n_j} \delta_{ij} \left\{ \sum_{l=1}^L [I(a_{l-1} < t_{ij} \leq a_l) \log \lambda_l - \lambda_l \nabla_l(t_{ij}) e^{\eta_{ij}}] + \theta_{ij} \exp \left[-\sum_{l=1}^L \lambda_l \nabla_l(t_{ij}) \right] \right\}.$$

Neste caso, os elementos do vetor $\partial l_p / \partial \boldsymbol{\lambda}$ são dados por

$$\frac{\partial l_p}{\partial \lambda_l} = \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \delta_{ij} \left\{ I(a_{l-1} < t_{ij} \leq a_l) \lambda_l^{-1} - \nabla_l(t_{ij}) e^{\eta_{ij}} \right\} - \theta_{ij} S_{ij} \nabla_l(t_{ij}) e^{\eta_{ij}} \right\}, \quad l = 1, \dots, L.$$

Os elementos da matriz $\mathbf{C}$ em (3.25) são dados por

$$C_{ll} = -\lambda_l^{-2} \sum_{j=1}^J \sum_{i=1}^{n_j} \delta_{ij} I(a_{l-1} < t_{ij} \leq a_l) + \sum_{j=1}^J \sum_{i=1}^{n_j} \theta_{ij} S_{ij} \nabla_l^2(t_{ij}) e^{2\eta_{ij}}, \quad l = 1, \dots, L,$$

$$C_{lk} = 0, \quad 1 \leq l, k \leq L, \quad l \neq k.$$

A.2.2 Modelo de mistura com efeitos aleatórios

A Tabela A.1 apresentada as f.d.a. da distribuição Logística, Normal e de Gumbel e as respectivas primeiras e segundas derivadas.

Distribuição	$\psi(u)$	$\psi'(u)$	$\psi''(u)$
Logística	$\frac{e^u}{1+e^u}$	$\frac{e^u}{(1+e^u)^2}$	$\frac{e^u(1-e^u)}{(1+e^u)^3}$
Normal	$\Phi(u)$	$\phi(u)$	$-u\phi(u)$
Gumbel	e^{-e^u}	$-e^{u-e^u}$	$e^{u-e^u}(e^u - 1)$

Tabela A.1: F.d.a. das distribuições Logística, Normal e de Gumbel e suas derivadas de primeiro e segundo ordem.

No modelo de Berkson e Gage com efeitos aleatórios bivariados e outras ligações na cura, o vetor escore e a matriz $\mathbf{B}$, menos a segunda derivada de l em relação a $\boldsymbol{\zeta}$, são dados por

$$\frac{\partial l^a}{\partial \boldsymbol{\zeta}} = \begin{bmatrix} \mathbf{X}^\top & 0 \\ 0 & \mathbf{Z}^\top \\ \mathbf{G}^\top & 0 \\ 0 & \mathbf{G}^\top \end{bmatrix} \begin{bmatrix} \partial l_1^a / \partial \boldsymbol{\eta} \\ \partial l_1^a / \partial \boldsymbol{\vartheta} \end{bmatrix} - \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 \\ 0 \\ \sigma_v^2 \mathbf{U} - \rho \sigma_u \sigma_v \mathbf{V} \\ \sigma_u^2 \mathbf{V} - \rho \sigma_u \sigma_v \mathbf{U} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{X}^\top & 0 \\ 0 & \mathbf{Z}^\top \\ \mathbf{G}^\top & 0 \\ 0 & \mathbf{G}^\top \end{bmatrix} \begin{bmatrix} -\partial^2 l_1^a / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top & -\partial^2 l_1^a / \partial \boldsymbol{\eta} \partial \boldsymbol{\vartheta}^\top \\ -\partial^2 l_1^a / \partial \boldsymbol{\vartheta} \partial \boldsymbol{\eta}^\top & -\partial^2 l_1^a / \partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top \end{bmatrix} \begin{bmatrix} \mathbf{X} & 0 & \mathbf{G} & 0 \\ 0 & \mathbf{Z} & 0 & \mathbf{G} \end{bmatrix}$$

$$+ \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_v^2 I_J & -\rho \sigma_u \sigma_v I_J \\ 0 & 0 & -\rho \sigma_u \sigma_v I_J & \sigma_u^2 I_J \end{bmatrix},$$

com $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{U}$ e $\boldsymbol{\vartheta} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{G}\mathbf{V}$. I_J é a matriz identidade de dimensão J e $\mathbf{X}$, $\mathbf{Z}$ e $\mathbf{G}$ são as matrizes de desenho de $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ e dos efeitos aleatórios $\mathbf{U}$ and $\mathbf{V}$ respectivamente. Para simplificar a notação, defina

$$\pi_i = \psi(\vartheta_i), \quad \pi'_i = \psi'(\vartheta_i), \quad \pi''_i = \psi''(\vartheta_i), \quad 1 \leq i \leq \sum_{j=1}^J n_j,$$

em que ψ, ψ' e ψ'' corresponde a alguma das distribuições na tabela A.1. As derivadas em relação

a $\boldsymbol{\eta}$ e $\boldsymbol{\vartheta}$ são dadas por

$$\begin{aligned} \frac{\partial l_1^a}{\partial \eta_i} &= \tilde{y}_i [\delta_i - e^{\eta_i} H_0(t_i | \boldsymbol{\lambda})], \\ -\frac{\partial^2 l_1^a}{\partial \eta_i^2} &= \tilde{y}_i e^{\eta_i} H_0(t_i | \boldsymbol{\lambda}), \\ \frac{\partial l_1^a}{\partial \vartheta_i} &= \tilde{y}_i \frac{\pi'_i}{\pi_i} + (1 - \tilde{y}_i) \frac{\pi'_i}{(1 - \pi_i)}, \\ -\frac{\partial^2 l_1}{\partial \vartheta_i^2} &= -\tilde{y}_i \left\{ \frac{\pi_i \pi''_i - \pi_i'^2}{\pi_i^2} \right\} + (1 - \tilde{y}_i) \left\{ \frac{(1 - \pi_i) \pi''_i + \pi_i'^2}{(1 - \pi_i)^2} \right\}, \\ -\frac{\partial^2 l_1}{\partial \eta_i \vartheta_i} &= 0, \quad \text{se } 1 \leq i \leq \sum_{j=1}^J n_j, \\ -\frac{\partial^2 l_1}{\partial \eta_i \eta_j} &= 0, -\frac{\partial^2 l_1}{\partial \vartheta_i \vartheta_j} = 0, \quad -\frac{\partial^2 l_1}{\partial \eta_i \vartheta_j} = 0, \quad \text{se } 1 \leq i \neq j \leq \sum_{j=1}^J n_j. \end{aligned}$$

Para estimar $S_0(\cdot | \boldsymbol{\lambda})$, considere primeiro que $S_0(\cdot | \boldsymbol{\lambda})$ provém da distribuição Weibull. Desta forma, a log-verossimilhança perfilada para o vetor $\boldsymbol{\lambda}$ é dada por

$$l_p^a = \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} \left[\delta_{ij} \{ \log \lambda + \log \nu + (\nu - 1) \log t_{ij} \} - \lambda e^{\eta_{ij}} t_{ij}^\nu \right].$$

Assim, os elementos do vetor escore em (3.28) são dados por

$$\begin{aligned} \frac{\partial l_p^a}{\partial \lambda} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} \left\{ \frac{\delta_{ij}}{\lambda} + e^{\eta_{ij}} t_{ij}^\nu \right\} \\ \frac{\partial l_p^a}{\partial \nu} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} \left\{ \delta_{ij} \left(\frac{1}{\nu} + \log t_{ij} \right) + \lambda t_{ij}^\nu \log t_{ij} \right\} \end{aligned}$$

e os elementos da matriz $\mathbf{C}$ são dados por

$$\begin{aligned} C_{11} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \frac{\tilde{y}_{ij} \delta_{ij}}{\lambda^2}, \\ C_{12} &= - \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} t_{ij}^\nu \log t_{ij}, \\ C_{22} &= \sum_{j=1}^J \sum_{i=1}^{n_j} \left\{ \frac{\tilde{y}_{ij} \delta_{ij}}{\nu^2} - \lambda t_{ij}^\nu \log^2 t_{ij} \right\}. \end{aligned}$$

Enquanto que se $S_0(\cdot | \boldsymbol{\lambda})$ provém da distribuição Exponencial Segmentada, então a log-verossimilhança perfilada para $\boldsymbol{\lambda}$ é

$$l_p^a = \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} \sum_{l=1}^L \left[\delta_{ij} I(a_{l-1} < t_{ij} \leq a_l) - e^{\eta_{ij}} \lambda_l \nabla_l(t_{ij}) \right].$$

Assim, o vetor escore em (3.28) é dado por

$$\frac{\partial l_p^a}{\partial \lambda_l} = \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{y}_{ij} \left\{ \frac{\delta_{ij}}{\lambda_l} I(a_{l-1} < t_{ij} \leq a_l) - e^{\eta_{ij}} \nabla_l(t_{ij}) \right\}, \quad l = 1, \dots, L,$$

e os elementos da matriz $\mathbf{C}$ são dados por

$$C_{ll} = \sum_{j=1}^J \sum_{i=1}^{n_j} \frac{\tilde{y}_{ij} \delta_{ij}}{\lambda_l^2} I(a_{l-1} < t_{ij} \leq a_l), \quad l = 1, \dots, L,$$

$$C_{jl} = 0, \quad 1 \leq j \neq l \leq L.$$

A.2.3 Modelos Destrutivos com efeitos aleatórios

Nos modelos destrutivos de Poisson Ponderado com efeitos aleatórios Normais bivariados, o vetor escore e a matriz $\mathbf{B}$, menos a segunda derivada de l em relação a $\boldsymbol{\zeta}$, dependem do parâmetro no qual são introduzidos os efeitos aleatórios associados à cura (equações 4.2 e 4.3). Para o caso em que os V_j 's são introduzidos através do parâmetro θ , tem-se que

$$\frac{\partial l}{\partial \boldsymbol{\zeta}} = \begin{bmatrix} \mathbf{X}^\top & 0 & 0 \\ 0 & \mathbf{Z}_1^\top & 0 \\ 0 & 0 & \mathbf{Z}_2^\top \\ \mathbf{G}^\top & 0 & 0 \\ 0 & \mathbf{G}^\top & 0 \end{bmatrix} \begin{bmatrix} \partial l_1 / \partial \boldsymbol{\eta} \\ \partial l_1 / \partial \boldsymbol{\vartheta}_1 \\ \partial l_1 / \partial \boldsymbol{\vartheta}_2 \end{bmatrix} - \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \sigma_v^2 \mathbf{U} - \rho \sigma_u \sigma_v \mathbf{V} \\ \sigma_u^2 \mathbf{V} - \rho \sigma_u \sigma_v \mathbf{U} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{X}^\top & 0 & 0 \\ 0 & \mathbf{Z}_1^\top & 0 \\ 0 & 0 & \mathbf{Z}_2^\top \\ \mathbf{G}^\top & 0 & 0 \\ 0 & \mathbf{G}^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{-\partial^2 l_1}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\eta} \partial \boldsymbol{\vartheta}_1^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\eta} \partial \boldsymbol{\vartheta}_2^\top} \\ \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_1 \partial \boldsymbol{\eta}^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_1 \partial \boldsymbol{\vartheta}_1^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_1 \partial \boldsymbol{\vartheta}_2^\top} \\ \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_2 \partial \boldsymbol{\eta}^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_2 \partial \boldsymbol{\vartheta}_1^\top} & \frac{-\partial^2 l_1}{\partial \boldsymbol{\vartheta}_2 \partial \boldsymbol{\vartheta}_2^\top} \end{bmatrix} \begin{bmatrix} \mathbf{X} & 0 & 0 & \mathbf{G} & 0 \\ 0 & \mathbf{Z}_1 & 0 & 0 & \mathbf{G} \\ 0 & 0 & \mathbf{Z}_2 & 0 & 0 \end{bmatrix}$$

$$+ \frac{1}{\sigma_u^2 \sigma_v^2 (1 - \rho^2)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_v^2 I_J & -\rho \sigma_u \sigma_v I_J \\ 0 & 0 & -\rho \sigma_u \sigma_v I_J & \sigma_u^2 I_J \end{bmatrix},$$

em que $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{U}$, $\boldsymbol{\vartheta}_1 = \mathbf{Z}_1\boldsymbol{\beta}_1 + \mathbf{G}\mathbf{V}$ e $\boldsymbol{\vartheta}_2 = \mathbf{Z}_2\boldsymbol{\beta}_2$. I_J é a matriz identidade de dimensão J e $\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2$ e $\mathbf{G}$ são as matrizes de desenho de $\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ e dos efeitos aleatórios $\mathbf{U}$, e $\mathbf{V}$ respectivamente.

Para o caso em que os V_j 's são introduzidos através do parâmetro p , só precisa-se modificar a matriz de desenho nas expressões anteriores pela seguinte matriz

$$\begin{bmatrix} \mathbf{X}^\top & 0 & 0 \\ 0 & \mathbf{Z}_1^\top & 0 \\ 0 & 0 & \mathbf{Z}_2^\top \\ \mathbf{G}^\top & 0 & 0 \\ 0 & 0 & \mathbf{G}^\top \end{bmatrix}.$$

Além disso, os vetores $\boldsymbol{\vartheta}_1$ e $\boldsymbol{\vartheta}_2$ são definidos como $\boldsymbol{\vartheta}_1 = \mathbf{Z}_1\boldsymbol{\beta}_1$ e $\boldsymbol{\vartheta}_2 = \mathbf{Z}_2\boldsymbol{\beta}_2 + \mathbf{G}\mathbf{V}$.

As componentes dos vetores $\partial l_1 / \partial \boldsymbol{\eta}$, $\partial l_1 / \partial \boldsymbol{\vartheta}_1$ e $\partial l_1 / \partial \boldsymbol{\vartheta}_2$ e as respectivas segundas derivadas dependem de cada modelo Destrutivo considerado. Por conta do espaço, serão proporcionadas só as derivadas de primer ordem. Tanto estas como as segundas derivadas estão implementadas no programa R ([R Core Team \[2012\]](#)). Para mais detalhes, ver o apêndice A.3.

Para simplificar a notação, defina $S_{ij} = S(t_{ij} | \boldsymbol{\lambda}, \eta_{ij})$, $F_{ij} = 1 - S_{ij}$ e $H_{0ij} = -\log S_0(t_{ij} | \boldsymbol{\lambda})$.

- Modelo DBLP.

Defina ainda $b_{ij} = H_{0ij} \exp \eta_{ij}$, $c_{ij} = (1 - p_{ij} F_{ij})^{-1}$ e $d_{ij} = (\theta_{ij} + c_{ij})^{-1}$. Então,

$$\begin{aligned}\partial l_1 / \partial \eta_{ij} &= \delta_{ij}(1 - b_{ij} + d_{ij} c_{ij}^2 p_{ij} b_{ij} S_{ij}) - b_{ij} p_{ij} S_{ij} d_{ij}^{-1}, \\ \partial l_1 / \partial \vartheta_{1ij} &= \theta_{ij}(\delta_{ij} d_{ij} - p_{ij} F_{ij}), \\ \partial l_1 / \partial \vartheta_{2ij} &= \delta_{ij}(1 - p_{ij} + d_{ij} c_{ij}^2 p_{ij}(1 - p_{ij}) F_{ij}) - p_{ij}(1 - p_{ij}) F_{ij} d_{ij}^{-1}.\end{aligned}$$

- Modelo DEWPM.

$$\begin{aligned}\partial l_1 / \partial \eta_{ij} &= \delta_{ij}(1 - H_{0ij} e^{\eta_{ij}}) - \theta_{ij} p_{ij} e^{\phi} H_{0ij} e^{\eta_{ij}} S_{ij}, \\ \partial l_1 / \partial \vartheta_{1ij} &= \delta_{ij} - \theta_{ij} p_{ij} e^{\phi} F_{ij}, \\ \partial l_1 / \partial \vartheta_{2ij} &= -\theta_{ij} p_{ij}(1 - p_{ij}) e^{\phi} F_{ij} + \delta_{ij}(1 - p_{ij}).\end{aligned}$$

- Modelo DBN.

$$\begin{aligned}\partial l_1 / \partial \eta_{ij} &= \delta_{ij}(1 - H_{0ij} e^{\eta_{ij}}) - (1 + \phi \delta_{ij}) \theta_{ij} p_{ij} H_{0ij} e^{\eta_{ij}} S_{ij} (1 + \phi \theta_{ij} p_{ij} F_{ij})^{-1}, \\ \partial l_1 / \partial \vartheta_{1ij} &= \delta_{ij} - (1 + \phi \delta_{ij}) \theta_{ij} p_{ij} F_{ij} (1 + \phi \theta_{ij} p_{ij} F_{ij})^{-1}, \\ \partial l_1 / \partial \vartheta_{2ij} &= \delta_{ij}(1 - p_{ij}) - (1 + \phi \delta_{ij}) \theta_{ij} p_{ij} (1 - p_{ij}) F_{ij} (1 + \phi \theta_{ij} p_{ij} F_{ij})^{-1}.\end{aligned}$$

A.3 Programas em R

Nesta seção, disponibilizamos alguns dos programas utilizados na implementação dos modelos desenvolvidos neste trabalho.

A.3.1 Estimação clássica dos modelos de Chen e dos modelos destrutivos com efeitos aleatórios bivariados Normais.

```
library(nlme)
library(msm)
#####
#####
#### dist=1: Distribuição Weibull para S0 #####
#### dist=2: Distribuição ES para S0 #####
#####
#### modelo=0 : Modelo de Chen #####
#### modelo=1 : Modelo destrutivo com viés de duração #####
#### modelo=2 : Modelo destrutivo com pesos exponenciais #####
#### modelo=3 : Modelo destrutivo Binomial Negativo #####
#####
#### dep=0: efeitos aleatórios independentes #####
#### dep=1: efeitos aleatórios dependentes #####
#####

logvero.zetab<-function(zetab)
{
if(modelo==0)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r)]
```

```

gamma<-matrix(gamma,ncol=1)
beta<-matrix(beta1,ncol=1)
U<-zetab[(q+r+1):(q+r+J)]
V<-zetab[(q+r+J+1):(q+r+2*J)]
}
if(modelo==1 || modelo==2 || modelo==3)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r1)]
beta2<-zetab[(q+r1+1):(q+r1+r2)]
gamma<-matrix(gamma,ncol=1)
beta1<-matrix(beta1,ncol=1)
beta2<-matrix(beta2,ncol=1)
U<-zetab[(q+r1+r2+1):(q+r1+r2+J)]
V<-zetab[(q+r1+r2+J+1):(q+r1+r2+2*J)]
}
if(dist==1)
{
lambda<-lambdab[1]
nu<-lambdab[2]
S0<-exp(-lambda*t^nu)
h0<-lambda*nu*(t^(nu-1))
H0<-lambda*t^nu
}
if(dist==2)
{
S0<-1-ppexp(t,rate=lambdab,t=c(0,as.vector(a)))
h0<-dpexp(t,rate=lambdab,t=c(0,as.vector(a)))/S0
H0<--log(S0)
}
xi<-t(x)%*%gamma+U[clinica]
S<-exp(-H0*exp(xi))
F<-1-S
f<-h0*exp(xi)*S
f<-ifelse(f==0,1e-300,f)
if(modelo==0)
{
theta=exp(t(z)%*%beta+V[clinica])
p=rep(1,length(t))
}
if(modelo==1 || modelo==2 || modelo==3)
{
if(forma==1)
{
theta=exp(t(z1)%*%beta1+V[clinica])
p=exp(t(z2)%*%beta2)/(1+exp(t(z2)%*%beta2))
}
if(forma==2)
{
theta=exp(t(z1)%*%beta1)
p=exp(t(z2)%*%beta2+V[clinica])/(1+exp(t(z2)%*%beta2+V[clinica]))
}
}

```

```

}
sigma2u<-diag(Sigma)[1]
sigma2v<-diag(Sigma)[2]
rho<-Sigma[1,2]
l2<--(1/(2*(1-rho^2)))*sum(U^2/sigma2u+V^2/sigma2v-2*rho*U*V/
sqrt(sigma2u*sigma2v))-(J/2)*(log(sigma2u)+log(sigma2v)+log(1-rho^2))
if(modelo==0)
{
hpop<-log(f)+log(theta)
Spop<--theta*F
}
if(modelo==1)
{
hpop<-log(p)+log(f)+log(theta+1/(1-p*F))
Spop<--theta*p*F+log(1-p*F)
}
if(modelo==2)
{
hpop<-phi+log(f)+log(theta)+log(p)
Spop<--theta*p*exp(phi)*F
}
if(modelo==3)
{
hpop<-log(theta)+log(p)+log(f)-log(1+phi*theta*p*F)
Spop<--(1/phi)*log(1+phi*theta*p*F)
}
log.vero<-sum(delta*hpop+Spop)+l2
if(is.nan(log.vero)==TRUE)
{log.vero=-1e300}
log.vero
}

logvero.lambdab<-function(loglambdab)
{
lambdab<-exp(loglambdab)
if(modelo==0)
{
gamma<-zetab[1:q]
beta<-zetab[(q+1):(q+r)]
gamma<-matrix(gamma,ncol=1)
beta<-matrix(beta,ncol=1)
U<-zetab[(q+r+1):(q+r+J)]
V<-zetab[(q+r+J+1):(q+r+2*J)]
}
if(modelo==1 || modelo==2 || modelo==3)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r1)]
beta2<-zetab[(q+r1+1):(q+r1+r2)]
gamma<-matrix(gamma,ncol=1)
beta1<-matrix(beta1,ncol=1)
beta2<-matrix(beta2,ncol=1)
}
}

```

```

U<-zetab[(q+r1+r2+1):(q+r1+r2+J)]
V<-zetab[(q+r1+r2+J+1):(q+r1+r2+2*J)]
}
if(dist==1)
{
lambda<-lambdab[1]
nu<-lambdab[2]
S0<-exp(-lambda*t^nu)
h0<-lambda*nu*(t^(nu-1))
H0<-lambda*t^nu
}
if(dist==2)
{
S0<-1-ppexp(t,rate=lambdab,t=c(0,as.vector(a)))
h0<-dpexp(t,rate=lambdab,t=c(0,as.vector(a)))/S0
H0<--log(S0)
}
xi<-t(x)%*%gamma+U[clinica]
S<-exp(-H0*exp(xi))
F<-1-S
f<-h0*exp(xi)*S
f<-ifelse(f==0,1e-300,f)
if(modelo==0)
{
theta=exp(t(z)%*%beta+V[clinica])
p=rep(1,length(t))
}
if(modelo==1 || modelo==2 || modelo==3)
{
if(forma==1)
{
theta=exp(t(z1)%*%beta1+V[clinica])
p=exp(t(z2)%*%beta2)/(1+exp(t(z2)%*%beta2))
}
if(forma==2)
{
theta=exp(t(z1)%*%beta1)
p=exp(t(z2)%*%beta2+V[clinica])/(1+exp(t(z2)%*%beta2+V[clinica]))
}
}
if(modelo==0)
{
hpop<-log(f)
Spop<--theta*F
}
if(modelo==1)
{
hpop<-log(f)+log(theta+1/(1-p*F))
Spop<--theta*p*F+log(1-p*F)
}
if(modelo==2)
{

```

```

hpop<-log(f)
Spop<--theta*p*exp(phi)*F
}
if(modelo==3)
{
hpop<-log(f)-log(1+phi*theta*p*F)
Spop<--(1/phi)*log(1+phi*theta*p*F)
}
log.vero<-sum(delta*hpop+Spop)
if(is.nan(log.vero)==TRUE)
{log.vero=-1e300}
log.vero
}

gradiente.zetab<-function(zetab)
{
if(modelo==0)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r)]
gamma<-matrix(gamma,ncol=1)
beta<-matrix(beta1,ncol=1)
U<-zetab[(q+r+1):(q+r+J)]
V<-zetab[(q+r+J+1):(q+r+2*J)]
}
if(modelo==1 || modelo==2 || modelo==3)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r1)]
beta2<-zetab[(q+r1+1):(q+r1+r2)]
gamma<-matrix(gamma,ncol=1)
beta1<-matrix(beta1,ncol=1)
beta2<-matrix(beta2,ncol=1)
U<-zetab[(q+r1+r2+1):(q+r1+r2+J)]
V<-zetab[(q+r1+r2+J+1):(q+r1+r2+2*J)]
}
if(dist==1)
{
lambda<-lambdab[1]
nu<-lambdab[2]
S0<-exp(-lambda*t^nu)
h0<-lambda*nu*(t^(nu-1))
H0<-lambda*t^nu
}
if(dist==2)
{
S0<-1-ppexp(t,rate=lambdab,t=c(0,as.vector(a)))
h0<-dpexp(t,rate=lambdab,t=c(0,as.vector(a)))/S0
H0<--log(S0)
}
xi<-t(x)%*%gamma+U[clinica]
S<-exp(-H0*exp(xi))

```

```

F<-1-S
f<-h0*exp(xi)*S
if(modelo==0)
{
theta=exp(t(z)*%*%beta+V[clinica])
p=rep(1,length(t))
}
if(modelo==1 || modelo==2 || modelo==3)
{
if(forma==1)
{
theta=exp(t(z1)*%*%beta1+V[clinica])
p=exp(t(z2)*%*%beta2)/(1+exp(t(z2)*%*%beta2))
}
if(forma==2)
{
theta=exp(t(z1)*%*%beta1)
p=exp(t(z2)*%*%beta2+V[clinica])/(1+exp(t(z2)*%*%beta2+V[clinica]))
}
}
sigma2u<-diag(Sigma)[1]
sigma2v<-diag(Sigma)[2]
rho<-Sigma[1,2]
if(modelo==0)
{
deriv.xi=delta*(1-H0*exp(xi))-theta*H0*exp(xi)*S
deriv.eta=delta-theta*F
}
if(modelo==1)
{
c=(1-p*F)^(-1)
d=(theta+c)^(-1)
b=exp(xi)*H0
m=p*(1-p)
deriv.xi<-delta*(1-b+d*c^2*p*b*S)-p*b*S*(theta+c)
deriv.eta1<-delta*theta*d-theta*p*F
deriv.eta2<-delta*(1-p+d*c^2*m*F)-m*F*(theta+c)
}
if(modelo==2)
{
deriv.xi<-delta*(1-H0*exp(xi))-theta*p*H0*exp(xi)*S*exp(phi)
deriv.eta1<-delta-theta*p*F*exp(phi)
deriv.eta2<--theta*p*(1-p)*exp(phi)*F+delta*(1-p)
}
if(modelo==3)
{
deriv.xi<-delta*(1-H0*exp(xi))-(phi*delta+1)*theta*p*H0*exp(xi)*S*
(1+phi*theta*p*F)^(-1)
deriv.eta1<-delta-(phi*delta+1)*theta*p*F*(1+phi*theta*p*F)^(-1)
deriv.eta2<-delta*(1-p)-(phi*delta+1)*theta*p*(1-p)*F*(1+phi*theta*p*F)^(-1)
}
G=matrix(model.matrix(~-1+as.factor(clinica)),nrow=J,byrow=T)

```

```

if(modelo==1 || modelo==2 || modelo==3)
{
  if(forma==1)
  {
    mm1=cbind(x,0*x,0*x)
    mm2=cbind(0*z1,z1,0*z1)
    mm3=cbind(0*z2,0*z2,z2)
    mm4=cbind(G,0*G,0*G)
    mm5=cbind(0*G,G,0*G)
  }
  if(forma==2)
  {
    mm1=cbind(x,0*x,0*x)
    mm2=cbind(0*z1,z1,0*z1)
    mm3=cbind(0*z2,0*z2,z2)
    mm4=cbind(G,0*G,0*G)
    mm5=cbind(0*G,0*G,G)
  }
  mm=rbind(mm1,mm2,mm3,mm4,mm5)
  deriv=matrix(c(deriv.xi,deriv.etal,deriv.eta2),ncol=1)
  grad.partel=mm%*%deriv
}
if(modelo==0)
{
  mm1=rbind(x,0*z,G,0*G)
  mm2=rbind(0*x,z,0*G,G)
  mm=cbind(mm1,mm2)
  deriv=matrix(c(deriv.xi,deriv.eta),ncol=1)
  grad.partel=mm%*%deriv
}
U=matrix(U,ncol=1)
V=matrix(V,ncol=1)
if(modelo==1 || modelo==2 || modelo==3)
{
  partel=matrix(0,ncol=1,nrow=(q+r1+r2))
}
if(modelo==0)
{
  partel=matrix(0,ncol=1,nrow=(q+r))
}
parte2=U*sigma2v-V*rho*sqrt(sigma2u*sigma2v)
parte3=V*sigma2u-U*rho*sqrt(sigma2u*sigma2v)
det=sigma2u*sigma2v*(1-rho^2)
grad.parte2=-(1/det)*rbind(partel,parte2,parte3)
grad=grad.partel+grad.parte2
grad
}

matrizB<-function(zetab,lambdab,Sigma)
{
  if(modelo==0)
  {

```

```

gamma<-zetab[1:q]
beta<-zetab[(q+1):(q+r)]
gamma<-matrix(gamma,ncol=1)
beta<-matrix(beta,ncol=1)
U<-zetab[(q+r+1):(q+r+J)]
V<-zetab[(q+r+J+1):(q+r+2*J)]
}
if(modelo==1 || modelo==2 || modelo==3)
{
gamma<-zetab[1:q]
beta1<-zetab[(q+1):(q+r1)]
beta2<-zetab[(q+r1+1):(q+r1+r2)]
gamma<-matrix(gamma,ncol=1)
beta1<-matrix(beta1,ncol=1)
beta2<-matrix(beta2,ncol=1)
U<-zetab[(q+r1+r2+1):(q+r1+r2+J)]
V<-zetab[(q+r1+r2+J+1):(q+r1+r2+2*J)]
}
if(dist==1)
{
lambda<-lambdab[1]
nu<-lambdab[2]
S0<-exp(-lambda*t^nu)
h0<-lambda*nu*(t^(nu-1))
H0<-lambda*t^nu
}
if(dist==2)
{
S0<-1-ppexp(t,rate=lambdab,t=c(0,as.vector(a)))
h0<-dpexp(t,rate=lambdab,t=c(0,as.vector(a)))/S0
H0<--log(S0)
}
xi<-t(x)%*%gamma+U[clinica]
S<-exp(-H0*exp(xi))
F<-1-S
f<-h0*exp(xi)*S
if(modelo==0)
{
theta=exp(t(z)%*%beta+V[clinica])
p=rep(1,length(t))
}
if(modelo==1 || modelo==2 || modelo==3)
{
if(forma==1)
{
theta=exp(t(z1)%*%beta1+V[clinica])
p=exp(t(z2)%*%beta2)/(1+exp(t(z2)%*%beta2))
}
if(forma==2)
{
theta=exp(t(z1)%*%beta1)
p=exp(t(z2)%*%beta2+V[clinica])/(1+exp(t(z2)%*%beta2+V[clinica]))
}
}

```



```

}
}
sigma2u<-diag(Sigma)[1]
sigma2v<-diag(Sigma)[2]
rho<-Sigma[1,2]
if(modelo==0)
{
deriv.xi.xi<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.xi[i,i]<-H0[i]*exp(xi[i])*(delta[i]+theta[i]*(1-H0[i]*
exp(xi[i]))*S[i])
}
deriv.xi.eta<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.eta[i,i]<-theta[i]*exp(xi[i])*H0[i]*S[i]
}
deriv.eta.eta<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta.eta[i,i]<-theta[i]*F[i]
}
}
if(modelo==1)
{
c=(1-p*F)^(-1)
d=(theta+c)^(-1)
b=exp(xi)*H0
m=p*(1-p)
deriv.xi.xi<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.xi[i,i]<-delta[i]*b[i]*(1-p[i]*S[i]*d[i]*c[i]^2*(-d[i]*c[i]^2*
p[i]*b[i]*S[i]+2*c[i]*p[i]*b[i]*S[i]+1-b[i]))+p[i]*b[i]*S[i]*
(theta[i]*(1-b[i])+c[i]*(c[i]*p[i]*b[i]*S[i]+1-b[i]))
}
deriv.xi.etal<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.etal[i,i]<-theta[i]*p[i]*b[i]*S[i]*(delta[i]*d[i]^2*c[i]^2+1)
}
deriv.xi.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.eta2[i,i]<-b[i]*S[i]*m[i]*(theta[i]+c[i]*(c[i]*p[i]*F[i]+1))
-delta[i]*b[i]*S[i]*d[i]*c[i]^2*m[i]*(-d[i]*c[i]^2*p[i]*F[i]
+2*c[i]*p[i]*F[i]+1)
}
deriv.etal.etal<-diag(length(t))
for(i in 1:length(t))
{

```

```

deriv.eta1.eta1[i,i]<-theta[i]*p[i]*F[i]-delta[i]*theta[i]*d[i]
*(1-theta[i]*d[i])
}
deriv.eta1.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta1.eta2[i,i]<-theta[i]*m[i]*F[i]*(delta[i]*d[i]^2*c[i]^2+1)
}
deriv.eta2.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta2.eta2[i,i]<-m[i]*F[i]*((theta[i]+c[i])*(1-2*p[i])+c[i]^2*F[i])
+delta[i]*m[i]-delta[i]*F[i]*d[i]*c[i]^2*m[i]*(-d[i]*c[i]^2*m[i]*F[i]
+2*c[i]*m[i]*F[i]+1-2*p[i])
}
}
if(modelo==2)
{
deriv.xi.xi<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.xi[i,i]<-exp(xi[i])*H0[i]*(delta[i]+theta[i]*p[i]*exp(phi)*S[i]
*(1-exp(xi[i])*H0[i]))
}
deriv.xi.eta1<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.eta1[i,i]<-theta[i]*exp(phi)*p[i]*exp(xi[i])*H0[i]*S[i]
}
deriv.xi.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.eta2[i,i]<-theta[i]*exp(phi)*p[i]*(1-p[i])*exp(xi[i])*H0[i]*S[i]
}
deriv.eta1.eta1<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta1.eta1[i,i]<-theta[i]*exp(phi)*p[i]*F[i]
}
deriv.eta1.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta1.eta2[i,i]<-theta[i]*exp(phi)*p[i]*(1-p[i])*F[i]
}
deriv.eta2.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta2.eta2[i,i]<-theta[i]*exp(phi)*F[i]*p[i]*(1-p[i])*(1-2*p[i])
+delta[i]*p[i]*(1-p[i])
}
}
if(modelo==3)

```

```

{
c=(1+phi*theta*p*F)^(-1)
b=exp(xi)*H0
m=p*(1-p)
deriv.xi.xi<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.xi[i,i]<-delta[i]*b[i]+(1+phi*delta[i])*theta[i]*p[i]*S[i]
*b[i]*c[i]*(phi*theta[i]*p[i]*H0[i]*c[i]+1-b[i])
}
deriv.xi.etal<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.etal[i,i]<-(1+phi*delta[i])*theta[i]*p[i]*b[i]*S[i]*c[i]^2
}
deriv.xi.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.xi.eta2[i,i]<-(1+phi*delta[i])*theta[i]*m[i]*b[i]*S[i]*c[i]^2
}
deriv.etal.etal<-diag(length(t))
for(i in 1:length(t))
{
deriv.etal.etal[i,i]<-(1+phi*delta[i])*theta[i]*p[i]*F[i]*c[i]^2
}
deriv.etal.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.etal.eta2[i,i]<-(1+phi*delta[i])*theta[i]*m[i]*F[i]*c[i]^2
}
deriv.eta2.eta2<-diag(length(t))
for(i in 1:length(t))
{
deriv.eta2.eta2[i,i]<-(1+phi*delta[i])*theta[i]*F[i]*m[i]*(1-2*p[i]
-phi*theta[i]*p[i]^2*F[i])*c[i]^2+delta[i]*m[i]
}
}
if(modelo==0)
{
deriv.eta.xi=deriv.xi.eta
deriv=rbind(cbind(deriv.xi.xi,deriv.xi.eta),cbind(deriv.eta.xi,
deriv.eta.eta))
}
if(modelo==1 || modelo==2 || modelo==3)
{
deriv.etal.xi=deriv.xi.etal
deriv.eta2.xi=deriv.xi.eta2
deriv.eta2.etal=deriv.etal.eta2
deriv=rbind(cbind(deriv.xi.xi,deriv.xi.etal,deriv.xi.eta2),
cbind(deriv.etal.xi,deriv.etal.etal,deriv.etal.eta2),
cbind(deriv.eta2.xi,deriv.eta2.etal,deriv.eta2.eta2))
}
}

```

```

G=matrix(model.matrix(~-1+as.factor(clinica)),nrow=J,byrow=T)
if(modelo==1 || modelo==2 || modelo==3)
{
  if(forma==1)
  {
    mm1=cbind(x,0*x,0*x)
    mm2=cbind(0*z1,z1,0*z1)
    mm3=cbind(0*z2,0*z2,z2)
    mm4=cbind(G,0*G,0*G)
    mm5=cbind(0*G,G,0*G)
  }
  if(forma==2)
  {
    mm1=cbind(x,0*x,0*x)
    mm2=cbind(0*z1,z1,0*z1)
    mm3=cbind(0*z2,0*z2,z2)
    mm4=cbind(G,0*G,0*G)
    mm5=cbind(0*G,0*G,G)
  }
  mm=rbind(mm1,mm2,mm3,mm4,mm5)
}
if(modelo==0)
{
  mm1=cbind(x,0*x)
  mm2=cbind(0*z,z)
  mm3=cbind(G,0*G)
  mm4=cbind(0*G,G)
  mm=rbind(mm1,mm2,mm3,mm4)
}
B.partel=mm%%deriv%%t(mm)
IJ=diag(J)
if(modelo==1 || modelo==2 || modelo==3)
{
  partel=matrix(0,ncol=(q+r1+r2+2*J),nrow=(q+r1+r2))
  parte2=cbind(matrix(0,ncol=(q+r1+r2),nrow=J),sigma2v*IJ,
    -rho*sqrt(sigma2u*sigma2v)*IJ)
  parte3=cbind(matrix(0,ncol=(q+r1+r2),nrow=J),
    -rho*sqrt(sigma2u*sigma2v)*IJ,sigma2u*IJ)
}
if(modelo==0)
{
  partel=matrix(0,ncol=(q+r+2*J),nrow=(q+r))
  parte2=cbind(matrix(0,ncol=(q+r),nrow=J),sigma2v*IJ,
    -rho*sqrt(sigma2u*sigma2v)*IJ)
  parte3=cbind(matrix(0,ncol=(q+r),nrow=J),
    -rho*sqrt(sigma2u*sigma2v)*IJ,sigma2u*IJ)
}
det=sigma2u*sigma2v*(1-rho^2)
B.parte2=(1/det)*rbind(partel,parte2,parte3)
B=B.partel+B.parte2
colnames(B)<-c();rownames(B)<-c()
return(B)

```

```

}

Sigma2REML<-function(zetab, lambdab, Sigma)
{
  if(modelo==0)
  {
    U.BLUP=zetab[(q+r+1):(q+r+J)]
    V.BLUP=zetab[(q+r+J+1):(q+r+2*J)]
  }
  if(modelo==1 || modelo==2 || modelo==3)
  {
    U.BLUP=zetab[(q+r1+r2+1):(q+r1+r2+J)]
    V.BLUP=zetab[(q+r1+r2+J+1):(q+r1+r2+2*J)]
  }
  d.BLUP=c(U.BLUP,V.BLUP)
  B=matrizB(zetab, lambdab, Sigma)
  if(modelo==0)
  {
    A=solve(B)[(q+r+1):(q+r+2*J),(q+r+1):(q+r+2*J)]
  }
  if(modelo==1 || modelo==2 || modelo==3)
  {
    A=solve(B)[(q+r1+r2+1):(q+r1+r2+2*J),(q+r1+r2+1):(q+r1+r2+2*J)]
  }
  IJ<-diag(J); zero.J<-0*diag(J)
  Q1=rbind(cbind(IJ, zero.J), cbind(zero.J, zero.J))
  Q2=rbind(cbind(zero.J, IJ), cbind(IJ, zero.J))
  Q3=rbind(cbind(zero.J, zero.J), cbind(zero.J, IJ))
  R1.aux=(A+d.BLUP%*%t(d.BLUP))%*%Q1
  R2.aux=(A+d.BLUP%*%t(d.BLUP))%*%Q2
  R3.aux=(A+d.BLUP%*%t(d.BLUP))%*%Q3
  R1<-sum(diag(R1.aux))
  R2<-0
  if(dep==1)
  {
    R2<-sum(diag(R2.aux))/2
  }
  R3<-sum(diag(R3.aux))
  Sigma2REML<-matrix(c(R1/J, R2/sqrt(R1*R3), R2/sqrt(R1*R3), R3/J), nrow=2)
  return(Sigma2REML)
}

varianzaREML<-function(zetab, lambdab, Sigma)
{
  if(modelo==1 || modelo==2 || modelo==3)
  {
    r=r1+r2
  }
  U.BLUP=zetab[(q+r+1):(q+r+J)]
  V.BLUP=zetab[(q+r+J+1):(q+r+2*J)]
  d.BLUP=c(U.BLUP,V.BLUP)
  IJ<-diag(J); zero.J<-0*diag(J)

```

```

Q1=rbind(cbind(IJ, zero.J), cbind(zero.J, zero.J))
Q2=rbind(cbind(zero.J, IJ), cbind(IJ, zero.J))
Q3=rbind(cbind(zero.J, zero.J), cbind(zero.J, IJ))
B=matrizB(zetab, lambdab, Sigma)
A=solve(B) [(q+r+1):(q+r+2*J), (q+r+1):(q+r+2*J)]
sigma2u=diag(Sigma)[1]
sigma2v=diag(Sigma)[2]
rho=Sigma[1,2]
deriv.1=1/(2*sigma2u^2*sigma2v*(1-rho^2))*matrix(c(-2*sigma2v,
rho*sqrt(sigma2u*sigma2v), rho*sqrt(sigma2u*sigma2v), 0), nrow=2)
deriv.2=1/(2*sigma2u*sigma2v^2*(1-rho^2))*matrix(c(0,
rho*sqrt(sigma2u*sigma2v), rho*sqrt(sigma2u*sigma2v), -2*sigma2u), nrow=2)
deriv.3=1/(sigma2u*sigma2v*(1-rho^2)^2)*matrix(c(2*rho*sigma2v,
-(1+rho^2)*sqrt(sigma2u*sigma2v), -(1+rho^2)*sqrt(sigma2u*sigma2v),
2*rho*sigma2u), nrow=2)
deriv.sigma2u=kronecker(deriv.1, IJ)
deriv.sigma2v=kronecker(deriv.2, IJ)
deriv.rho=kronecker(deriv.3, IJ)
Sigma.aux=matrix(c(sigma2u, rho*sqrt(sigma2u*sigma2v),
rho*sqrt(sigma2u*sigma2v), sigma2v), nrow=2)
Omega=kronecker(Sigma.aux, IJ)
K1=A%%deriv.sigma2u
K2=Omega%%deriv.sigma2u
K3=A%%deriv.sigma2v
K4=Omega%%deriv.sigma2v
K5=A%%deriv.rho
K6=Omega%%deriv.rho
w.11=sum(diag((K1-K2)%%(K1-K2)))
w.22=sum(diag((K3-K4)%%(K3-K4)))
w.33=sum(diag((K5-K6)%%(K5-K6)))
w.12=sum(diag(K1%%K3+K2%%K4-2*K1%%K4))
w.13=sum(diag(K1%%K5+K2%%K6-2*K1%%K6))
w.23=sum(diag(K3%%K5+K4%%K6-2*K3%%K6))
K=matrix(c(w.11, w.12, w.13, w.12, w.22, w.23, w.13, w.23, w.33), nrow=3)
if(dep==1)
{
K=matrix(c(w.11, w.12, w.13, w.12, w.22, w.23, w.13, w.23, w.33), nrow=3)
}
return(2*solve(K))
}

```

Referências Bibliográficas

- Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The Annals of Statistics* **2**:1152-1174. [35](#), [129](#)
- Azevedo, R. (2011). *Modelos de fração de cura com fatores latentes competitivos e fragilidade*. Dissertação de Mestrado - IME - USP. São Paulo. [17](#)
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**:171-178. [79](#)
- Basu, S., Chib, S. (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models, *Journal of the American Statistical Association* **98**:224-235. [39](#), [122](#)
- Berkson, J., Gage, R. (1952). Survival curve for cancer patients following treatment, *Journal of the American Statistical Association* **47**:501-515. [2](#), [11](#), [16](#), [27](#), [46](#), [78](#)
- Blackwell, D., Macqueen, J. (1973) Ferguson distributions via Pólya urn schemes, *The Annals of Statistics* **1**:353-355. [32](#), [36](#)
- Borges, P., Rodrigues, J., Louzada, F., Balakrishnan, N. (2012) A cure rate survival model under a hybrid latent activation scheme, *Statistical Methods in Medical Research*, **doi:10.1177/0962280212469682**. [27](#)
- Breslow, N.E. (1974) Covariate analysis of censored survival data, *Biometrics* **30**:89-99. [45](#)
- Bush, C.A., MacEachern, S.N. (1996) A semiparametric Bayesian model for randomised block designs, *Biometrika* **83**:275-285. [37](#)
- Cancho, V.G., de Castro, M., Rodrigues, J. (2012a) A Bayesian analysis of the Conway-Maxwell-Poisson cure rate model, *Statistical Papers* **53**:165-176. [27](#)
- Cancho, V.G., Louzada, F., Barriga, G.D.C. (2012b) The Geometric Birnbaum-Saunders regression model with cure rate, *Journal of Statistical Planning and Inference* **142**:993-1000. [17](#)
- Cancho, V.G., Bandyopadhyay, D., Louzada, F., Yiqi, B. (2013a) The destructive negative binomial cure rate model with a latent activation scheme, *Statistical Methodology* **13**:48-68. [27](#), [129](#)
- Cancho, V.G., de Castro, M., Dey, D.K. (2013b) Long-term survival models with latent activation under a flexible family of distributions, *Brazilian Journal of Probability and Statistics* **27**(4):585-600. [27](#)
- Cancho, V.G., Louzada, F., Ortega, E.M. (2013c) The power series cure rate model: an application to a cutaneous melanoma data, *Communications in Statistics - Simulation and Computation* **42**:586-602. [27](#)
- Chen, M.H., Ibrahim, J.G., Sinha, D. (1999a). A new Bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**:909-919. [3](#), [4](#), [13](#), [15](#), [16](#), [17](#), [18](#), [27](#), [28](#), [46](#), [78](#), [108](#)
- Chen, M.H., Dey, D.K., Shao, Q. (1999b) A new skewed link model for dichotomous quantal response data, *Journal of the American Statistical Association*, **94**:1172-1186. [79](#)

- Colosimo, E. A., Giolo, S. R. (2006). *Análise de Sobrevida Aplicada*. Edgard Blücher, São Paulo. 106
- de Castro, M., Cancho, V.G., Rodrigues, J. (2009) A Bayesian long-term survival model parametrized in the cured fraction, *Biometrical Journal* **51**: 443-455. 17
- Escobar, M.D., West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**:577-588. 35
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* **39**:1-38. 44
- Donohue, M.C., Overholser, R., Xu, R., Vaida, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models, *Biometrika* **98**(3):685-700. 31
- Fahrmeir, L, Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. New York, Springer. 30
- Farewell, V.T. (1986). Mixture models in survival analysis: Are they worth the risk?, *Canadian Journal of Statistics* **14**(3):257-262. 28
- Feigl, P. e Zelen, M. (1965). Estimation of exponential probabilities with concomitant, *Biometrics* **21**:826-838. 8
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**:209-230. 31, 34, 85
- Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, **10**:101-113. 8
- Gallardo, D.I., Bolfarine, H., Pedroso-de-Lima, A.C. (2013a) Promotion time cure rate model with random effects: an application to a multi-centre clinical trial of carcinoma. *Statistics Research Letters* **2**:44-52. 4
- Gallardo, D.I., Bolfarine, H., Pedroso-de-Lima, A.C. (2013b) Promotion time cure rate model with bivariate random effects. *Communications in Statistics - Simulation and Computations*. Accepted for publication. 4
- Gamerman, D., Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman & Hall/CRC, Boca Raton, second edition. 90
- Gelman, A, Rubin, D.R. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science* **7**(4):457-511. 122
- Ghosh, J.K. & Ramamoorthi, R.V. (2003). *Bayesian nonparametric*. New York, USA: Springer. 34
- Hastings, W.K. (1970). Monte Carlo sampling methods using markov chains and their applications, *Biometrika* **57**:97-109. 38
- Jara, A., García-Zattera, M.J., Lesaffre, E. (2007) A Dirichlet process mixture model for the analysis of correlated binary responses, *Computational Statistics & Data Analysis* **51**:5402-5415. 92
- Jara, A. (2008). *Bayesian semiparametric methods for the analysis of complex data*. Tese de Doutorado - Catholic University of Leuven. Belgium. 32
- Kalbfleisch J.D., Prentice R.L. (2002) *The statistical analysis of failure data*. Wiley, New York, second edition. 2, 4, 41, 44, 105, 106, 107
- Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**:457-481.

- Klein, J.P., Houwelingen, H.C., Ibrahim, J.G., Scheike, T.H. (2013). *Handbook of survival analysis*. Chapman & Hall, New York, first edition. 28, 42
- Korwar, R.M. & Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *The Annals of Probability* 1:705-711. 34
- Kuk, Y.C. (1994). Asymptotically unbiased estimation in generalized linear models with random effects, *Journal of the Royal Society, Series B* 57:395-407. 29
- Lagakos, S.W. (1979). A stochastic model for censored-survival data in the presence of an auxiliary variable, *Biometrics* 35:139-156. 13
- Lai, X., Yau, K. (2008). Long-term survivor model with bivariate random effects: applications to bone marrow transplant and carcinoma study data, *Statistics in Medicine* 27:5692-5708. 3, 4, 44, 45, 50, 83
- Lee, Y., Nelder, Y.A. (2005). Likelihood for random-effect models, *Sort* 29(2):141-164. 30
- Li, C.s., Taylor, J., Sy, J. (2001). Identifiability of cure models, *Statistics & Probability Letters* 54:389-395. 3, 4, 26, 28, 49, 78, 83
- Li, Y., Müller, P., Lin, X. (2011). Center-adjusted inference for a nonparametric Bayesian random effect distribution, *Statistica Sinica* 21:1201-1223. 39
- Lopes, C.M., Bolfarine, H. (2012). Random effects in promotion time cure rate models, *Computational Statistics and Data Analysis* 56:75-87. 3, 47, 48, 129
- Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction, *Statistica Sinica* 20:661-674. 11
- MacEachern, S.N. (1994). Estimating normal means with a conjugate style Dirichlet process prior, *Communication in Statistics: Simulation and Computation*, 23:727-741. 37
- MacEachern, S.N., Muller, P. (1998). Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics*, 7:223-238. 38
- Maller, R.A., Zhou, S. (1996). *Survival analysis with long term survivors*. Wiley, New York, first edition. 11, 28, 105
- McGilchrist, C.A., Yau, K.K. (1995). The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models. *Communications in Statistics - Theory and Methods*, 24:2963-2980. 29, 30, 42, 44, 85, 86
- Miller, R.G. (1974). The jackknife: a review. *Biometrika*, 61:1-15 43, 79, 88
- Mizoi, M.F. (2004). *Influência local em modelos de sobrevivência com fração de cura*. Tese de Doutorado - IME - USP. São Paulo. 14
- Morais, A.L., Barreto-Souza, W. (2011) A compound class of Weibull and power series distributions, *Computational Statistics and Data Analysis*, 55:1410-1425. 15
- Nadarajah, S. (2009). The skew logistic distribution, *Advances in Statistical Analysis* 93(2):187-203. 79
- Neal, R.M. (2000). Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics* 9(2):249-265. 36, 37, 38
- Pan, R., Kozakai, Y. (2013). Semiparametric model and Bayesian analysis for clustered accelerate life testing data, *Statistical Research Letters* 2(1):1-11. 8

- Peng, Y. (2003). Estimating baseline distribution in proportional hazards cure models, *Computational Statistics & Data Analysis* **42**:187-201. [11](#)
- Peng, Y., Zhang, J. (2008). Identifiability of a mixture cure frailty model, *Statistics and Probability Letters* **78**:2604-2608. [28](#)
- Pinheiro, J.C., Chao, E.C., Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models, *Journal of Computational and Graphical Statistics*, **15**:58-81. [30](#)
- R Core Team (2012). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. ISBN 3-900051-07-0. [122](#), [140](#)
- Rodrigues, J., Cancho, V.G., De Castro, M. (2008) Teoria unificada de análise de sobrevivência. *Associação Brasileira de Estatística*, São Paulo-SP. [17](#)
- Rodrigues, J., Cancho, V.G., De Castro, M., Louzada-Neto, F. (2009a). On the unification of the long-term survival models, *Statistics and probability Letters* **79**:753-759. [16](#), [17](#), [18](#), [27](#)
- Rodrigues, J., De Castro, M., Cancho, V.G., Balakrishnan, N. (2009b). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data, *Journal of Statistical Planning and Inference* **139**:3605-3611. [27](#)
- Rodrigues, J., De Castro, M., Balakrishnan, N., Cancho, V.G. (2011). Destructive weighted Poisson cure rate models, *Lifetime Data Analysis* **17**:333-346. [3](#), [18](#), [20](#), [24](#), [26](#), [85](#), [129](#)
- Sahu, S., Dey, D., Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *The Canadian Journal of Statistics* **31**:129-150 [92](#)
- Sen, P.K., Singer, J.M., Pedroso-de-Lima, A.C. (2010) *From finite sample to asymptotic methods in statistics*. New York: Cambridge University Press. [109](#)
- Sethuraman, J. (1994). A constructive definition of Dirichlet prior. *Statistica Sinica* **2**:639-650. [34](#)
- Vaida, F., Blanchard, S. (2005) Conditional Akaike information for mixed-effects models, *Biometrika* **92**(2):351-370. [31](#)
- Vieira, C.C., Loschi, R.H., Duarte, D. (2013) Nonparametric mixtures based on skew-normal distributions: an application to density estimation, *Communications in Statistics - Theory and Methods*, doi:10.1080/03610926.2013.771745. [92](#)
- Weibull, W. (1939). A statistical theory of strength of materials, *Ingeniors Vetenskaps Akademien* **151**:45-55. [5](#)
- Weibull, W. (1951). A statistical distribution function of wide applicability, *Journal of Applied Mechanics* **3**:293-297. [5](#)
- West, M., Muller, P., Escobar, M.D. (1994). *Hierarchical priors and mixture models with application in regression and density estimation in Aspects of Uncertainty*, eds. P.R. Freeman and A.F.M. Smith, New York: Wiley. [37](#)
- Williams, J.S., Lagakos, S.W. (1977). Models for censored survival analysis: constant-sum and variable-sum models, *Biometrika* **64**:215-224. [13](#)
- Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population, *Biometrika* **20A**:32-52. [10](#)
- Yakovlev, A.Y., Tsodikov, A.D., (1996). Stochastic model of tumor latency and their biostatistical applications. *World Scientific*, Singapore. [3](#), [13](#)

- Yau K., Ng, A. (2001) Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma, *Statistics in Medicine* **20**:1591-1607. [41](#), [42](#), [43](#), [44](#), [47](#), [78](#), [105](#)
- Yin, G., Ibrahim, J.G. (2005). Cure rate models: a unified approach, *Canadian Journal of Statistics* **33**(4):559-570. [16](#), [17](#)
- Yu, B., Tiwari, R.C., Cronin, K.A., Feuer, E.J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data, *Statistics in Medicine* **23**:1733-1747. [28](#)
- Zhang, J., Peng, Y. (2009). Accelerated hazards mixture cure model, *Lifetime Data Analysis* **15**:455-467. [27](#)