

Modelos de Sobrevivência Bivariados Baseados na Cópula FGM: Uma Abordagem Bayesiana

Adriano Kamimura Suzuki

Orientador: Prof. Dr. Francisco Louzada Neto

Co-orientador: Prof. Dr. Vicente Garibay Cancho

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

São Carlos

Fevereiro de 2012

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

S968ms

Suzuki, Adriano Kamimura.

Modelos de sobrevivência bivariados baseados na cópula
FGM : uma abordagem bayesiana / Adriano Kamimura
Suzuki. -- São Carlos : UFSCar, 2012.
91 f.

Tese (Doutorado) -- Universidade Federal de São Carlos,
2012.

1. Análise de sobrevivência. 2. Inferência bayesiana. 3.
Cópula de Farlie-Gumbel-Morgenstern. 4. Medidas de
divergência. I. Título.

CDD: 519.9 (20^a)

Adriano Kamimura Suzuki

Modelos de Sobrevivência Bivariados Baseados na Cópula FGM: Uma Abordagem Bayesiana

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

Aprovada em 07 de fevereiro de 2012.

BANCA EXAMINADORA

Presidente



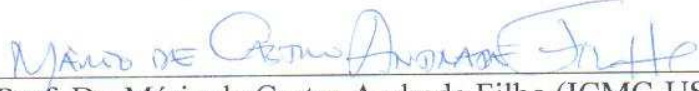
Prof. Dr. Francisco Louzada Neto (ICMC-USP / Orientador)

1º Examinador



Prof. Dr. Jorge Luis Bazán Guzmán (PUC-Perú)

2º Examinador



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

3º Examinador



Prof. Dr. Nikolai Valtchev Kolev (IME-USP)

4º Examinador



Prof. Dr. Vicente Garibay Cancho (ICMC-USP / Co-Orientador)

5º Examinador



Prof. Dr. Victor Hugo Lachos Dávila (UNICAMP)

Agradecimentos

A Deus pela força durante todo esses anos de estudos.

Aos meus pais, Luis e Ruth, pelo apoio e incentivo em todos os momentos, por serem pessoas honradas que com dedicação e trabalho honesto conseguiram proporcionar a mim a oportunidade de estudar. Sempre me estimularam a prosseguir meus estudos e me deram o conforto para cumpri-los.

À minha noiva e meu grande amor Meire pelo carinho, apoio, compreensão e seu imprescindível amor.

A toda minha família, em especial à minha irmã Daniela, ao meu cunhado Jessé e ao meu sobrinho Heitor, pelo apoio e ajuda em todos os momentos.

Ao meu orientador Prof. Dr. Francisco Louzada-Neto e ao meu coorientador Prof. Dr. Vicente Garibay Cancho pela orientação, ideias e principalmente pelo exemplo de dedicação e disciplina de trabalho. Agradeço ainda o apoio, a paciência e o incentivo à pesquisa durante todo o processo de elaboração deste trabalho. Minha eterna gratidão e admiração.

Aos professores Mário de Castro Andrade Filho, Victor Hugo Lachos Dávila, Jorge Luis Bazán Guzmán e Nikolai Valtchev Kolev, pelas sugestões e correções importantes à melhoria deste trabalho.

Aos professores do programa de pós-graduação do Departamento de Estatística da UFSCar, pois todos de forma direta ou indireta contribuíram para a realização desta tese.

À amiga e professora Juliana Cobre pela sua amizade e pelas valiosas sugestões e comentários apresentados durante o exame de qualificação.

Aos amigos Luís Ernesto Bueno Salazar e Erlandson Ferreira Saraiva, pelos bons momentos que passamos juntos, discutindo sobre estatística ou conversando sobre diversos assuntos.

Aos funcionários do DEs-UFSCar, em especial a Maria Isabel Rinaldo Pessoa de Araujo, pelo carinho e atenção.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Resumo

Neste trabalho apresentamos uma análise bayesiana para dados de sobrevivência bivariados na presença de covariáveis e observações censuradas. Propomos uma distribuição bivariada para os tempos de sobrevivência baseada na cópula de Farlie-Gumbel-Morgenstern (FGM) para modelar dados com fraca dependência. Alguns modelos de sobrevivência com e sem fração de cura foram assumidos para as distribuições marginais. Para fins inferenciais foi considerada uma abordagem bayesiana usando métodos Monte Carlo em Cadeias de Markov (MCMC). Além disso, algumas discussões sobre os critérios de seleção de modelos são apresentadas e comparações com outras cópulas foram realizadas. A fim de detectar observações influentes nos dados analisados foi utilizado o método bayesiano de análise de influência caso a caso baseado na divergência ψ . Os sistemas OpenBUGS e R foram utilizados para simular amostras da distribuição *a posteriori* de interesse. Ilustrações numéricas são apresentadas considerando conjunto de dados artificiais e reais.

Palavras chaves: Análise de Sobrevivência, Inferência Bayesiana, Cópula de Farlie-Gumbel-Morgenstern, Divergência ψ .

Abstract

In this work we present a Bayesian analysis for bivariate survival data in the presence of a covariate and censored observations. We propose a bivariate distribution for the bivariate survival times based on the Farlie-Gumbel-Morgenstern (FGM) copula to model data with weak dependence. Some survival models with and without cure rate have been assumed for the marginal distributions. For inferential purpose a Bayesian approach via Markov Chain Monte Carlo (MCMC) was considered. Further, some discussions on model selection criteria are given and comparisons with other copula models were performed. To detect influential observations in the data we consider a Bayesian case deletion influence diagnostics based on the ψ -divergence. The OpenBUGS and R systems were used to simulate samples of the posterior distribution. Numerical illustrations are presented considering artificial and real data sets.

Key words: Survival Analysis, Bayesian Inference, Farlie-Gumbel-Morgenstern Copula, ψ -Divergence.

Sumário

1	Introdução	1
1.1	Funções cópulas	3
1.2	Critérios de comparação de modelos	8
1.3	Diagnóstico	9
1.3.1	Análise de influência caso a caso	10
1.4	Dados reais utilizados	13
1.4.1	Dados de retinopatia diabética	13
1.4.2	Dados de HIV	14
1.5	Organização dos capítulos	16
2	Modelo de sobrevivência bivariado baseado na cópula FGM	18
2.1	Modelo de sobrevivência FGM bivariado	18
2.2	Inferência	19
2.2.1	Distribuições a <i>priori</i> e a <i>posteriori</i>	20
2.2.2	Implementação Computacional	21
2.3	Aplicações	23
2.3.1	Propriedades frequentistas	23
2.3.2	Diagnóstico de observações influentes	25
2.3.3	Dados reais	30
3	Modelo de sobrevivência bivariado mistura padrão baseado na cópula FGM	35
3.1	Função de sobrevivência de longa duração	36

3.2	Modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM	38
3.2.1	Inferência	39
3.2.2	Implementação Computacional - Covariáveis utilizadas no parâmetro de escala da distribuição Weibull	40
3.2.3	Implementação Computacional - Covariáveis utilizadas no parâmetro de fração de cura	48
4	Modelo de sobrevivência bivariado tempo de promoção baseado na cópula FGM	61
4.1	Modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM	62
4.2	Inferência	62
4.2.1	Implementação Computacional	63
4.3	Aplicações	64
4.3.1	Dados artificiais	64
4.3.2	Diagnóstico de observações influentes	65
4.3.3	Dados reais	68
5	Considerações finais e extensões	75
	Referências Bibliográficas	81

Lista de Figuras

1.1	Estimativas de Kaplan-Meier da função de sobrevivência e gráficos TTT para os dados de retinopatia diabética.	14
1.2	Estimativas de Kaplan-Meier da função de sobrevivência e gráficos TTT para os dados de HIV.	16
2.1	Dados simulados. Gráficos de índices das medidas de divergência para o caso (i).	30
2.2	Dados de retinopatia diabética. Gráfico de dispersão de T_1 e T_2 (painel acima). Curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull.	32
2.3	Dados de retinopatia diabética. Gráficos de índices das medidas de divergência.	33
2.4	Dados de retinopatia diabética. Gráficos de contorno para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência FGM com marginais Weibull.	34
3.1	Gráficos de índices das medidas de divergência para o caso (e).	46
3.2	Gráficos de índices das medidas de divergência.	48

3.3	Dados de retinopatia diabética. Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Exponencial baseado na cópula FGM.	49
3.4	Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência mistura padrão Exponencial baseado na cópula FGM.	50
3.5	Gráficos de índices das medidas de divergência.	57
3.6	Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.	58
3.7	Gráficos de índices das medidas de divergência.	59
3.8	Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.	60
4.1	Gráficos de índices das medidas de divergência para o caso (i).	69
4.2	Dados de retinopatia diabética. Gráficos de índices das medidas de divergência.	71
4.3	Dados de retinopatia diabética perturbado. Gráficos de índices das medidas de divergência.	72
4.4	Dados de retinopatia diabética. Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM.	73
4.5	Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM.	74

Lista de Tabelas

2.1	Dados simulados. Média MC e EQM MC para os parâmetros do modelo de sobrevivência FGM bivariado baseados sobre as 100 amostras geradas para as duas configurações de censuras.	25
2.2	Dados simulados. Média Monte Carlo dos critérios bayesianos baseados sobre as 100 amostras geradas para as duas configurações de censuras.	25
2.3	Dados simulados. Média e desvio padrão (DP) para os parâmetros do modelo de sobrevivência FGM bivariado para cada conjunto de dados simulados.	27
2.4	Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência FGM bivariado para cada conjunto de dados simulados.	28
2.5	Dados simulados. Medidas de divergência.	29
2.6	Dados de retinopatia diabética. Média <i>a posteriori</i> , desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência FGM bivariado.	31
2.7	Dados de retinopatia diabética. Medidas de divergência.	33
2.8	Dados de retinopatia diabética. Critérios bayesianos.	34
3.1	Dados simulados. Média <i>a posteriori</i> (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.	42
3.2	Dados simulados. Critérios bayesianos.	43

3.3	Dados simulados. Média <i>a posteriori</i> e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.	44
3.4	Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.	44
3.5	Dados simulados. Medidas de Divergência.	45
3.6	Dados de retinopatia diabética. Média <i>a posteriori</i> e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão baseado na cópula FGM.	47
3.7	Dados de retinopatia diabética. Medidas de diagnósticos para o caso número 100.	47
3.8	Dados de retinopatia diabética. Critérios bayesianos.	50
3.9	Dados simulados. Média <i>a posteriori</i> (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.	52
3.10	Dados simulados. Critérios bayesianos.	52
3.11	Dados simulados. Média <i>a posteriori</i> e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.	54
3.12	Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.	55
3.13	Dados Simulados. Medidas de diagnósticos.	56
3.14	Dados de HIV. Média <i>a posteriori</i> e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull e Exponencial baseado na cópula FGM.	57
3.15	Dados de HIV. Medidas de divergência.	58

3.16	Dados de HIV. Critérios bayesianos.	60
4.1	Dados simulados. Média <i>a posteriori</i> (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM.	65
4.2	Dados simulados. Critérios bayesianos.	65
4.3	Dados simulados. Média <i>a posteriori</i> e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM para cada conjunto de dados simulados.	66
4.4	Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM para cada conjunto de dados simulados.	67
4.5	Dados simulados. Medidas de divergência.	68
4.6	Dados de retinopatia diabética. Média <i>a posteriori</i> , desvio padrão (DP) e HPD (95%) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção baseado na cópula FGM.	70
4.7	Dados de retinopatia diabética. Medidas de divergência.	70
4.8	Dados de retinopatia diabética. Critérios bayesianos.	74

Capítulo 1

Introdução

Na análise de sobrevivência e de confiabilidade ocorrem situações em que observamos dois tempos de vida para um mesmo paciente ou equipamento, ou seja, dados de sobrevivência bivariados. Exemplos de situações na área médica ocorrem quando o interesse é estudar os tempos de vida de órgãos humanos emparelhados como rins e olhos, o tempo de reincidência do câncer de dois diferentes órgãos, o tempo até a primeira e a segunda infecção ou internação, etc. Aplicações industriais ocorrem por exemplo em sistemas cujo tempo de duração depende da durabilidade de dois componentes, como o tempo de vida de motores em um avião bimotor.

Um conjunto de dados de sobrevivência multivariados geralmente apresenta associação entre os tempos de sobrevivência. Para o estudo desta dependência, a abordagem mais popular são os modelos de fragilidade, propostos por Vaupel *et al.* (1979). Nos modelos de fragilidade um ou mais efeitos aleatórios são incluídos para modelar a dependência entre as observações. Neste caso, os tempos marginais são condicionalmente independentes dada a variável de fragilidade.

Os primeiros modelos de fragilidade para dados multivariados foram considerados por Clayton (1978) e Oakes (1982) e a primeira abordagem bayesiana foi considerado por Clayton (1991). Hougaard (1995) apresenta uma revisão dos modelos de fragilidades multiplicativos numa perspectiva frequentista. Sinha & Dey (1997) apresentam uma revisão sob uma abordagem bayesiana. Sahu & Dey (2000) consideram um estudo comparativo entre modelos de fragilidade e modelos bivaria-

dos baseados nas distribuições Exponencial e Weibull para dados de sobrevivência bivariados sob uma perspectiva bayesiana. A literatura sobre estes modelos é extensa, dentre as quais citamos os trabalhos de Hougaard (1986a, b), Self & Prentice (1986), Longini & Halloran (1996), Aalen & Hjort (2002), Wienke *et al.* (2003), Tomazella *et al.* (2006), Balakrishnan & Peng (2006) e Duchateau & Janssen (2008). Também, podemos citar os livros de Therneau & Grambsch (2000), Hanagal (2011) e Wienke (2011).

Recentemente, modelos de cópulas tem-se tornado uma ferramenta popular para modelar a dependência entre dados multivariados, especialmente em áreas biológicas, ciências atuariais e finanças (ver por exemplo Shih & Louis, 1995; Embrechts *et al.*, 2003; Cherubini *et al.*, 2004; Trivedi & Zimmer, 2005; Nelsen, 2006; Kolev *et al.*, 2006; Salvadori *et al.*, 2007; Jaworski, 2010). Oakes (2001) e Goethals *et al.* (2008) exibem algumas relações entre modelos com fragilidade e modelos de cópulas.

Uma cópula é uma função que conecta as distribuições marginais univariadas com sua distribuição multivariada conjunta. Diferentes funções cópulas representam diferentes estruturas de dependência entre as variáveis (Nelsen, 2006). Um modelo de cópula é uma ferramenta conveniente para estudar a estrutura de dependência e é flexível em aplicações, uma vez que quando a dispersão dos dados não se ajusta a nenhuma família conhecida de distribuições conjuntas, pode ser difícil especificar a distribuição conjunta. Usando cópulas, no entanto, podemos estimar primeiro distribuições marginais e, em seguida, estimar os parâmetros da função cópula. Uma outra vantagem da modelagem cópula é a sua relativa simplicidade matemática. Além disso, é possível construir uma variedade de estruturas de dependência com base em modelos paramétricos ou não paramétricos para as distribuições marginais.

Em análise de sobrevivência, modelos baseados em cópulas são considerados, por exemplo em Hougaard (1989), Oakes (1989), Shih & Louis (1995), Gustafson *et al.* (2003) e Zhang *et al.* (2010). Particularmente, sob uma perspectiva bayesiana, Romeo *et al.* (2006) consideraram uma aplicação de cópulas arquimedianas para modelar a dependência de dados de sobrevivência bivariados com distribuições marginais Exponencial e Weibull.

Neste trabalho, consideramos o modelo cópula baseada na distribuição de Farlie-Gumbel-Morgenstern (FGM) (Conway, 1983; Hutchinson & Lai, 1990) para modelar a dependência de dados de sobrevivência bivariados.

O principal objetivo deste trabalho é apresentar o modelo de sobrevivência FGM bivariado. Assumindo uma abordagem bayesiana utilizando métodos Monte Carlo em Cadeias de Markov (MCMC), apresentamos o procedimento inferencial para dados de sobrevivência bivariados na presença de covariáveis e observações censuradas. Para as distribuições marginais assumimos modelos de sobrevivência sem e com fração de cura. Além disso, realizamos comparações com outras cópulas e com o caso assumindo independência entre os tempos marginais.

Depois de ajustar o modelo é importante verificar as suposições do modelo e realizar estudos de sensibilidade para detectar possíveis observações influentes que podem causar distorções nos resultados da análise. Métodos de diagnóstico têm sido uma importante ferramenta na análise de regressão. Seguindo o trabalho pioneiro de Cook (1986), diagnósticos de influência local e eliminação de casos têm sido amplamente aplicados a muitos modelos de regressão. Neste trabalho discutimos diagnóstico de influência sob uma abordagem bayesiana. O objetivo é desenvolver medidas de diagnósticos baseadas na divergência ψ (Peng & Dey, 1995; Weiss, 1996) entre as distribuições *a posteriori* dos parâmetros do modelo proposto. Mostramos a aplicabilidade dos modelos propostos a conjuntos de dados simulados e reais.

A seguir, na Subseção 1.1 apresentamos uma breve introdução sobre as funções cópulas. Nas Subseções 1.2 e 1.3 apresentamos, respectivamente, os critérios de comparação de modelos bayesianos e o procedimento de análise de influência utilizado. Na Subseção 1.4 apresentamos os dois conjuntos de dados utilizados neste trabalho. Por fim, na Subseção 1.5 apresentamos a organização dos capítulos subsequentes.

1.1 Funções cópulas

Nesta seção apresentamos uma breve introdução de alguns conceitos básicos das funções cópulas. Informações complementares podem ser encontradas, por exemplo,

em Anjos *et al.* (2004), Trivedi & Zimmer (2005), Nelsen (2006) e Jaworski (2010).

Definição 1. Uma cópula é uma distribuição multivariada cujas marginais são $U(0, 1)$. Considere o vetor aleatório $\mathbf{U} = (U_1, \dots, U_n) \in \mathbf{I}^n$ com cópula n -dimensional C , temos

$$C(u_1, \dots, u_n; \phi) = Pr(U_1 \leq u_1, \dots, U_n \leq u_n; \phi), \quad (u_1, \dots, u_n) \in \mathbf{I}^n, \quad (1.1)$$

em que ϕ é o parâmetro associado à função cópula.

Teorema 1. (Teorema de Sklar.) Seja H uma função de distribuição conjunta com marginais $F_1(t_1), \dots, F_n(t_n)$. Então, existe uma cópula n -dimensional C tal que

$$H(t_1, \dots, t_n; \phi) = C(F_1(t_1), \dots, F_n(t_n); \phi). \quad (1.2)$$

Se $F_1(t_1), \dots, F_n(t_n)$ são todas contínuas, então C é única.

Demonstração: Ver Nelsen (2006).

O teorema de Sklar é um dos resultados mais importantes na teoria e aplicações de cópulas. A partir deste temos que a cópula conecta as distribuições marginais univariadas formando uma distribuição multivariada, ou então que uma função de distribuição multivariada pode ser decomposta nas marginais univariadas e na estrutura de dependência dada pela cópula.

Para todas as cópulas temos os limites de Fréchet-Hoeffding, que são os limites de variação, e são dados por

$$\max(0, \sum_{i=1}^n F_i(t_i) - n + 1) \leq C(F_1(t_1), \dots, F_n(t_n)) \leq \min(F_1(t_1), \dots, F_n(t_n)). \quad (1.3)$$

A classe das cópulas arquimedianas (Genest & Mackay, 1986), amplamente utilizada na prática, permite uma grande variedade de estruturas de dependência. Todas as cópulas Arquimedianas usualmente encontradas possuem expressões com forma fechada. Tem sido utilizada com sucesso em conexão com a noção de fragilidade (Oakes, 1989). A representação da cópula arquimediana permite reduzir o estudo de cópula multivariada ao estudo de uma função univariada φ , comumente chamada de gerador de uma cópula arquimediana.

Definição 2. Considere $\varphi : [0, 1] \mapsto [0, \infty]$ tal que φ é uma função contínua e estritamente decrescente, com $\varphi(1) = 0$. Também considere a pseudo-inversa de φ sendo a função $\varphi^{[-1]} : [0, \infty] \mapsto [0, 1]$ dado por

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \quad (1.4)$$

Note que $\varphi^{[-1]}$ é contínua e não crescente em $[0, \infty]$, e estritamente decrescente em $[0, \varphi(0)]$. Também $\varphi^{[-1]}(\varphi(t)) = t$ em $[0, 1]$, e $\varphi(\varphi^{[-1]}(t)) = \min(t, \varphi(0))$. Claramente se $\varphi(0) = \infty$, então $\varphi^{[-1]} = \varphi^{-1}$.

Teorema 2. Considere $C : [0, 1] \times [0, 1] \mapsto [0, 1]$ dada por

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \quad (1.5)$$

em que φ e $\varphi^{[-1]}$ são como na Definição 2. Então, C é uma cópula se, e somente se, φ é uma função convexa.

Demonstração: Ver Nelsen (2006).

Note que φ é convexa, se e somente se, $\varphi^{[-1]}$ é convexa.

Definição 3. Cópulas arquimedianas. Cópulas da forma apresentada no Teorema 2 são denominadas cópulas arquimedianas com gerador φ . Se $\varphi(0) = \infty$, dizemos que φ é um gerador estrito. Se $\varphi(0) < \infty$, φ é denominado gerador não estrito.

É denominada cópula arquimediana estrita se $\varphi^{[-1]} = \varphi^{-1}$ e $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$.

Considere $C(u, v)$ uma cópula arquimediana bivariada. Temos as seguintes propriedades:

- i) $C(u, v) = C(v, u)$, ou seja, C é simétrica (permutável);
- ii) $C(C(u, v), w) = C(u, C(v, w))$ para todo u, v, w em $[0, 1]$, ou seja, C é associativa;
- iii) Seja φ a geradora de C . Então para qualquer constante $a > 0$ tem-se que $a\varphi$ também é uma geradora de C .

A seguir apresentamos as cópulas arquimedianas bivariadas de Clayton, Frank e Estável Positiva (PS) e, a cópula de Farlie-Gumbel-Morgenstern (FGM), que utilizamos neste trabalho.

Cópula de Clayton

A cópula de Clayton (Clayton, 1978), originalmente estuda por Kimeldorf & Sampson (1975), tem a forma

$$C(u, v; \phi) = (u^{-\phi} + v^{-\phi} - 1)^{-\frac{1}{\phi}}, \quad \phi \in \mathbb{R}^+, \quad (1.6)$$

e função geradora dada por $\varphi(t) = \frac{1}{\phi}(t^{-\phi} - 1)$.

O valor $\phi = 0$ representa independência, ou seja, $C_\phi(u, v) = uv$. Quando $\phi \rightarrow +\infty$ a cópula atinge o limite superior de Fréchet, mas para nenhum valor atinge o limite inferior. A cópula de Clayton não contempla dependência negativa.

A relação entre a medida τ de Kendall e o parâmetro de dependência ϕ da cópula de Clayton é dada por $\tau = \frac{\phi}{\phi+2}$.

Cópula de Frank

A cópula de Frank, considerada por Frank (1979) e estudadas por Nelsen (1986) e Genest (1987), é amplamente utilizada em aplicações empíricas (Meester & MacKay, 1994). Permite dependência negativa e é simétrica em ambas as caldas. Tem a forma

$$C(u, v; \phi) = \log_\phi \left(1 + \frac{(\phi^u - 1)(\phi^v - 1)}{\phi - 1} \right), \quad \phi \in (0, 1) \quad (1.7)$$

e função geradora dada por $\varphi(t) = -\ln \left(\frac{1-\phi^t}{1-\phi} \right)$.

O valor $\phi = 1$ representa independência, ou seja, $C_\phi(u, v) = uv$. Os coeficientes de dependência da cauda inferior e superior para os membros desta família são iguais a 0 (Salvadori *et al.*, 2007).

A relação entre a medida τ de Kendall e o parâmetro de dependência ϕ da cópula de Frank não tem forma fechada e é dada por

$$\tau = 1 + \frac{4}{\ln(\phi)} \left(\frac{1}{\ln(\phi)} \int_0^{-\ln(\phi)} \frac{t}{e^t - 1} dt + 1 \right).$$

Cópula Estável Positiva

A cópula Estável Positiva (PS), também conhecida como família de Gumbel-Hougaard, foi considerada por Hougaard (1986a, b) e Hutchinson & Lai (1990). Tem a forma

$$C(u, v; \phi) = \exp\{-[(-\ln(u))^{1/\phi} + (-\ln(v))^{1/\phi}]^\phi\}, \quad \phi \in (0, 1), \quad (1.8)$$

e função geradora dada por $\varphi(t) = (-\ln(t))^{1/\phi}$.

Os valores 0 e 1 correspondem ao limite superior de Fréchet e independência, respectivamente, mas esta cópula não atinge o limite inferior de Fréchet para nenhum valor de ϕ . Como a cópula de Clayton, a cópula PS não permite dependência negativa, mas em contraste à de Clayton, apresenta forte dependência na cauda direita e relativamente fraca dependência na cauda esquerda.

A relação entre a medida τ de Kendall e o parâmetro de dependência ϕ da cópula PS é dada por $\tau = 1 - \phi$.

Cópula de Farlie-Gumbel-Morgenstern

A cópula de Farlie-Gumbel-Morgenstern (FGM) (Gumbel, 1958; Farlie, 1960) é dada por

$$C(u, v; \phi) = uv[1 + \phi(1 - u)(1 - v)], \quad -1 \leq \phi \leq 1. \quad (1.9)$$

O valor $\phi = 0$ representa independência, ou seja, $C_\phi(u, v) = uv$. A cópula FGM é atraente devido à sua forma analítica simples e tem sido muito utilizada em modelagem para testes de associação e no estudo da eficiência de procedimentos não paramétricos (Nelsen, 2006). Para uma lista de exemplo de aplicações e referências, vide Conway (1983) e Hutchinson & Lai (1990).

A relação entre a medida τ de Kendall e o parâmetro de dependência ϕ da cópula FGM é dada por $\tau = \frac{2\phi}{9}$. Note que $-\frac{2}{9} \leq \tau \leq \frac{2}{9}$.

1.2 Critérios de comparação de modelos

Na literatura encontramos diversas metodologias que se propõem a analisar a adequabilidade de um modelo a um certo conjunto de dados, além de, dentre uma coleção de modelos, selecionar o melhor.

Neste trabalho utilizamos quatro critérios de seleção de modelos, os quais especificamente são usados na metodologia bayesiana em que as amostras das distribuições *a posteriori* para os parâmetros do modelo são obtidas usando métodos MCMC. A seguir descrevemos cada um deles.

O critério DIC (*Deviance Information Criterion*) proposto por Spiegelhalter *et al.* (2002), o EAIC (*Expected Akaike Information Criterion*) por Brooks (2002) e o EBIC (*Expected Bayesian (ou Schwarz) Information Criterion*) por Carlin & Louis (2001) são critérios baseados na média *a posteriori* da deviança, $E\{D(\theta)\}$, que é uma medida de ajuste e que pode ser aproximada por

$$\bar{D} = \frac{1}{V} \sum_{v=1}^V D(\theta_v),$$

sendo v o índice que indica a v -ésima realização de um total de V realizações (após o *burn-in*) e

$$D(\theta) = -2 \sum_{i=1}^n \ln(g(t_{1i}, t_{2i}|\theta)),$$

em que $g(\cdot)$ é a função densidade de probabilidade correspondente ao nosso modelo. Os critérios EAIC, EBIC e DIC podem ser calculados, respectivamente, por $\widehat{EAIC} = \bar{D} + 2q$, $\widehat{EBIC} = \bar{D} + q \ln(n)$ e $\widehat{DIC} = \bar{D} + \widehat{\rho}_D = 2\bar{D} - \widehat{D}$, em que q é o número de parâmetros no modelo, ρ_D é o número efetivo de parâmetros definido como $E\{D(\theta)\} - D\{E(\theta)\}$. $D\{E(\theta)\}$ pode ser estimada por

$$\widehat{D} = D\left(\frac{1}{V} \sum_{v=1}^V \theta_v\right).$$

Comparando modelos alternativos, o modelo preferido é aquele com os menores valores destes critérios.

Um outro critério, um dos mais utilizados em trabalhos aplicados, é derivado das ordenadas da densidade preditiva condicional (CPO), que é uma ferramenta de avaliação do modelo muito útil e intensamente usada na literatura estatística sob vários contextos (Ibrahim *et al.*, 2001). Seja \mathcal{D} os dados completos e $\mathcal{D}^{(-i)}$ os dados com a i -ésima observação excluída. Denotamos a densidade *a posteriori* de θ dado $\mathcal{D}^{(-i)}$ por $\pi(\theta|\mathcal{D}^{(-i)})$, para $i = 1, \dots, n$, e θ é um vetor de parâmetros.

Para a i -ésima observação, a CPO_i pode ser escrita como

$$CPO_i = \int_{\Theta} g(t_{1i}, t_{2i}|\theta)\pi(\theta|\mathcal{D}^{(-i)})d\theta = \left\{ \int_{\Theta} \frac{\pi(\theta|\mathcal{D})}{g(t_{1i}, t_{2i}|\theta)}d\theta \right\}^{-1}, i = 1, \dots, n. \quad (1.10)$$

e $g(t_{1i}, t_{2i}|\theta)$ é a função densidade de probabilidade. Na comparação de vários modelos, para valores altos de CPO_i temos um melhor ajuste do modelo.

Para o nosso modelo proposto uma forma fechada de CPO_i não é possível. Entretanto, uma estimativa Monte Carlo de CPO_i pode ser obtida através de uma simples amostra MCMC a partir da distribuição *a posteriori* $\pi(\theta|\mathcal{D})$. Considere $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(V)}$ uma amostra de tamanho V de $\pi(\theta|\mathcal{D})$ após o *burn-in*. Uma aproximação Monte Carlo de CPO_i (Chen *et al.*, 2000) é dada por

$$\widehat{CPO}_i = \left\{ \frac{1}{q} \sum_{q=1}^V \frac{1}{g(t_{1i}, t_{2i}|\theta^{(q)})} \right\}^{-1}. \quad (1.11)$$

Como em Ibrahim *et al.* (2001), utilizamos a estatística $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$ na seleção de modelos, em que maiores valores de LPML indicam o melhor modelo.

1.3 Diagnóstico

Uma maneira de detectar a existência de pontos influentes pode ser feita por meio da análise de diagnóstico que teve início com a análise de resíduos (Cox & Snell, 1968).

Na literatura, uma forma utilizada de avaliação da influência de uma observação no ajuste de um modelo é por meio da exclusão de casos (Cook & Weisberg, 1982).

Sob pequenas perturbações no modelo e/ou nos dados, Cook (1986) propôs a avaliação da influência conjunta das observações, ao invés da retirada individual ou

conjunta de pontos. Caso essas perturbações causem efeitos desproporcionais, pode haver evidências de que há mau ajuste do modelo ou sérios afastamentos de suas suposições.

Peng & Dey (1995) apresentam duas abordagens bayesianas distintas para detectar observações influentes no ajuste de modelos de regressão. A primeira é baseada na distribuição *a posteriori* e a segunda na distribuição preditiva, em que a influência de uma observação é estudada observando as mudanças no fator de Bayes, pseudo-Bayes e Bayes *a posteriori*. Quatro medidas específicas, dentre elas a divergência de Kullback-Leibler (K-L), são propostas, as quais transmitem os efeitos de uma única observação ou covariável na distribuição *a posteriori*.

Atualmente, técnicas de influência local têm sido amplamente utilizados (Ortega *et al.*, 2003; Labra *et al.*, 2005; Ortega *et al.*, 2006; Vidal & Castro, 2010).

Cho *et al.* (2009) propõem um método bayesiano de análise de influência caso a caso para modelos de sobrevivência. Desenvolveram medidas de diagnóstico para avaliar a influência de um caso nas distribuições *a posteriori* conjuntas e marginais baseadas na divergência de Kullback-Leibler (K-L). Neste trabalho, Cho *et al.* (2009) apresentam uma expressão simplificada para o cálculo da divergência de Kullback-Leibler entre a distribuição *a posteriori* com os dados completos e a distribuição *a posteriori* sob a exclusão de um simples caso, como também investigar estas relações para as ordenadas da densidade preditiva condicional. Aplicações desta metodologia podem ser encontradas em Cancho *et al.* (2010) e Cancho *et al.* (2011b).

1.3.1 Análise de influência caso a caso

Neste trabalho vamos considerar a análise de influência caso a caso baseado na divergência ψ .

Seja $D_\psi(P, P_{(-i)})$ a divergência ψ entre P e $P_{(-i)}$, em que P indica a distribuição *a posteriori* de θ para os dados completos e, $P_{(-i)}$ a distribuição *a posteriori* sem o i -ésimo caso. Especificamente,

$$D_\psi(P, P_{(-i)}) = \int \psi \left(\frac{\pi(\theta|\mathcal{D}^{(-i)})}{\pi(\theta|\mathcal{D})} \right) \pi(\theta|\mathcal{D}) d\theta, \quad (1.12)$$

em que ψ é uma função convexa com $\psi(1) = 0$. Várias escolhas de ψ são dadas em Dey & Birmiwal (1994). Por exemplo, $\psi(z) = -\log(z)$ define a divergência de Kullback-Leibler (K-L), $\psi(z) = (z - 1) \log(z)$ a distância J (ou a versão simétrica da divergência de K-L), $\psi(z) = 0,5|z - 1|$ a distância variacional ou norma L_1 e $\psi(z) = (z - 1)^2$ define a divergência χ^2 . A relação entre a CPO (1.10) e a medida de divergência ψ é dada na seguinte proposição.

Proposição 1. A medida divergência ψ pode ser escrita como

$$D_\psi(P, P_{(-i)}) = E_{\theta|\mathcal{D}} \left[\psi \left(\frac{CPO_i}{g(t_{1i}, t_{2i}|\theta)} \right) \right], \quad (1.13)$$

em que o valor esperado é com relação à distribuição *a posteriori* conjunta $\pi(\theta|\mathcal{D})$.

Prova: Do teorema de Bayes, a distribuição *a posteriori* de θ é dada por

$$\pi(\theta|\mathcal{D}) = \frac{\pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta)}{\int_{\theta \in \Theta} \pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta) d\theta},$$

em que $\pi(\theta)$ e $\prod_{i=1}^n g(t_{1i}, t_{2i}|\theta)$ representam, respectivamente, a distribuição *a priori* e a função verossimilhança de θ . A razão das distribuições *a posteriori* é dada por

$$\begin{aligned} \frac{\pi(\theta|\mathcal{D}^{(-i)})}{\pi(\theta|\mathcal{D})} &= \frac{\pi(\theta) \prod_{j \in \mathcal{D}^{(i)}} g(t_{1i}, t_{2i}|\theta)}{\int_{\theta \in \Theta} \pi(\theta) \prod_{j \in \mathcal{D}^{(i)}} g(t_{1j}, t_{2j}|\theta) d\theta} \times \frac{\int_{\theta \in \Theta} \pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta) d\theta}{\pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta)} \\ &= \frac{1}{g(t_{1i}, t_{2i}|\theta)} \times \frac{\int_{\theta \in \Theta} \pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta) d\theta}{\int_{\theta \in \Theta} \frac{1}{g(t_{1i}, t_{2i}|\theta)} \pi(\theta) \prod_{j \in \mathcal{D}} g(t_{1j}, t_{2j}|\theta) d\theta} \\ &= \frac{\left(\int_{\theta \in \Theta} \frac{1}{g(t_{1i}, t_{2i}|\theta)} \pi(\theta|\mathcal{D}) d\theta \right)^{-1}}{g(t_{1i}, t_{2i}|\theta)} = \frac{CPO_i}{g(t_{1i}, t_{2i}|\theta)}. \quad \square \end{aligned}$$

Da proposição 1, a divergência K-L pode ser expressa como uma esperança *a posteriori*,

$$\begin{aligned} D_{\text{K-L}}(P, P_{(-i)}) &= \log E_{\theta|\mathcal{D}} \{ [g(t_{1i}, t_{2i}|\theta)]^{-1} \} + E_{\theta|\mathcal{D}} \{ \log [g(t_{1i}, t_{2i}|\theta)] \} \\ &= -\log(CPO_i) + E_{\theta|\mathcal{D}} \{ \log [g(t_{1i}, t_{2i}|\theta)] \}. \end{aligned} \quad (1.14)$$

De (1.13) podemos calcular $D_\psi(P, P_{(-i)})$ considerando uma amostra da distribuição *a posteriori* de θ via métodos MCMC. Considere $\theta^{(1)}, \dots, \theta^{(V)}$ uma amostra de tamanho V de $\pi(\theta|\mathcal{D})$. Então, uma estimativa Monte Carlo é dada por

$$\widehat{D}_\psi(P, P_{(-i)}) = \frac{1}{V} \sum_{q=1}^V \psi \left(\frac{\pi(\theta^{(q)}|\mathcal{D}^{(-i)})}{\pi(\theta^{(q)}|\mathcal{D})} \right). \quad (1.15)$$

De (1.14) uma estimativa Monte Carlo da divergência $D_{K-L}(P, P_{(-i)})$ é dada por

$$\widehat{D}_{K-L}(P, P_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{V} \sum_{q=1}^V \log [g(t_{1i}, t_{2i} | \theta^{(q)})]. \quad (1.16)$$

A medida $D_\psi(P, P_{(-i)})$ pode ser interpretada como a divergência ψ do efeito da exclusão do i -ésimo caso dos dados completos na distribuição *a posteriori* de θ .

Como apontado por Peng & Dey (1995) e Weiss (1996) (vide também Cancho *et al.*, 2010), pode ser difícil para um profissional (por exemplo, um médico) avaliar o ponto de corte da medida de divergência, de modo a determinar se uma observação ou um pequeno subconjunto de observações é influente ou não. Neste contexto, usaremos a proposta dada por Peng & Dey (1995) e Weiss (1996). Considere uma moeda viesada com probabilidade de sucesso p . Então, a divergência ψ entre a moeda viesada e a não viesada é

$$D_\psi(f_0, f_1) = \int \psi \left(\frac{f_0(x)}{f_1(x)} \right) f_1(x) dx, \quad (1.17)$$

em que $f_0(x) = p^x(1-p)^{1-x}$ e $f_1(x) = 0,5$, $x = 0, 1$. Se $D_\psi(f_0, f_1) = d_\psi(p)$, então pode ser facilmente verificado que d_ψ satisfaz a seguinte equação

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1-p))}{2}. \quad (1.18)$$

Não é difícil notar que, para as medidas de divergência consideradas, d_ψ aumenta à medida que p afasta-se de 0,5. Além disso, $d_\psi(p)$ é simétrica em torno de $p = 0,5$ e d_ψ atinge seu mínimo em $p = 0,5$. Neste ponto, $d_\psi(0,5) = 0$ e $f_0 = f_1$. Portanto, se considerarmos $p > 0,80$ (ou $p \leq 0,20$) como uma moeda muito viciada, então $d_{L_1}(0,80) = 0,30$. Esta relação implica que o i -ésimo caso é considerado influente quando $d_{L_1}(0,80) > 0,30$.

Assim, se usarmos a divergência de Kullback-Leibler, podemos considerar que uma observação é influente quando $d_{K-L} > 0,223$. Da mesma forma, usando a distância J ou a divergência χ^2 , uma observação na qual $d_J > 0,416$ ou $d_{\chi^2}(0,80) > 0,36$ pode ser considerada influente.

1.4 Dados reais utilizados

1.4.1 Dados de retinopatia diabética

No conjunto de dados de retinopatia diabética (The Diabetic Retinopathy Study Research Group, 1976), o principal objetivo do estudo é avaliar a eficácia do tratamento com fotocoagulação para retinopatia proliferativa. O tratamento foi aleatoriamente atribuído para um olho de cada paciente e, o outro foi considerado como controle. O primeiro componente é um vetor de tempos de vidas que representa os tempos até a perda visual para o olho de tratamento (T_1), enquanto o segundo componente (T_2) é o vetor de tempos até a perda visual para o olho controle. Este conjunto de dados é composto por 197 pacientes e, contém observações censuradas, o que aconteceu em 73% dos olhos tratados e 49% dos olhos não tratados.

A idade no início da diabetes foi considerada como covariável. Para criar dois grupos foi considerado um ponto de corte de 20 anos (58% dos pacientes tinham menos de 20 anos de idade).

Na literatura, muitas abordagens com a finalidade de identificar a forma da função de risco podem ser encontradas (vide, por exemplo, Glaser, 1980). Com a finalidade de identificar o modelo adequado para os tempos de vida, utilizamos neste trabalho um método gráfico baseado no tempo total em teste (TTT) transformado (Barlow & Campo, 1975). A versão empírica do gráfico TTT transformado (Aarset, 1985) é dada por

$$G(r/n) = \frac{\sum_{i=1}^r t_{(i)} - (n-r)t_{(r)}}{\sum_{i=1}^r t_{(i)}}, \quad (1.19)$$

$r = 1, \dots, n$, e $t_{(i)}$ representa as estatísticas de ordem dos tempos, $i = 1, \dots, n$. A curva TTT pode exibir várias formas. Aarset (1985) mostrou que se a curva é côncava, a função de risco é crescente. Na Figura 1.1 apresentamos as estimativas de Kaplan-Meier da função de sobrevivência e o gráfico TTT para os dados de retinopatia diabética. Nesta figura, podemos observar pelo gráfico de Kaplan-Meier a existência de uma fração de indivíduos que poderiam ser ditos curados. Pelo gráfico TTT notamos algumas das observações estão localizadas ao redor da reta identidade, mas

em sua maioria acima, concluindo que a função de risco é crescente, sugerindo que o modelo Weibull é uma alternativa adequada para a modelagem destes dados.

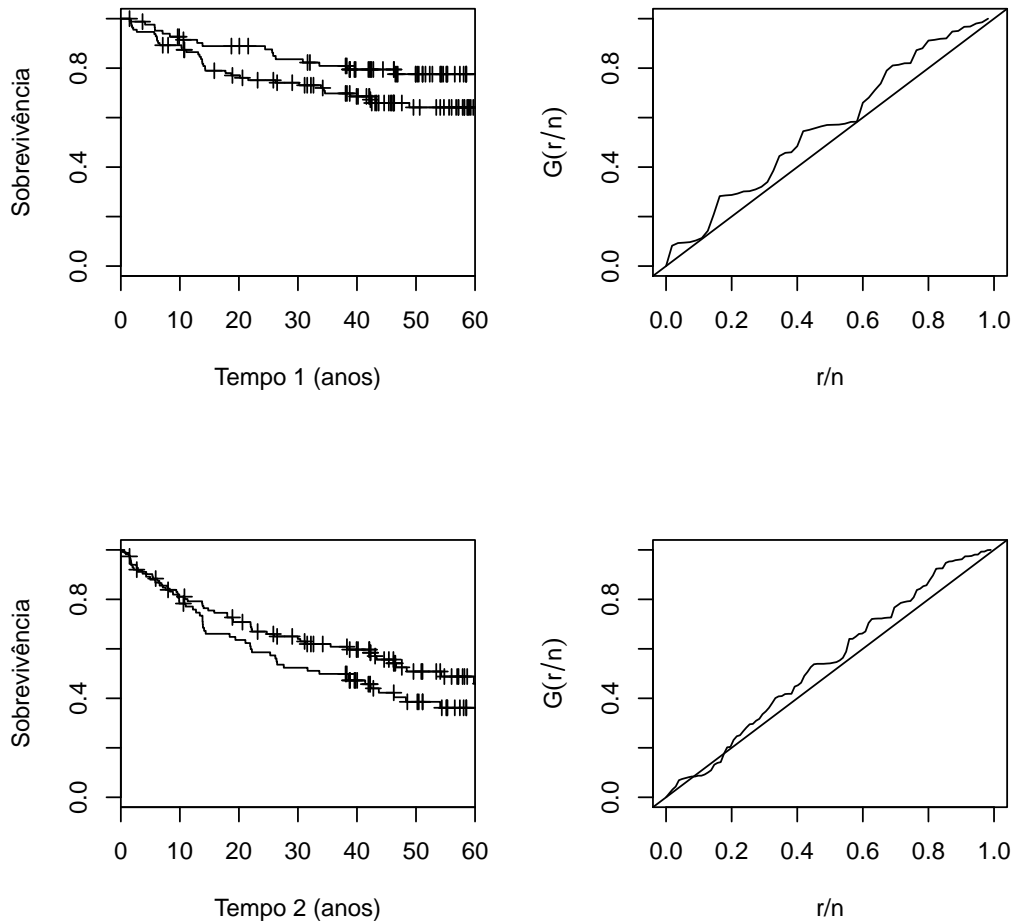


Figura 1.1: Estimativas de Kaplan-Meier da função de sobrevivência e gráficos TTT para os dados de retinopatia diabética.

1.4.2 Dados de HIV

O conjunto de dados de pacientes portadores de HIV é composto por 135 pacientes com idade superior a 18 anos de idade, atendidos no Serviço de Doenças Infecciosas e Parasitárias (DIP) da Universidade Federal do Triângulo Mineiro (UFTM), diagnosticados com HIV entre janeiro de 1996 e dezembro de 1998, os quais foram

gentilmente fornecidos por Gilberto A. Pereira e Fatima L. Prince (UFTM).

Infecções oportunistas são importantes causas de morbidade e mortalidade em pessoas com o vírus da imunodeficiência humana (HIV) e mortes e hospitalizações são os principais eventos resultantes dessas infecções.

Além disso, as infecções oportunistas são importantes indicadores do impacto das intervenções na população infectada pelo HIV (Candiani *et al.*, 2007). Pessoas infectadas pelo HIV estão sujeitas a inúmeras infecções e complicações neoplásicas relacionadas à sua doença. Em geral, além de infecções adquiridas ao viver fora do hospital, quando hospitalizado o indivíduo infectado pelo HIV também é mais suscetível a infecções hospitalares por sua possível condição de saúde debilitada.

No Brasil, segundo Candiani *et al.* (2007), o tempo médio de permanência no hospital, que foi igual a 15 dias antes de 1996, reduziu de 40% após este ano.

Neste conjunto de dados, os tempos da primeira (T_1) e da segunda (T_2) internação (em dias) foram modelados de acordo com o sexo do paciente. Para os tempos de vida 1 menores que 25 dias, 20% foram censurados. Para o tempo de vida 2 menores que 20 dias, 40% foram censurados. Tempos maiores que estes valores foram censurados.

Na Figura 1.2 apresentamos as estimativas de Kaplan-Meier da função de sobrevivência e os gráficos TTT para os dados de HIV. Nesta figura, podemos observar pelo gráfico de Kaplan-Meier a existência de uma fração de indivíduos que poderiam ser ditos curados, pelo menos no que concerne ao intervalo de tempo abrangido pelo estudo. Também constatamos pelo gráfico TTT que a função de risco é crescente, sugerindo que o modelo Weibull é uma alternativa adequada para a modelagem destes dados.

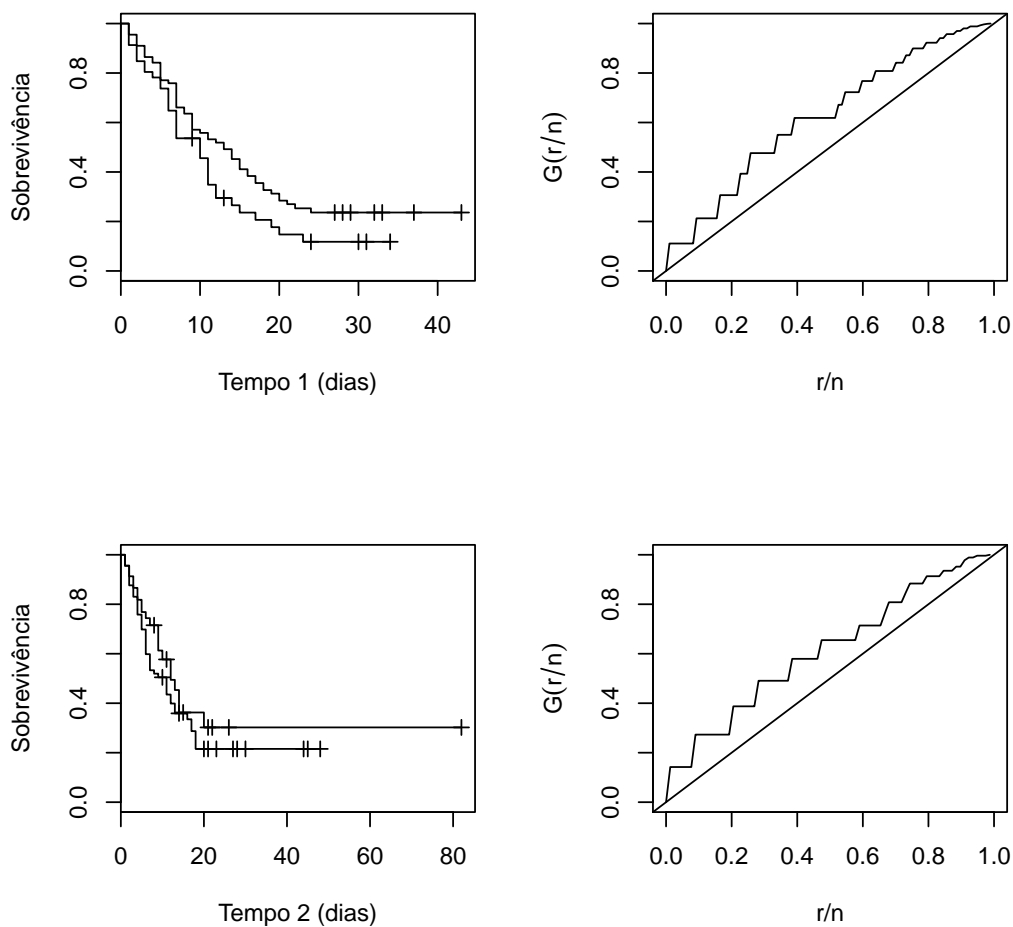


Figura 1.2: Estimativas de Kaplan-Meier da função de sobrevivência e gráficos TTT para os dados de HIV.

1.5 Organização dos capítulos

No Capítulo 2 apresentamos o modelo de sobrevivência FGM bivariado assumindo distribuições marginais Weibull.

No Capítulo 3 apresentamos o modelo de sobrevivência FGM bivariado considerando distribuições marginais mistura padrão Weibull em duas situações: a primeira com a inclusão de covariáveis no parâmetro do modelo Weibull e na segunda na proporção de curados.

No Capítulo 4 apresentamos o modelo de sobrevivência FGM bivariado assumindo modelos de tempo de promoção Weibull para as distribuições marginais.

Em todos estes capítulos, exibimos os procedimentos inferenciais sob um enfoque bayesiano, comparamos com outros modelos de cópulas e com o caso assumindo independência entre os tempos marginais. Também realizamos o procedimento de análise de influência caso a caso baseado na divergência ψ . Ilustrações numéricas são apresentadas considerando conjunto de dados artificiais e reais.

Por fim, no Capítulo 5 apresentamos algumas considerações finais e extensões.

Capítulo 2

Modelo de sobrevivência bivariado baseado na cópula FGM

Neste capítulo propomos um modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull. A construção do modelo é apresentada na Seção 2.1. Na Seção 2.2 apresentamos o procedimento inferencial. Aplicações com dados simulados e reais são realizadas na Seção 2.3.

2.1 Modelo de sobrevivência FGM bivariado

Com o objetivo de definir a função de sobrevivência bivariado, primeiramente supomos que C_ϕ é uma função distribuição com função densidade c_ϕ em $[0, 1]^2$ para $\phi \in \mathbb{R}$. Então, considere (T_1, T_2) os tempos de falhas pareados e, S_j e f_j , respectivamente, a função de sobrevivência e a função densidade de T_j , $j = 1, 2$.

Se (T_1, T_2) provém da cópula C_ϕ para algum ϕ , então as funções de sobrevivência e densidade conjunta de (T_1, T_2) são dadas por (Shih & Louis, 1995)

$$S(t_1, t_2) = C_\phi(S_1(t_1), S_2(t_2)), \quad t_1, t_2 > 0 \quad (2.1)$$

e

$$f(t_1, t_2) = c_\phi(S_1(t_1), S_2(t_2))f_1(t_1)f_2(t_2), \quad t_1, t_2 > 0, \quad (2.2)$$

respectivamente. Note que as distribuições marginais e a estrutura de dependência

podem ser visualizadas separadamente e esta estrutura de dependência é representada por uma cópula.

A função de sobrevivência bivariada induzida pela cópula FGM (1.9) é dada por

$$S(t_1, t_2) = S_1(t_1)S_2(t_2)[1 + \phi\{1 - S_1(t_1)\}\{1 - S_2(t_2)\}], \quad (2.3)$$

em que ϕ , $\phi \in (-1, 1)$, é o parâmetro que mede a intensidade da dependência entre os tempos de vida. Observe que quando $\phi = 0$, $S(t_1, t_2) = S_1(t_1)S_2(t_2)$, levando à conclusão de que as variáveis aleatórias T_1 e T_2 são independentes.

Para a cópula FGM, o tau de Kendall (Nelsen, 2006) é dado por $\tau_\phi = \frac{2\phi}{9}$. Para $\phi = 0$, τ_ϕ é igual a 0. Outras medidas de dependência entre as variáveis T_1 e T_2 podem ser encontradas em Nelsen (2006). Por exemplo, se as variáveis aleatórias T_1 e T_2 têm uma distribuição Weibull com função densidade

$$f(t_j) = \alpha_j \lambda_j t_j^{\alpha_j - 1} \exp\{-\lambda_j t_j^{\alpha_j}\}, \quad j = 1, 2, \quad (2.4)$$

e função de sobrevivência

$$S(t_j) = \exp\{-\lambda_j t_j^{\alpha_j}\}, \quad j = 1, 2, \quad (2.5)$$

pode ser provado que o coeficiente de correlação de Pearson entre T_1 e T_2 é dado por

$$\rho = \frac{\phi \left(1 - 2^{-\frac{1}{\alpha_1}}\right) \left(1 - 2^{-\frac{1}{\alpha_2}}\right)}{\sqrt{\left(\frac{\Gamma(\frac{2}{\alpha_1} + 1)}{\Gamma(\frac{1}{\alpha_1} + 1)^2} - 1\right) \left(\frac{\Gamma(\frac{2}{\alpha_2} + 1)}{\Gamma(\frac{1}{\alpha_2} + 1)^2} - 1\right)}}. \quad (2.6)$$

Observe que quando $\phi = 0$, o coeficiente de correlação ρ é igual a 0. Quando T_j tem uma distribuição Exponencial com parâmetro $\alpha_j = 1$, $j = 1, 2$, o coeficiente de correlação é dado por $\rho = \phi/4$.

2.2 Inferência

Para inferência, adotamos uma abordagem bayesiana. A função verossimilhança, as distribuições *a priori* para os parâmetros do modelo e os detalhes do algoritmo Monte Carlo em Cadeias de Markov (MCMC) são descritos a seguir.

Considere (T_{i1}, T_{i2}) e (C_{i1}, C_{i2}) os i -ésimos tempos de vida e de censura bivariados, para $i = 1, \dots, n$. Suponha que (T_{i1}, T_{i2}) e (C_{i1}, C_{i2}) são independentes. Para cada indivíduo i , as quantidades individuais são representadas pelas variáveis aleatórias $t_{ij} = \min(T_{ij}, C_{ij})$ e $\delta_{ij} = I(t_{ij} = T_{ij})$, que denota o indicador de falha, $j = 1, 2$. Sejam $S(t_1|\boldsymbol{\gamma}_1)$ e $S(t_2|\boldsymbol{\gamma}_2)$ as funções de sobrevivência de T_{i1} e T_{i2} , respectivamente, em que $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$ são vetores de parâmetros de q_1 e q_2 elementos associados a cada uma das distribuições marginais.

Considerando a função de sobrevivência bivariada $S(t_1, t_2|\phi, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ dada em (2.3), a contribuição do i -ésimo indivíduo para a log-verossimilhança de $\boldsymbol{\theta} = (\phi, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ é dada por (Lawless, 2003)

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) = & \delta_{i1}\delta_{i2} \log \left(\frac{\partial^2 S(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i1} \partial t_{i2}} \right) + \delta_{i1}(1 - \delta_{i2}) \log \left(\frac{-\partial S(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i1}} \right) \\ & + \delta_{i2}(1 - \delta_{i1}) \log \left(\frac{-\partial S(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i2}} \right) + (1 - \delta_{i1})(1 - \delta_{i2}) \log S(t_1, t_2|\boldsymbol{\theta}). \end{aligned} \quad (2.7)$$

2.2.1 Distribuições *a priori* e *a posteriori*

O uso do método bayesiano permite a incorporação de conhecimentos prévios dos parâmetros através de uma distribuição *a priori* informativa. Quando não há conhecimentos prévios pode-se considerar uma estrutura não informativa.

Por exemplo, a fim de realizar um procedimento de inferência bayesiana para o modelo de sobrevivência bivariado dado em (2.3), considere que a família de distribuições paramétricas dos tempos de vida marginal T_1 e T_2 é conhecida e indexada pelos vetores de parâmetros $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$, respectivamente.

Consideramos uma distribuição *a priori* conjunta própria para os parâmetros do modelo com o intuito de garantir que a distribuição *a posteriori* conjunta seja própria (Ibrahim *et al.*, 2001). A densidade *a priori* conjunta de $\boldsymbol{\theta} = (\phi, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ é dada por

$$\pi(\boldsymbol{\theta}) \propto (1 - \phi)^{r_1 - 1} (1 + \phi)^{r_2 - 1} \prod_{j=1}^2 \pi(\boldsymbol{\gamma}_j). \quad (2.8)$$

A expressão (2.8) implica que $(1 - \phi)/2$ segue uma distribuição *Beta*($r_1; r_2$), $\pi(\boldsymbol{\gamma}_j)$ é a distribuição *a priori* de $\boldsymbol{\gamma}_j$ e que $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ e ϕ são mutuamente independentes *a priori*.

Combinando (2.8) com a função de verossimilhança, $L(\boldsymbol{\theta}) = \exp(\sum_{i=1}^n \ell_i(\boldsymbol{\theta}))$, em que $\ell_i(\boldsymbol{\theta})$ é dada em (2.7), obtemos diretamente a distribuição conjunta *a posteriori* de $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\mathcal{D})$, em que \mathcal{D} é o conjunto de dados observados. As estimativas dos parâmetros são dadas pelas médias da distribuição *a posteriori*.

Para especificação paramétrica das marginais é possível ligar covariáveis em cada um dos componentes, isto é, $\gamma_j = \varphi(x_j)$. Por simplicidade e para evitar restrições no espaço paramétrico, consideramos $\varphi(x_j) = \exp(\mathbf{x}_j^\top \boldsymbol{\beta}_j)$, em que $\boldsymbol{\beta}_j$ corresponde ao vetor de coeficientes desconhecidos de ordem $q_j \times 1$, associado às covariáveis \mathbf{x}_j , $j = 1, 2$.

2.2.2 Implementação Computacional

Na abordagem bayesiana, a distribuição alvo para a inferência é a distribuição *a posteriori* dos parâmetros de interesse. Para isto, precisamos obter a densidade *a posteriori* marginal de cada parâmetro, que são obtidas através da integração da densidade conjunta *a posteriori* com relação aos demais parâmetros.

O amostrador de Gibbs é um processo iterativo de uma ampla classe de métodos genericamente nomeados Monte Carlo em Cadeias de Markov (MCMC). Muitos aspectos práticos da metodologia MCMC são descritos em Gelfand & Smith (1990) e Gamerman & Lopes (2006). Este método é aplicável em situações em que não somos capazes de gerar amostras diretamente da distribuição *a posteriori* conjunta. Contudo, requer as densidades condicionais completas para gerar amostras.

Ressaltamos que para qualquer distribuição marginal de T_1 e T_2 , a distribuição *a posteriori* conjunta não é tratável analiticamente, mas métodos MCMC tal como o amostrador de Gibbs podem ser usados para obter amostras, a partir da quais características das distribuições *a posteriori* marginais de interesse podem ser inferidas (Gilks *et al.*, 1996). Para o procedimento de estimação, consideramos a estimação conjunta em que todos os parâmetros do modelo são estimados simultaneamente no algoritmo MCMC.

Assumindo que T_{ij} segue uma distribuição Weibull dada em (2.4) com parâmetros α_j e $\lambda_{ij} = \exp(\beta_{0j} + \beta_{1j}x_i)$, $j = 1, 2$. Escolhemos as seguintes distribuições *a priori*

independentes: $\beta_{kj} \sim N(\mu_{kj}; \sigma_{kj}^2)$ e $\alpha_j \sim \text{Gama}(a_j; b_j)$, em que $k = 0, 1$ e $j = 1, 2$.

A distribuição condicional para cada parâmetro é dada por

$$\pi(\phi | \mathcal{D}, \theta_{(-\phi)}) \propto (1 - \phi)^{r_1 - 1} (1 + \phi)^{r_2 - 1} \left(\prod_{i=1}^n \Delta_i \right),$$

$$\pi(\beta_{0j} | \mathcal{D}, \theta_{(-\beta_{0j})}) \propto \left(\prod_{i=1}^n \Delta_i \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} + \frac{\mu_{0j}}{\sigma_{0j}^2} \right\} \beta_{0j} - \frac{\beta_{0j}^2}{2\sigma_{0j}^2} - \sum_{i=1}^n e^{(\beta_{0j} + \beta_{1j} x_i)} t_{ji}^{\alpha_j} \right),$$

$$\pi(\beta_{1j} | \mathcal{D}, \theta_{(-\beta_{1j})}) \propto \left(\prod_{i=1}^n \Delta_i \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} x_i + \frac{\mu_{1j}}{\sigma_{1j}^2} \right\} \beta_{1j} - \frac{\beta_{1j}^2}{2\sigma_{1j}^2} - \sum_{i=1}^n e^{(\beta_{0j} + \beta_{1j} x_i)} t_{ji}^{\alpha_j} \right)$$

e

$$\pi(\alpha_j | \mathcal{D}, \theta_{(-\alpha_j)}) \propto \left(\prod_{i=1}^n \Delta_i t_{ji}^{(\alpha_j - 1) \delta_{ji}} \right) \alpha_j^{(\alpha_j - 1) (\sum_{i=1}^n \delta_{ji})} \exp \left(- \sum_{i=1}^n (\beta_{0j} + \beta_{1j} x_i) t_{ji}^{\alpha_j} - b_j \alpha_j \right),$$

para $j = 1, 2$ com $\mathcal{D} = (\mathbf{t}_1, \mathbf{t}_2, \delta_1, \delta_2, \mathbf{x})$,

$$\Delta_i = \nu_{1i}^{(1-\delta_{1i})(1-\delta_{2i})} \nu_{2i}^{\delta_{1i}\delta_{2i}} \nu_{3i}^{\delta_{1i}(1-\delta_{2i})} \nu_{4i}^{\delta_{2i}(1-\delta_{1i})},$$

$$\nu_{1i} = 1 + \phi \prod_{j=1}^2 (1 - S_j(t_{ji})),$$

$$\nu_{2i} = 1 + \phi \left(1 + \prod_{j=1}^2 (1 - 2S_j(t_{ji})) \right),$$

$$\nu_{3i} = 1 + \phi \left(1 - 2S_1(t_{ji}) - S_2(t_{ji}) + 2 \prod_{j=1}^2 S_j(t_{ji}) \right)$$

e

$$\nu_{4i} = 1 + \phi \left(1 - S_1(t_{ji}) - 2S_2(t_{ji}) + 2 \prod_{j=1}^2 S_j(t_{ji}) \right).$$

As densidades condicionais acima não pertencem a nenhuma família de densidades paramétricas conhecidas. A fim de gerar as amostras temos então que implementar um passo de Metropolis nas iterações Gibbs (Chib & Greenberg, 1995). Para o parâmetro ϕ , a distribuição condicional $\pi(\phi | \mathcal{D}, \theta_{(-\phi)})$ é uma função log-côncava. Assim, o método de rejeição adaptativa pode ser usada.

2.3 Aplicações

Nesta seção resultados de estudos de simulação e um exemplo de dados reais são apresentados com o objetivo de ilustrar a performance da metodologia proposta.

2.3.1 Propriedades frequentistas

Inicialmente empregamos dados simulados para estudar as propriedades frequentistas dos estimadores bayesianos quando os parâmetros do modelo são conhecidos. O objetivo deste estudo de simulação é mostrar o bom comportamento das estimativas bayesianas, com base no erro quadrático médio frequentista (EQM), na média frequentista (Média) e nas medidas utilizadas para comparação de modelos (EAIC, EBIC, DIC, LPML).

Os tempos de censura C_{ij} foram gerados a partir de uma distribuição Uniforme $U(0; \tau_j)$, com $\tau_j, j = 1, 2$, controlando o percentual de observações censuradas. Neste estudo consideramos o modelo de sobrevivência dado pela expressão (2.3), assumindo que T_j tem distribuição Weibull com parâmetros α_j e $\lambda_{ij} = \exp(\beta_{0j} + \beta_{1j}x_i)$, $j = 1, 2$, em que as covariáveis x_i foram geradas de uma distribuição Bernoulli com parâmetro 0,5. Para os dados bivariados (T_{i1}, T_{i2}) , $i = 1, \dots, n$, foram gerados $n = 100$ observações de acordo com os seguintes passos. Primeiro geramos $T_{i1} = (-\log(1 - u_{i1})/\lambda_{i1})^{1/\alpha_1}$ em que $u_{i1} \sim U(0; 1)$. Então, T_{i1} foi comparado com o valor de censura C_{i1} a fim de determinar o indicador de censura δ_{i1} e o valor observado dado por $t_{i1} = \min(T_{i1}, C_{i1})$. Depois T_{i2} foi gerado usando a variável aleatória $u_{i2} \sim U(0; 1)$ e a solução da equação não linear, $w_i + \phi(2u_{i2} - 1)(w_i^2 - w_i) - u_{i2} = 0$, considerando $T_{i2} = (-\log(1 - w_i)/\lambda_{i2})^{1/\alpha_2}$. Então, T_{i2} foi comparado com o valor de censura C_{i2} a fim de determinar o indicador de censura δ_{i2} e o valor observado dado por $t_{i2} = \min(T_{i2}, C_{i2})$.

Em nosso estudo consideramos os seguintes valores para os parâmetros do modelo $\beta_{01} = 1$, $\beta_{11} = 1$, $\alpha_1 = 2$, $\beta_{02} = 1,5$, $\beta_{12} = 0,5$, $\alpha_2 = 1,5$ e $\phi = 0,5$. Simulamos os conjuntos de dados assumindo (0%; 0%) e (15%; 15%) de censuras.

Portanto, duas configurações diferentes de simulação foram realizadas de acordo

com o percentual de censura, cada um com 100 conjuntos Monte Carlo de dados gerados.

As seguintes distribuições *a priori* independentes foram consideradas para o amostrador de Gibbs: $\beta_{ij} \sim N(0; 1000)$ e $\alpha_j \sim \text{Gama}(1; 0,001)$, $i = 0, 1$ e $j = 1, 2$. Assumimos para o parâmetro da cópula PS que $\phi \sim \text{Beta}(1; 1)$ e para a cópula FGM que $(1 - \phi)/2 \sim \text{Beta}(1; 1)$, tal escolha garante que $\phi \in (-1; 1)$.

Para cada conjunto de dados gerados simulamos duas cadeias de tamanho 50.000 para cada parâmetro, desconsiderando as primeiras 10.000 iterações para eliminar o efeito dos valores iniciais e, para evitar problemas de autocorrelação, consideramos um espaçamento de tamanho 20, obtendo uma amostra efetiva de tamanho 4.000 sobre a qual a inferência *a posteriori* é baseada. Para cada amostra, a média *a posteriori* dos parâmetros e os valores de EAIC, EBIC, DIC e LPML são gravados. As simulações foram realizadas utilizando os sistemas OpenBUGS (Spiegelhalter *et al.*, 2007) e R (R Development Core Team, 2007) por meio do pacote BRugs (BUGS team, 2007).

A convergência das cadeias foi monitorada de acordo com os métodos recomendados por Cowless & Carlin (1996) (pacote CODA (Plummer *et al.*, 2006)). Em todos os casos, a convergência foi verificada por meio do diagnóstico de Gelman-Rubin (Gelman & Rubin, 1992) sendo muito próximo a 1 (≤ 1.02).

A Tabela 2.1 mostra as estatísticas resumo da simulação para os parâmetros ajustando o modelo baseado na cópula FGM para as duas configurações de censuras. Os verdadeiros valores são dados entre parênteses, a Média MC denota a média aritmética das 100 estimativas dada por $\sum_{j=1}^{100} \hat{\theta}_{kj}/100$, e o EQM MC denota o erro quadrático médio frequentista dado por $\sum_{j=1}^{100} (\hat{\theta}_{kj} - \theta_k)^2/100$, em que $\hat{\theta}_k$ é o valor estimado do parâmetro fixo θ_k . Podemos observar que em ambos os casos (com e sem a presença de dados censurados) as estimativas obtidas estão próximas, em média, do verdadeiro valor e têm um baixo valor de EQM.

A Tabela 2.2 apresenta a média Monte Carlo (MC) dos quatro critérios de comparação de modelos discutidos na Seção 1.2 para comparar os modelos de sobrevivência bivariado baseado na cópula FGM e PS com marginais Weibull. Podemos

Tabela 2.1: Dados simulados. Média MC e EQM MC para os parâmetros do modelo de sobrevivência FGM bivariado baseados sobre as 100 amostras geradas para as duas configurações de censuras.

Parâmetro (verdadeiro valor)	% de censurados			
	(0; 0)		(15; 15)	
	Média MC	EQM MC	Média MC	EQM MC
$\beta_{01}(1)$	1,004	0,012	0,884	0,020
$\beta_{11}(1)$	1,051	0,033	1,033	0,034
$\alpha_1(2)$	2,040	0,015	2,096	0,022
$\beta_{02}(1, 5)$	1,536	0,019	1,398	0,021
$\beta_{12}(0, 5)$	0,522	0,026	0,491	0,024
$\alpha_2(1, 5)$	1,546	0,008	1,567	0,011
$\phi(0, 5)$	0,475	0,021	0,374	0,036

observar que para as duas configurações de censura, o modelo baseado na cópula FGM supera o modelo PS em todas as médias dos critérios considerados.

Tabela 2.2: Dados simulados. Média Monte Carlo dos critérios bayesianos baseados sobre as 100 amostras geradas para as duas configurações de censuras.

% de censurados	Cópula	Critérios bayesianos			
		EAIC	EBIC	DIC	LPML
(0; 0)	FGM	-99,203	-80,967	-106,486	53,079
	PS	-98,115	-79,879	-105,463	52,564
(15; 15)	FGM	-27,325	-9,089	-34,578	17,028
	PS	-26,728	-8,492	-34,182	16,822

2.3.2 Diagnóstico de observações influentes

Para examinar o desempenho da medida de diagnóstico proposta neste trabalho, geramos uma amostra de tamanho 100 com parâmetros fixos $\beta_{01} = 1$, $\beta_{11} = 1$, $\alpha_1 = 2$, $\beta_{02} = 1,5$, $\beta_{12} = 0,5$, $\alpha_2 = 1,5$ e $\phi = 0,5$. Nesta amostra, consideramos

também que 15% de cada tempo foi censurado.

Selecionamos os casos 12, 28, 50 (ambos os tempos observados) e 77 (ambos os tempos censurados) para perturbação. Para criar observações artificialmente influentes no conjunto de dados, escolhemos um, dois ou três desses casos selecionados. Para cada caso, perturbamos um ou ambos os tempos da seguinte forma: $\tilde{t}_i = t_i + 4D_t, i = 1, 2$, em que D_t é o desvio padrão dos t_i 's. Para o caso 12 foi perturbado apenas o tempo de vida t_1 , para o caso 50 o tempo de vida t_2 e, para ambos os casos 28 e 77, ambos os tempos de vidas foram perturbados.

Para a implementação do algoritmo MCMC, assim como a verificação da convergência das cadeias, realizamos os mesmos procedimentos descritos na Subseção 2.2.2.

A Tabela 2.3 mostra que as inferências *a posteriori* são sensíveis à perturbação do(s) caso(s) selecionado(s). Na Tabela 2.3 o conjunto de dados (a) denota os dados originais simulados sem perturbação e os conjuntos de dados (b) a (i) denotam os conjuntos de dados com casos perturbados.

Tabela 2.3: Dados simulados. Média e desvio padrão (DP) para os parâmetros do modelo de sobrevivência FGM bivariado para cada conjunto de dados simulados.

Nome dos dados	Identificação dos casos perturbados	β_{01} Média [DP]	β_{11} Média [DP]	α_1 Média [DP]	β_{02} Média [DP]	β_{12} Média [DP]	α_2 Média [DP]	ϕ Média [DP]
a	Amostra original	0,939 [0,156]	0,856 [0,230]	2,111 [0,184]	1,115 [0,174]	0,594 [0,225]	1,400 [0,120]	0,386 [0,28]
b	12	0,768 [0,154]	0,939 [0,234]	1,982 [0,167]	1,112 [0,173]	0,607 [0,224]	1,403 [0,122]	0,419 [0,273]
c	50	0,941 [0,160]	0,853 [0,230]	2,106 [0,185]	0,974 [0,166]	0,681 [0,228]	1,340 [0,112]	0,328 [0,277]
d	77	0,847 [0,154]	0,472 [0,219]	1,825 [0,147]	1,063 [0,166]	0,382 [0,223]	1,313 [0,112]	0,508 [0,272]
e	{12, 50}	0,770 [0,151]	0,932 [0,228]	1,984 [0,164]	0,977 [0,166]	0,673 [0,226]	1,336 [0,115]	0,360 [0,287]
f	{12, 77}	0,717 [0,152]	0,564 [0,222]	1,757 [0,139]	1,060 [0,167]	0,385 [0,222]	1,311 [0,111]	0,536 [0,272]
g	{12, 50, 77}	0,718 [0,149]	0,573 [0,225]	1,753 [0,139]	0,946 [0,165]	0,460 [0,223]	1,262 [0,106]	0,481 [0,279]
h	28	0,797 [0,156]	0,403 [0,220]	1,700 [0,134]	1,054 [0,171]	0,374 [0,223]	1,293 [0,110]	0,491 [0,283]
i	{28, 77}	0,776 [0,151]	0,234 [0,218]	1,639 [0,132]	1,025 [0,169]	0,219 [0,228]	1,249 [0,105]	0,602 [0,263]

A Tabela 2.4 mostra os critérios bayesianos do ajuste de diferentes casos de conjuntos de dados perturbados. Podemos observar que o conjunto de dados (a) (conjunto dos dados originais simulados) teve o melhor ajuste.

Vamos considerar as amostras da distribuição *a posteriori* dos parâmetros do modelo de sobrevivência FGM bivariado para obter uma estimativa das quatro medidas de divergência. Os resultados na Tabela 2.5 mostram, antes da perturbação (conjunto de dados (a)), que todos os casos selecionados não são influentes, com pequenas medidas de divergência. Entretanto, após perturbações (conjunto de dados (b) a (i)) as quatro medidas aumentam, indicando que os casos são influentes. Os

Tabela 2.4: Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência FGM bivariado para cada conjunto de dados simulados.

Nome dos dados	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
a	-1,685	16,551	-8,779	4,387
b	7,424	25,660	0,280	-0,396
c	6,581	24,817	-0,506	0,210
d	27,382	45,618	20,110	-11,871
e	15,742	33,979	8,599	-4,631
f	34,590	52,826	27,250	-15,411
g	42,460	60,696	35,180	-19,144
h	35,063	53,299	27,790	-17,542
i	52,696	70,932	45,320	-24,926

valores que ultrapassam os pontos de cortes apresentados na Subseção 1.3.1 estão destacados com *.

Tabela 2.5: Dados simulados. Medidas de divergência.

Nome dos dados	Identificação dos casos	Medidas de divergência			
		d_{K-L}	d_J	d_{L_1}	d_{χ^2}
a	12	0,012	0,022	0,059	0,023
	28	0,126	0,244	0,193	0,319
	50	0,018	0,037	0,076	0,038
	77	0,008	0,015	0,050	0,016
b	12	0,485*	1,038*	0,393*	2,170*
c	50	0,344*	0,728*	0,328*	1,462*
d	77	2,589*	5,725*	0,799*	79,903*
e	12	0,498*	1,386*	0,418*	13,387*
	50	0,319*	0,695*	0,323*	1,221*
f	12	0,303*	0,732*	0,321*	2,055*
	77	2,373*	7,958*	0,900*	855,628*
g	12	0,270*	0,570*	0,293*	0,920*
	50	0,278*	0,555*	0,289*	0,948*
	77	2,077*	4,101*	0,708*	24,424*
h	28	4,603*	10,855*	1,029*	518,161*
i	28	2,415*	5,410*	0,798*	55,341*
	77	0,866*	1,910*	0,512*	7,589*

A Figura 2.1 mostra os gráficos de índices das quatro medidas de divergência para o conjunto de dados (i). Claramente, podemos ver que as quatro medidas de divergência detectam os pontos influentes.

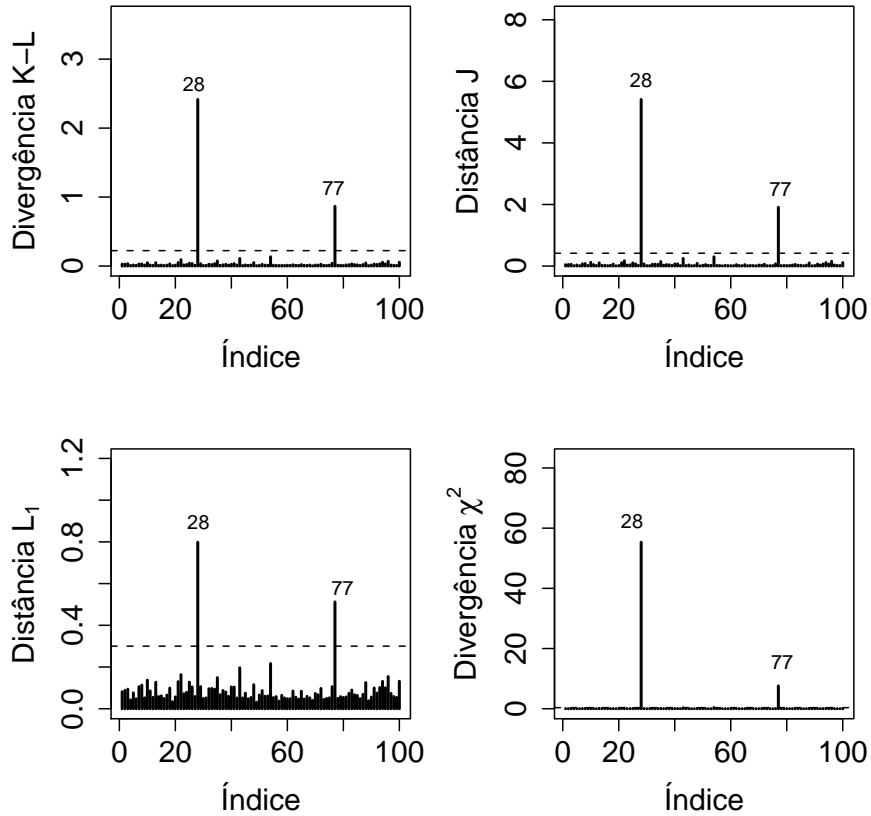


Figura 2.1: Dados simulados. Gráficos de índices das medidas de divergência para o caso (i).

2.3.3 Dados reais

Nesta seção, ilustramos a metodologia proposta em uma amostra composta por 162 pacientes, extraída dos dados de retinopatia diabética (Subseção 1.4.1).

Ao analisarmos este conjunto de dados, não detectamos nenhum ponto influente. Dessa forma, para criar uma observação influente, escolhemos o caso 100 e perturbamos ambos os tempos de vida da seguinte forma $\tilde{t}_u = t_u + 10S_t$, $u = 1, 2$, em que S_t é o desvio-padrão de t_u 's.

Simulamos duas cadeias de tamanho 50.000 para cada parâmetro, desconsiderando as primeiras 10.000 iterações para eliminar o efeito dos valores iniciais e, para evitar problemas de autocorrelação, foi considerado um espaçamento de tamanho 20,

obtendo uma amostra efetiva de tamanho 4.000 sobre a qual a inferência *a posteriori* é baseada.

As seguintes distribuições *a priori* independentes foram consideradas para realizar o amostrador de Gibbs: $\beta_{ji} \sim N(0; 1000)$, $\alpha_j \sim Gama(1; 0,001)$, $i = 0, 1$, $j = 1, 2$ e $(1 - \phi)/2 \sim Beta(1; 1)$, tal escolha garante que $\phi \in (-1; 1)$. A convergência das cadeias foi monitorada de acordo com os métodos recomendados por Cowless & Carlin (1996).

Na Tabela 2.6 apresentamos os resumos *a posteriori* para os parâmetros do modelo baseado na Cópula FGM com marginais Weibull e Exponencial. Além de avaliar a robustez do modelo relacionado às escolhas dos hiperparâmetros das distribuições *a priori*, foi realizado um estudo de sensibilidade no qual constatamos que as estimativas dos parâmetros *a posteriori* não apresentaram diferenças significativas e não alteraram os resultados da Tabela 2.6.

Tabela 2.6: Dados de retinopatia diabética. Média *a posteriori*, desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência FGM bivariado.

Parâmetro	Weibull			Exponencial			
	Média	DP	HPD (95%)	Média	DP	HPD (95%)	
	α_1	0,817	0,102	(0,629; 1,032)	-	-	-
Tempo 1	β_{01}	-3,917	0,416	(-4,767; -3,148)	-4,606	0,169	(-4,948; -4,294)
	β_{11}	-0,676	0,309	(-1,297; -0,097)	-0,682	0,317	(-1,323; -0,078)
	α_2	0,821	0,070	(0,687; 0,964)	-	-	-
Tempo 2	β_{02}	-3,489	0,292	(-4,076; -2,933)	-4,154	0,143	(-4,446; -3,887)
	β_{12}	0,332	0,203	(-0,065; 0,729)	0,325	0,203	(-0,071; 0,722)
Cópula	ϕ	0,655	0,221	(0,173; 0,979)	0,641	0,212	(0,178; 0,971)

A Figura 2.2 mostra o gráfico de dispersão entre T_1 e T_2 (painel acima). Também, as curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita) dicotomizadas pela idade do paciente juntamente com o ajuste do modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull.

Na Tabela 2.7 apresentamos o valor das quatro medidas de divergência para o caso 100. Em comparação com pontos de cortes apresentados na Subseção 1.3.1,

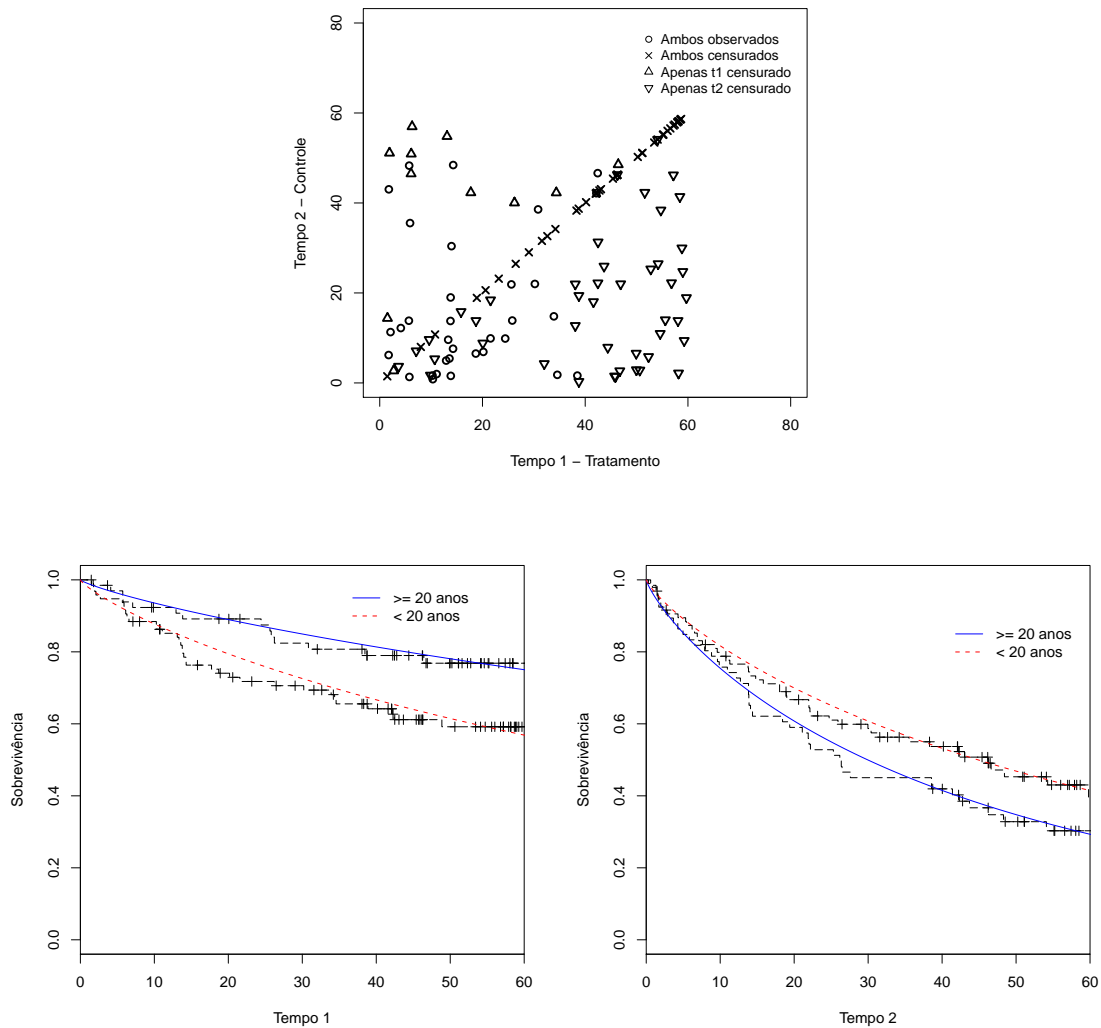


Figura 2.2: Dados de retinopatia diabética. Gráfico de dispersão de T_1 e T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull.

podemos observar que todas as medidas detectam o ponto influente, que também pode ser observado no gráfico de índices (Figura 2.3).

A Tabela 2.8 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull e Exponencial com os modelos baseados nas cópulas de Clayton, Frank e Positive Stable, como também com o caso assumindo independência entre os

Tabela 2.7: Dados de retinopatia diabética. Medidas de divergência.

Identificação dos casos	Medidas de divergência			
	d_{K-L}	d_J	d_{L_1}	d_{χ^2}
100	0,292	0,651	0,307	1,339

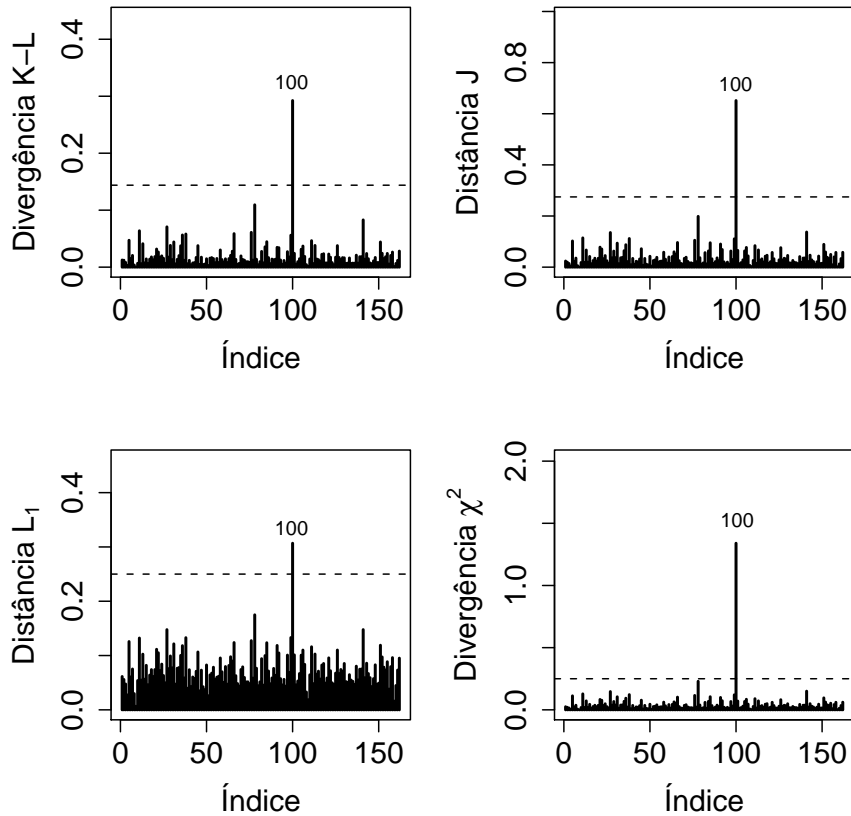


Figura 2.3: Dados de retinopatia diabética. Gráficos de índices das medidas de divergência.

tempos. Com base em todos os critérios bayesianos, há evidências positivas em favor da modelagem FGM, indicando que o modelo de sobrevivência bivariado baseado na cópula FGM pode ser visto como um concorrente para os conhecidos modelos de sobrevivência bivariados induzido pelas cópulas PS, Frank e Clayton comumente utilizados na literatura para ajuste dos dados de tempo de vida bivariado.

A Figura 2.4 mostra os gráficos de contorno da função de sobrevivência para o modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull.

Tabela 2.8: Dados de retinopatia diabética. Critérios bayesianos.

	Modelo	Critérios bayesianos			
		EAIC	EBIC	DIC	LPML
Weibull	FGM	1532,521	1554,134	1525	-762,365
	PS	1533,459	1555,072	1526	-763,078
	Frank	1532,527	1554,14	1525	-762,514
	Clayton	1532,57	1554,183	1525	-762,592
	Independência	1536,889	1555,414	1531	-765,294
Exponencial	FGM	1534,743	1550,181	1530	-765,3462
	PS	1537,341	1552,779	1532	-766,6769
	Frank	1535,463	1550,901	1530	-765,7647
	Clayton	1535,46	1550,898	1530	-765,5114
	Independência	1540,179	1552,529	1536	-768,5528

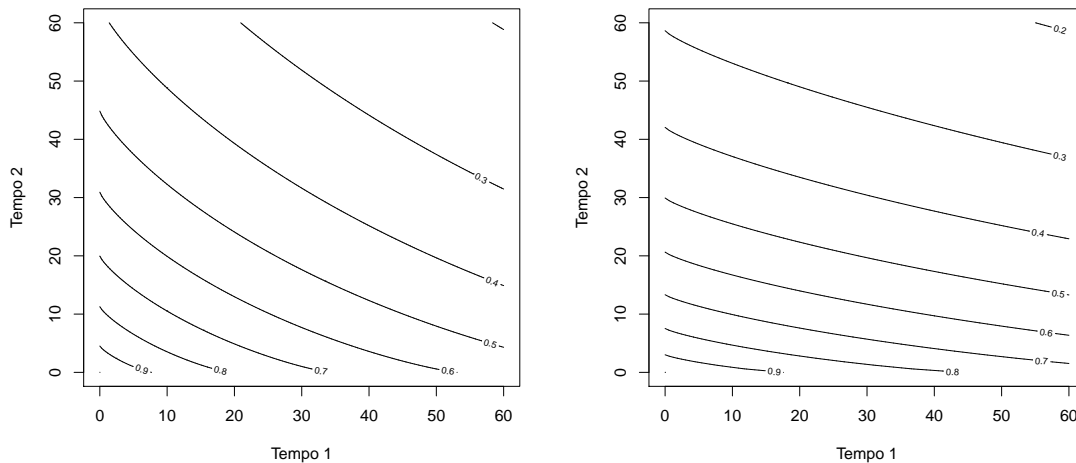


Figura 2.4: Dados de retinopatia diabética. Gráficos de contorno para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência FGM com marginais Weibull.

Capítulo 3

Modelo de sobrevivência bivariado mistura padrão baseado na cópula FGM

Os modelos de longa duração (Maller & Zhou, 1996; Rodrigues *et al.*, 2009), também conhecidos como modelos com fração de cura, têm sido amplamente desenvolvidos nos últimos tempos, sendo utilizados em várias aplicações a dados de estudos clínicos (Yakovlev & Tsodikov, 1996), como por exemplo dados de câncer em que uma proporção dos pacientes são “curados”.

Na análise de sobrevivência tradicional (Colosimo & Giolo, 2006; Kalbfleisch & Prentice, 2002; Lawless, 2003; Louzada-Neto *et al.*, 2002), os modelos paramétricos assumem que a fração de cura é zero. Os indivíduos que não apresentam o evento de interesse até certo momento são considerados censurados. Tais modelos não levam em consideração se os indivíduos deixaram de ser suscetíveis a tal evento. Neste caso, a censura pode significar tanto os indivíduos que por algum motivo abandonaram o estudo quanto os que deixaram de ser suscetíveis.

A presença de uma grande quantidade de observações com tempos mais à direita (os maiores) censurados, um conjunto de dados pode ser uma evidência de que exista uma fração de curados na população em estudo (Goldman, 1984). Caso o gráfico do estimador de Kaplan-Meier da função de sobrevivência apresente cauda longa

em um nível visivelmente acima de zero, as observações censuradas determinam a presença de uma fração de curados.

Os primeiros modelos com fração de cura foram desenvolvidos por Boag (1949) e Berkson & Gage (1952). Propuseram um modelo de mistura em que há uma proporção de indivíduos curados na população em estudo. Este modelo, também conhecido como modelo de mistura padrão, supõe a existência de apenas uma possível causa interferindo para a ocorrência do evento e , que esta causa possa ser manifestada ou não segundo uma probabilidade a ser estimada. Posteriormente, trabalhos como Yakovlev & Tsodikov (1996), Chen *et al.* (1999), Ibrahim *et al.* (2001) e Yin & Ibrahim (2005a, b) propõem modelos que permitem que não apenas uma, mas várias causas estejam relacionadas à ocorrência do evento de interesse.

Neste capítulo apresentamos a função de sobrevivência de longa duração apresentado por Rodrigues *et al.* (2009) (Seção 3.1). Propomos um modelo de sobrevivência bivariado baseado na cópula FGM com marginais mistura padrão (Berkson & Gage, 1952) em duas situações: a primeira em que consideramos a inclusão de covariáveis no parâmetro de escala do modelo Weibull (Subseção 3.2.2) e a segunda situação na proporção de curados (Subseção 3.2.3).

3.1 Função de sobrevivência de longa duração

Definição 4. (Feller, 1968) Seja $\{a_m\}$, $m = 0, 1, 2, \dots$, uma sequência de números reais. Se

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

converge, para $s \in [0, 1]$, então $A_a(s)$ é definida como função geradora da sequência $\{a_m\}$.

Considere M uma variável aleatória representando o número de causas ou riscos da ocorrência de um particular evento de interesse com distribuição de probabilidade

$$p_m = P[M = m], \quad m = 0, 1, 2, \dots \quad (3.1)$$

Dado $M = m$, sejam Z_j , $j = 1, \dots, m$, variáveis aleatórias contínuas (não-negativas) independentes e identicamente distribuídas (i.i.d.) representando o tempo de ocorrência

do evento de interesse devido à j -ésima causa ou risco. A função de sobrevivência de um indivíduo da população em risco associado à j -ésima causa é dada por

$$S(t) = P(Z_j > t), \quad j = 1, \dots, m, \quad t \geq 0, \quad (3.2)$$

em que $S(t)$ é uma função de sobrevivência própria, ou seja, $S(t) \rightarrow 0$ para $t \rightarrow \infty$. O tempo de ocorrência do evento de interesse é definido como

$$T = \min\{Z_0, Z_1, \dots, Z_M\}, \quad (3.3)$$

em que $P[Z_0 = \infty] = 1$. Esta suposição admite a possibilidade de uma proporção p_0 da população não apresentar a ocorrência do evento de interesse.

As variáveis aleatórias Z_j e M são variáveis latentes e T é uma variável aleatória observável ou censurada com função de sobrevivência dada por

$$S_{pop}(t) = P[T > t]. \quad (3.4)$$

Teorema 3. Dada uma função de sobrevivência própria $S(t)$ para os indivíduos em risco da população, temos que a função de sobrevivência de longa duração da variável aleatória T é dada por

$$S_{pop}(t) = A_p(S(t)) = \sum_{m=0}^{\infty} p_m \{S(t)\}^m, \quad (3.5)$$

sendo $A_p(\cdot)$ a função geradora da sequência $\{p_m\}$.

Podemos notar que $S_{pop}(t)$ é uma função de sobrevivência imprópria, ou seja, $\lim_{t \rightarrow \infty} S_{pop}(t) > 0$. A fração de cura da população pode ser obtida por meio do seguinte teorema.

Teorema 4. Dada uma função de sobrevivência própria $S(t)$, para os elementos da população em risco, temos que

$$\lim_{t \rightarrow \infty} S_{pop}(t) = P[M = 0] = p_0, \quad (3.6)$$

em que p_0 denota a proporção de curados (ou imunes) da população.

O modelo com fração de cura unificado é dado pelo seguinte teorema.

Teorema 5. A função de sobrevivência da v.a. T correspondente a um modelo de longa duração unificado é dada por

$$S_{pop}(t) = p_0 + (1 - p_0)S^*(t), \quad (3.7)$$

em que $S^*(t) = \sum_{m=1}^{\infty} p_m^* \{S(t)\}^m$ e $p_m^* = \frac{p_m}{1-p_0}$.

Podemos notar que $S^*(t)$ é uma função de sobrevivência própria associada aos indivíduos da população que estão em risco.

A função densidade imprópria associada à função de sobrevivência de longa duração (3.7) é dada por

$$f_{pop}(t) = f(t) \frac{dA_p(s)}{ds} \Big|_{s=S(t)}, \quad (3.8)$$

em que $A_p(\cdot)$ é a função geradora da sequência de probabilidades $\{p_m\}$.

Mais detalhes sobre o modelo de longa duração unificado podem ser obtidos em Tsodikov *et al.* (2003) e Rodrigues *et al.* (2009).

3.2 Modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM

Um caso particular do modelo de longa duração unificado é o modelo de mistura padrão (Berkson & Gage, 1952). Assume que o número de causas M segue uma distribuição de Bernoulli com parâmetro $1 - p_0$ e é dado por

$$S_{pop}(t) = p_0 + (1 - p_0)S(t), \quad (3.9)$$

em que $S(t) = P(T > t | M = 1)$ é uma função de sobrevivência própria. Para $S(t)$ sendo uma função de sobrevivência de uma distribuição Weibull, temos o modelo denominado de mistura padrão Weibull.

Então, considere (T_1, T_2) os tempos de falha pareados e S_{pop_j} e f_{pop_j} , respectivamente, a função de sobrevivência de longa duração e a função densidade de longa duração de T_j , $j = 1, 2$.

Considere que (T_1, T_2) provém da cópula C_ϕ para algum ϕ . Então, as funções de sobrevivência e densidade conjuntas de (T_1, T_2) são dadas por

$$S_{pop}(t_1, t_2) = C_\phi(S_{pop1}(t_1), S_{pop2}(t_2)), \quad t_1, t_2 > 0 \quad (3.10)$$

e

$$f_{pop}(t_1, t_2) = c_\phi(S_{pop1}(t_1), S_{pop2}(t_2))f_{pop1}(t_1)f_{pop2}(t_2), \quad t_1, t_2 > 0, \quad (3.11)$$

respectivamente.

Vamos assumir o modelo de mistura padrão para T_j , $j = 1, 2$. Dessa forma, a função de sobrevivência bivariada de (T_1, T_2) induzida pela cópula FGM (expressão 1.9) é dada por

$$S_{pop}(t_1, t_2) = S_{pop1}(t_1)S_{pop2}(t_2)[1 + \phi\{1 - S_{pop1}(t_1)\}\{1 - S_{pop2}(t_2)\}], \quad (3.12)$$

em que ϕ é o parâmetro que mede a intensidade da dependência entre os tempos de vida.

3.2.1 Inferência

Com as mesmas considerações da Seção 2.2, sejam $S_{pop}(t_1|\boldsymbol{\gamma}_1)$ e $S_{pop}(t_2|\boldsymbol{\gamma}_1)$ as funções de sobrevivência de longa duração de T_{i1} e T_{i2} , respectivamente, em que $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$ são vetores de parâmetros de q_1 e q_2 elementos associados a cada uma das distribuições marginais.

Considerando a função de sobrevivência de longa duração $S_{pop}(t_1, t_2|\phi, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ dada em (3.12), a contribuição do i -ésimo indivíduo para a log-verossimilhança de $\boldsymbol{\theta} = (\phi, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ é dada por

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) = & \delta_{i1}\delta_{i2} \log \left(\frac{\partial^2 S_{pop}(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i1} \partial t_{i2}} \right) + \delta_{i1}(1 - \delta_{i2}) \log \left(\frac{-\partial S_{pop}(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i1}} \right) \\ & + \delta_{i2}(1 - \delta_{i1}) \log \left(\frac{-\partial S_{pop}(t_1, t_2|\boldsymbol{\theta})}{\partial t_{i2}} \right) + (1 - \delta_{i1})(1 - \delta_{i2}) \log S_{pop}(t_1, t_2|\boldsymbol{\theta}). \end{aligned} \quad (3.13)$$

Consideramos as distribuições *a priori* de forma análoga à Subseção 2.2.1.

As covariáveis foram utilizadas no modelo de duas formas que serão apresentadas nas Subseções 3.2.2 e 3.2.3.

3.2.2 Implementação Computacional - Covariáveis utilizadas no parâmetro de escala da distribuição Weibull

Nesta primeira implementação, assumimos que T_j tem uma distribuição mistura padrão Weibull com parâmetros p_j , α_j and $\lambda_{ij} = \exp(\beta_{0j} + \beta_{1j}x_i)$, $i = 1, \dots, n$ e $j = 1, 2$.

Escolhemos as seguintes distribuições *a priori* independentes: $\beta_{kj} \sim N(\mu_{kj}; \sigma_{kj}^2)$, $\alpha_j \sim Gama(a_j; b_j)$, $p_{0j} \sim Beta(e_j; f_j)$ e $(1 - \phi)/2 \sim Beta(1; 1)$, em que $k = 0, 1$ e $j = 1, 2$.

A distribuição condicional para cada parâmetro é dada por

$$\pi(\phi | \mathcal{D}, \theta_{(-\phi)}) \propto (1 - \phi)^{r_1 - 1} (1 + \phi)^{r_2 - 1} \left(\prod_{i=1}^n \Delta_i \right),$$

$$\pi(p_j | \mathcal{D}, \theta_{(-p_j)}) \propto \left(\prod_{i=1}^n \Delta_i S_{popj}^{(1-\delta_{ji})} \right) p_j^{(e_j-1)} (1 - p_j)^{(\sum_{i=1}^n \delta_{ji} + f_j - 1)},$$

$$\pi(\beta_{0j} | \mathcal{D}, \theta_{(-\beta_{0j})}) \propto \left(\prod_{i=1}^n \Delta_i S_{popj}^{(1-\delta_{ji})} \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} + \frac{\mu_{0j}}{\sigma_{0j}^2} \right\} \beta_{0j} - \frac{\beta_{0j}^2}{2\sigma_{0j}^2} - \sum_{i=1}^n e^{(\beta_{0j} + \beta_{1j}x_i)} t_{ji}^{\alpha_j} \delta_{ji} \right),$$

$$\pi(\beta_{1j} | \mathcal{D}, \theta_{(-\beta_{1j})}) \propto \left(\prod_{i=1}^n \Delta_i S_{popj}^{(1-\delta_{ji})} \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} x_i + \frac{\mu_{1j}}{\sigma_{1j}^2} \right\} \beta_{1j} - \frac{\beta_{1j}^2}{2\sigma_{1j}^2} - \sum_{i=1}^n e^{(\beta_{0j} + \beta_{1j}x_i)} t_{ji}^{\alpha_j} \delta_{ji} \right)$$

e

$$\pi(\alpha_j | \mathcal{D}, \theta_{(-\alpha_j)}) \propto \left(\prod_{i=1}^n \Delta_i t_{ji}^{(\alpha_j-1)\delta_{ji}} S_{popj}^{(1-\delta_{ji})} \right) \alpha_j^{(a_j-1)(\sum_{i=1}^n \delta_{ji})} \exp \left(- \sum_{i=1}^n (\beta_{0j} + \beta_{1j}x_i) t_{ji}^{\alpha_j} \delta_{ji} - b_j \alpha_j \right),$$

para $j = 1, 2$ com $\mathcal{D} = (\mathbf{t}_1, \mathbf{t}_2, \delta_1, \delta_2, \mathbf{x})$,

$$\Delta_i = \nu_{1i}^{(1-\delta_1)(1-\delta_2)} \nu_{2i}^{\delta_1 \delta_2} \nu_{3i}^{\delta_1(1-\delta_2)} \nu_{4i}^{\delta_2(1-\delta_1)},$$

$$\nu_{1i} = 1 + \phi \prod_{j=1}^2 (1 - S_{popj}(t_{ji})),$$

$$\nu_{2i} = 1 + \phi \left(1 + \prod_{j=1}^2 (1 - 2S_{popj}(t_{ji})) \right),$$

$$\nu_{3i} = 1 + \phi \left(1 - 2S_{pop1}(t_{ji}) - S_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right)$$

$$e$$

$$\nu_{4i} = 1 + \phi \left(1 - s_{pop1}(t_{ji}) - 2s_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right).$$

As densidades condicionais acima não pertencem a nenhuma família de densidades paramétricas conhecidas. A fim de gerar as amostras temos então que implementar um passo de Metropolis nas iterações Gibbs (Chib & Greenberg, 1995). Para o parâmetro ϕ , a distribuição condicional $\pi(\phi|\mathcal{D}, \theta_{(-\phi)})$ é uma função log-côncava. Assim, o método de rejeição adaptativa pode ser usada.

3.2.2.1 Aplicação a dados artificiais

Nesta seção vamos considerar uma amostra artificial assumindo que a função sobrevivência própria do modelo (3.9) para T_j , $j = 1, 2$, tem uma distribuição Weibull com parâmetros α_j e $\lambda_{ij} = \exp(\beta_{0j} + \beta_{1j}x_i)$, $j = 1, 2$, em que as covariáveis x_i foram geradas de uma distribuição Bernoulli com parâmetro 0,5. Os dados bivariados artificiais (T_{i1}, T_{i2}) , $i = 1, \dots, n$, foram gerados assumindo $n = 150$ de acordo com os seguintes passos: Primeiro geramos $T_{i1} = (-\log(1 - u_{i1})/\lambda_{i1})^{1/\alpha_1}$ em que $u_{i1} \sim U(0; 1)$. Depois T_{i2} foi gerado usando a variável aleatória $u_{i2} \sim U(0; 1)$ e a solução da equação não linear, $w_i + \phi(2u_{i1} - 1)(w_i^2 - w_i) - u_{i2} = 0$, considerando $T_{i2} = (-\log(1 - w_i)/\lambda_{i2})^{1/\alpha_2}$.

No passo seguinte fixamos os valores de p_j , representando a fração de cura, para $j = 1, 2$. Um ponto de corte t_j^* , foi escolhido de forma que os tempos mais elevados representam a quantidade $p_j\%$ de curados e, foram assumidos censurados.

Em nosso estudo consideramos os seguintes valores para os parâmetros do modelo: $p_1 = 0,2$, $\beta_{01} = -1$, $\beta_{11} = 1$, $\alpha_1 = 2$, $p_2 = 0,1$, $\beta_{02} = 1$, $\beta_{12} = 0,5$, $\alpha_2 = 1,5$ e $\phi = 0,6$.

As seguintes distribuições *a priori* independentes foram consideradas para o amostrador de Gibbs: $p_j \sim Beta(1; 1)$, $\beta_{ij} \sim N(0; 1000)$ e $\alpha_j \sim Gama(1; 0,001)$, $i = 0, 1$ e $j = 1, 2$. Assumimos para o parâmetro da cópula PS que $\phi \sim Beta(1; 1)$ e para a cópula FGM que $1 - 2\pi(\phi) \sim Beta(1; 1)$, tal escolha garante que $\phi \in (-1, 1)$ e a não informatividade.

Simulamos duas cadeias de tamanho 50.000 para cada parâmetro, desconsiderando as primeiras 10.000 iterações para eliminar o efeito dos valores iniciais e, para evitar problemas de autocorrelação, consideramos um espaçamento de tamanho 20, obtendo uma amostra efetiva de tamanho 4.000 sobre a qual a inferência *a posteriori* é baseada.

A Tabela 3.1 apresenta o resumo para os parâmetros do modelo de sobrevivência de longa duração bivariado baseado na cópula FGM. Os verdadeiros valores são dados entre parênteses.

Tabela 3.1: Dados simulados. Média *a posteriori* (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.

	Parâmetro	Média	DP	HPD (95%)
Tempo 1	α_1 (2)	1,880	0,194	(1,532; 2,277)
	β_{01} (-1)	-0,550	0,265	(-1,241; -0,230)
	β_{11} (1)	0,955	0,283	(0,400; 1,483)
	p_1 (0,2)	0,116	0,054	(0,014; 0,215)
Tempo 2	α_2 (1,5)	1,587	0,130	(1,332; 1,843)
	β_{02} (1)	1,270	0,223	(0,797; 1,675)
	β_{12} (0,5)	0,279	0,209	(-0,122; 0,686)
	p_2 (0,1)	0,077	0,029	(0,017; 0,133)
Cópula	ϕ (0,6)	0,490	0,231	(0,056; 0,927)

A Tabela 3.2 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar os modelos de sobrevivência bivariados baseados na cópula FGM e PS com marginais mistura padrão Weibull. Podemos observar que o modelo baseado na cópula FGM supera o PS em todos os critérios considerados.

3.2.2.2 Diagnóstico de observações influentes

Para examinar o desempenho das medidas de diagnósticos propostas neste trabalho, utilizamos as mesmas amostras previamente simuladas. Selecionamos os casos 6, 98

Tabela 3.2: Dados simulados. Critérios bayesianos.

Modelo	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
FGM	317,502	344,597	308,000	-154,761
PS	320,811	347,907	310,700	-156,098

e 127 para perturbação. Para criar observações artificialmente influentes no conjunto de dados, escolhemos um, dois ou três desses casos selecionados. Para cada caso, perturbamos um ou ambos os tempos de vida da seguinte forma $\tilde{t}_i = t_i + 4S_t, i = 1, 2$, em que S_t é o desvio padrão dos t_i 's. Para o caso 6 foi perturbado apenas o tempo de vida t_1 , para o caso 98 o tempo de vida t_2 e, para o caso 127 ambos os tempos de vidas foram perturbados.

A Tabela 3.3 mostra que as inferências *a posteriori* são sensíveis à perturbação do(s) caso(s) selecionado(s). Na Tabela 3.3, o conjunto de dados (a) denota os dados originais simulados sem perturbação, e os conjuntos de dados (b) a (g) denotam os conjuntos de dados com casos perturbados.

A Tabela 3.4 mostra os critérios bayesianos do ajuste de diferentes casos de conjunto de dados perturbados. Podemos observar que os dados originais simulados (conjunto de dados (a)) tiveram o melhor ajuste.

Vamos considerar a amostra da distribuição *a posteriori* dos parâmetros do modelo de sobrevivência FGM bivariado para obter uma estimativa das quatro medidas de divergência. Os resultados na Tabela 3.5 mostram, antes da perturbação (conjunto de dados (a)), que todos os casos selecionados não são influentes, com pequenas medidas de divergência. Entretanto, após perturbações (conjunto de dados (b) a (g)) as quatro medidas aumentam, indicando que os casos são influentes. Os valores que ultrapassam os pontos de cortes apresentados na Subseção 1.3.1 estão destacados com *.

A Figura 3.1 mostra os gráficos de índices das quatro medidas de divergência para o conjunto de dados (e). Claramente, podemos ver que as quatro medidas de divergência detectam o ponto influente.

Tabela 3.3: Dados simulados. Média *a posteriori* e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Nome dos dados	Identificação dos casos perturbados	β_{01} Média [DP]	β_{11} Média [DP]	r_1 Média [DP]	p_1 Média [DP]	β_{02} Média [DP]	β_{12} Média [DP]	r_2 Média [DP]	p_2 Média [DP]	ϕ Média [DP]
a	Amostra original	-0,755 [0,26]5	0,955 [0,283]	1,880 [0,194]	0,116 [0,054]	1,270 [0,223]	0,279 [0,209]	1,587 [0,130]	0,077 [0,029]	0,490 [0,231]
b	6	-1,011 [0,192]	1,185 [0,223]	1,704 [0,142]	0,066 [0,038]	1,239 [0,221]	0,315 [0,206]	1,583 [0,132]	0,077 [0,029]	0,468 [0,233]
c	98	-0,749 [0,267]	0,943 [0,286]	1,881 [0,187]	0,117 [0,052]	1,048 [0,218]	0,216 [0,217]	1,407 [0,111]	0,056 [0,031]	0,514 [0,240]
d	127	-0,898 [0,190]	0,888 [0,220]	1,524 [0,115]	0,050 [0,038]	1,027 [0,215]	0,237 [0,211]	1,413 [0,109]	0,057 [0,031]	0,569 [0,235]
e	{6,98}	-1,043 [0,184]	1,197 [0,219]	1,663 [0,136]	0,057 [0,036]	1,005 [0,219]	0,269 [0,212]	1,403 [0,111]	0,055 [0,031]	0,498 [0,244]
f	{6,127}	-0,994 [0,170]	0,976 [0,205]	1,477 [0,108]	0,035 [0,029]	1,018 [0,220]	0,252 [0,218]	1,407 [0,110]	0,056 [0,031]	0,562 [0,234]
g	{6,98,127}	-0,993 [0,168]	0,978 [0,194]	1,477 [0,107]	0,034 [0,028]	0,904 [0,209]	0,202 [0,216]	1,318 [0,100]	0,043 [0,029]	0,584 [0,242]

Tabela 3.4: Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Conjunto de dados	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
a	317,502	344,598	308,000	-154,761
b	323,052	350,148	313,500	-157,342
c	333,519	360,615	323,400	-163,714
d	351,246	378,342	340,700	-172,834
e	341,084	368,180	331,100	-169,199
f	356,127	383,223	345,800	-175,109
g	367,601	394,697	356,900	-180,600

Tabela 3.5: Dados simulados. Medidas de Divergência.

Nome dos dados	Identificação dos casos	Medidas de divergência			
		d_{K-L}	d_J	d_{L_1}	d_{χ^2}
a	6	0,009	0,017	0,053	0,018
	98	0,007	0,014	0,047	0,014
	127	0,007	0,013	0,046	0,013
b	6	0,675*	1,928*	0,477*	24,970*
c	98	1,340*	3,204*	0,625*	42,944*
d	127	2,793*	5,821*	0,786*	110,601*
e	6	2,602*	9,357*	0,879*	2850,702*
	98	1,302*	2,992*	0,616*	21,363*
f	6	0,225*	0,573*	0,265*	3,988*
	127	2,518*	5,270*	0,768*	73,330*
g	6	0,182*	0,403*	0,241*	0,721*
	98	0,613*	1,457*	0,441*	7,645*
	27	0,010*	0,021*	0,057*	0,021*

3.2.2.3 Dados reais

Nesta seção ilustramos a metodologia proposta com a mesma amostra de dados de retinopatia diabética utilizada na Subseção 2.3.3.

Na Tabela 3.6 apresentamos os resumos *a posteriori* para os parâmetros do modelo de sobrevivência bivariado mistura padrão baseado na Cópula FGM. Além de avaliar a robustez do modelo relacionado às escolhas dos hiperparâmetros das distribuições *a priori*, foi realizado um estudo de sensibilidade no qual constatamos que as estimativas dos parâmetros *a posteriori* não apresentaram diferenças significativas e não alteraram os resultados da Tabela 3.6. Consideramos distribuições marginais com mistura padrão Weibull e Exponencial.

Na Tabela 3.7 apresentamos o valor das quatro medidas de divergência para o caso 100 ajustando o modelo de sobrevivência bivariado mistura padrão Exponencial baseado na cópula FGM. Claramente observamos que todas as medidas detectam

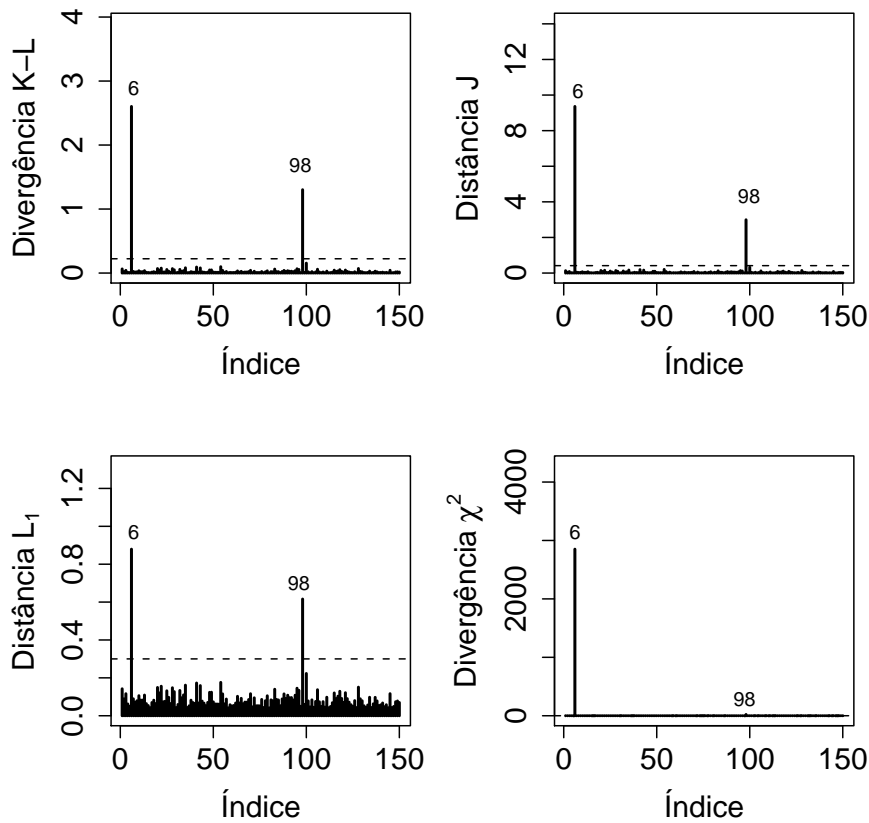


Figura 3.1: Gráficos de índices das medidas de divergência para o caso (e).

o ponto influente, o que também pode ser observado no gráfico de índices (Figura 3.2).

A Figura 3.3 mostra o gráfico de dispersão entre T_1 e T_2 (painel acima) e as curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), dicotomizadas pela idade do paciente juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Exponencial baseado na cópula FGM.

A Tabela 3.8 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar os modelos de sobrevivências bivariados mistura padrão Weibull e Exponencial baseados na cópula FGM com os modelos baseados nas cópulas de Clayton, Frank e Positive Stable (PS), como também com o caso assumindo independência entre os tempos.

Tabela 3.6: Dados de retinopatia diabética. Média *a posteriori* e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão baseado na cópula FGM.

	Parâmetro	Mistura padrão Weibull			Mistura padrão Exponencial		
		Média	DP	HPD (95%)	Média	DP	HPD (95%)
Tempo 1	α_1	1,107	0,170	(0,784; 1,454)	-	-	-
	β_{01}	-3,630	0,486	(-4,605; -2,742)	-3,371	0,353	(-4,212; -2,812)
	β_{11}	-0,937	0,428	(-1,765; -0,103)	-0,922	0,412	(-1,712; -0,123)
	p_{01}	0,537	0,110	(0,196; 0,681)	0,517	0,101	(0,250; 0,659)
Tempo 2	α_2	0,846	0,076	(0,704; 0,998)	-	-	-
	β_{02}	-3,414	0,311	(-4,023; -2,794)	-3,879	0,246	(-4,304; -3,340)
	β_{12}	0,297	0,227	(-0,160; 0,726)	0,254	0,245	(-0,233; 0,743)
	p_{02}	0,093	0,071	(0,004; 0,265)	0,133	0,081	(0,007; 0,302)
Cópula	ϕ	0,644	0,227	(0,135; 0,976)	0,639	0,227	(0,133; 0,975)

Tabela 3.7: Dados de retinopatia diabética. Medidas de diagnósticos para o caso número 100.

Identificação dos casos	Medidas de divergência			
	d_{K-L}	d_J	d_{L_1}	d_{χ^2}
100	0,919	2,088	0,535	9,539

Com base em todos os critérios bayesianos, há evidências positivas em favor da modelagem FGM, indicando que o modelo de sobrevivência mistura padrão bivariado baseado na cópula FGM pode ser visto como um concorrente para os conhecidos modelos de sobrevivência bivariados induzido pelas cópulas PS, Frank e Clayton comumente utilizados na literatura para ajuste dos dados de tempo de vida bivariados. Para os dados de retinopatia, obtivemos melhor ajuste assumindo marginais mistura padrão Exponencial.

A Figura 3.4 mostra os gráficos de contorno da função sobrevivência para o modelo de sobrevivência bivariado mistura padrão Exponencial baseado na cópula FGM.

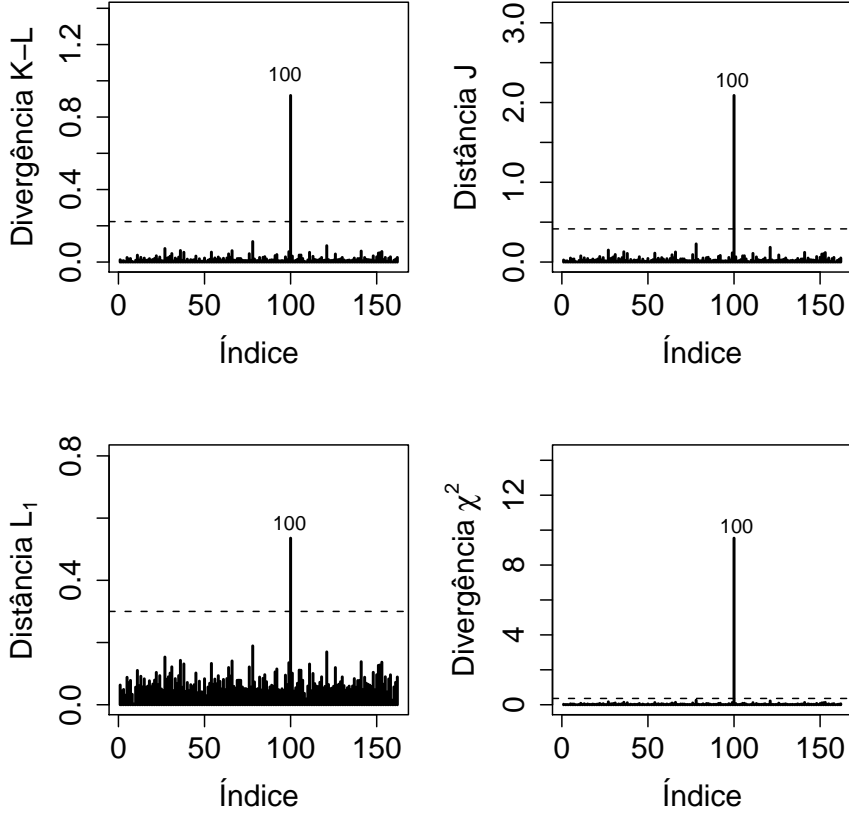


Figura 3.2: Gráficos de índices das medidas de divergência.

3.2.3 Implementação Computacional - Covariáveis utilizadas no parâmetro de fração de cura

Nesta segunda implementação, assumimos que T_j tem uma distribuição mistura padrão Weibull com parâmetros α_j e λ_j , e $p_j = \exp(\beta_{0j} + \beta_{1j}\mathbf{x}) / (1 + \exp(\beta_{0j} + \beta_{1j}\mathbf{x}))$, $j = 1, 2$.

A distribuição condicional para cada parâmetro é dada por

$$\pi(\phi | \mathcal{D}, \theta_{(-\phi)}) \propto (1 - \phi)^{r_1 - 1} (1 + \phi)^{r_2 - 1} \left(\prod_{i=1}^n \Delta_i \right),$$

$$\pi(\beta_{0j} | \mathcal{D}, \theta_{(-\beta_{0j})}) \propto \left(\prod_{i=1}^n f_{pop_j}^{\delta_i} (1 + e^{\beta_{0j} + \beta_{1j}x_i})^{-(1 - \delta_{ji})} \Delta_i \right) e^{\frac{-1}{2\sigma_{0j}^2}(\beta_{0j}^2 - 2\mu_{0j}\beta_{0j})},$$

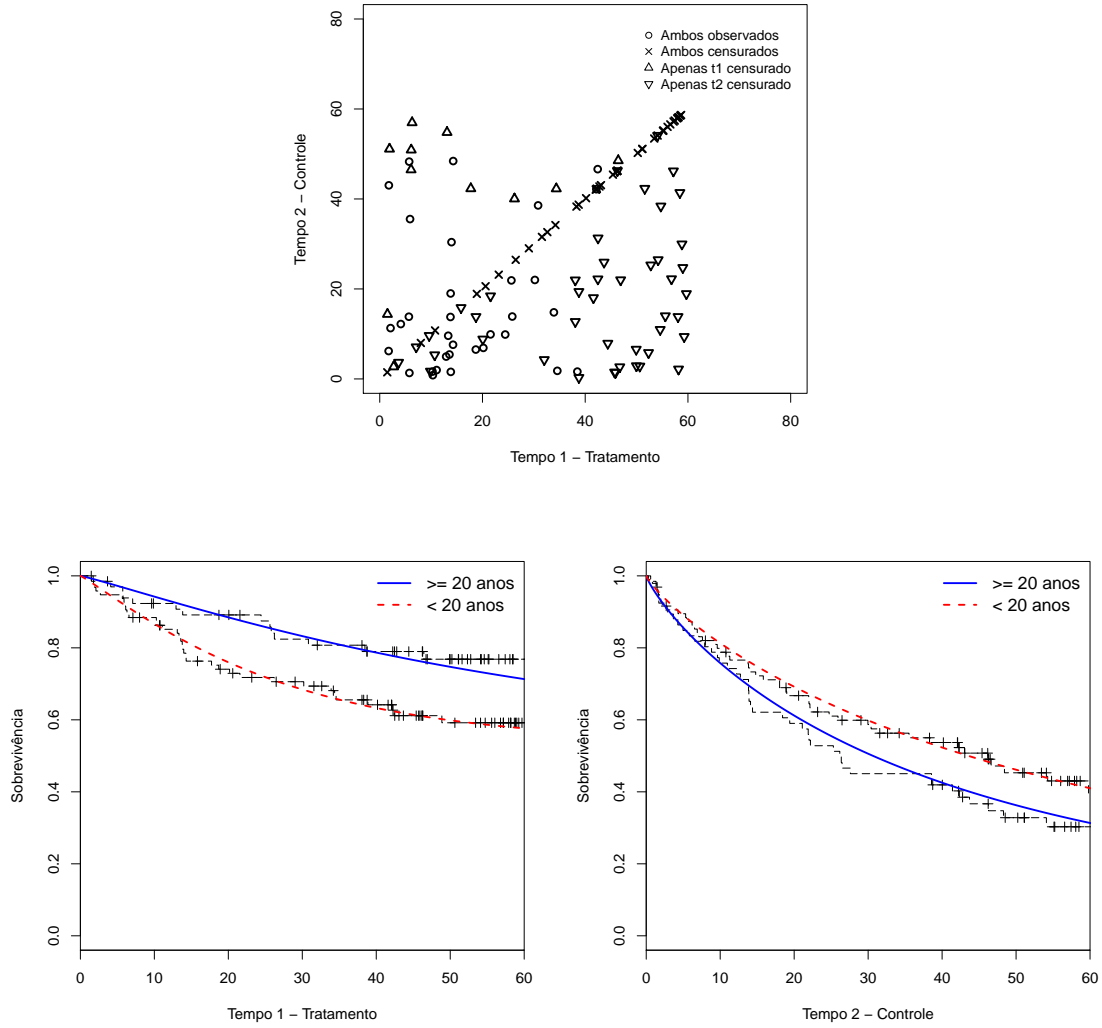


Figura 3.3: Dados de retinopatia diabética. Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Exponencial baseado na cópula FGM.

$$\pi(\beta_{1j} | \mathcal{D}, \theta_{(-\beta_{1j})}) \propto \left(\prod_{i=1}^n f_{pop_j}^{\delta_{ji}} (1 + e^{\beta_{0j} + \beta_{1j} x_i})^{-(1-\delta_{ji})} \Delta_i \right) e^{\frac{-1}{2\sigma_{1j}^2} (\beta_{1j}^2 - 2\mu_{1j} \beta_{1j})},$$

$$\pi(\lambda_j | \mathcal{D}, \theta_{(-\lambda_j)}) \propto \left(\prod_{i=1}^n f_{pop_j}^{\delta_{ji}} (1 - p_{ji})^{(1-\delta_{ji})} \Delta_i \right) \lambda_j^{n+a_j - (1 + \sum_{i=1}^n \delta_{ji})} e^{-\lambda_j \left[b_j + \sum_{i=1}^n t_{ji}^{\alpha_j} (1-\delta_{ji}) \right]}$$

e

$$\pi(\alpha_j | \mathcal{D}, \theta_{(-\alpha_j)}) \propto \left(\prod_{i=1}^n f_{pop_j}^{\delta_{ji}} (1 - p_{ji})^{(1-\delta_{ji})} \Delta_i t_{ji}^{(\alpha_j-1)(1-\delta_{ji})} \right) \alpha_j^{c_j-1} e^{-\left[d_j \alpha_j + \sum_{i=1}^n \lambda_j t_{ji}^{\alpha_j} (1-\delta_{ji}) \right]},$$

Tabela 3.8: Dados de retinopatia diabética. Critérios bayesianos.

	Modelo	Critérios bayesianos			
		EAIC	EBIC	DIC	LPML
Weibull	FGM	1532,876	1560,664	1522,000	-761,810
	PS	1534,943	1556,556	1527,000	-764,888
	Frank	1532,999	1560,787	1522,000	-762,043
	Clayton	1533,405	1561,194	1523,000	-762,467
	Independência	1536,773	1561,474	1528,000	-764,554
Exponencial	FGM	1532,589	1554,202	1525,000	-763,460
	PS	1534,813	1562,602	1524,000	-762,866
	Frank	1532,730	1554,343	1525,000	-763,689
	Clayton	1532,996	1554,609	1525,000	-763,930
	Independência	1536,260	1554,780	1530,000	-765,930

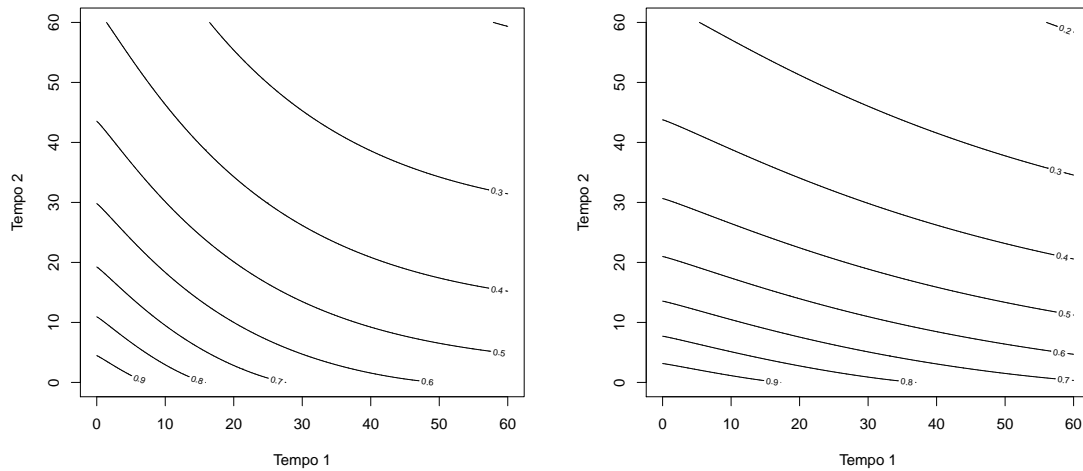


Figura 3.4: Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência mistura padrão Exponencial baseado na cópula FGM.

para $j = 1, 2$ com $\mathcal{D} = (\mathbf{t}_1, \mathbf{t}_2, \delta_1, \delta_2, \mathbf{x})$, $f_{pop_j} = -S'_{pop_j}(t_{ji})$,

$$\Delta_i = \nu_{1i}^{(1-\delta_1)(1-\delta_2)} \nu_{2i}^{\delta_1 \delta_2} \nu_{3i}^{\delta_1(1-\delta_2)} \nu_{4i}^{\delta_2(1-\delta_1)},$$

$$\nu_{1i} = 1 + \phi \prod_{j=1}^2 (1 - S_{popj}(t_{ji})),$$

$$\nu_{2i} = 1 + \phi \left(1 + \prod_{j=1}^2 (1 - 2S_{popj}(t_{ji})) \right),$$

$$\nu_{3i} = 1 + \phi \left(1 - 2S_{pop1}(t_{ji}) - S_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right)$$

e

$$\nu_{4i} = 1 + \phi \left(1 - s_{pop1}(t_{ji}) - 2s_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right).$$

3.2.3.1 Aplicação para dados artificiais

Nesta seção vamos considerar uma amostra artificial assumindo que a sobrevivência própria do modelo (3.9) para T_j , $j = 1, 2$, tem uma distribuição Weibull com parâmetros α_j e λ_j , $j = 1, 2$. Os dados bivariados artificiais (T_{i1}, T_{i2}) , $i = 1, \dots, n$, foram gerados assumindo $n = 150$. No primeiro passo, geramos os tempos T_j , $j = 1, 2$, de forma análoga à Subseção 3.2.2.1.

No passo seguinte, para cada tempo de vida t_j e para cada nível de covariável \mathbf{x} (0 ou 1), temos uma proporção de curados. Fixamos os valores p_{kj} , representando a taxa de cura e para o nível k do tempo de vida j , em que $k = 0, 1$ e $j = 1, 2$. Para cada nível de covariável de cada tempo vida gerada t_j foi escolhido um ponto de corte t_{kj}^* , de forma que os tempos mais elevados representam a quantidade $p_{kj}\%$ de curados e foram assumidos censurados.

Em nosso estudo consideramos os seguintes valores para os parâmetros do modelo: $p_{01} = 0,5$, $p_{11} = 0,1$, $p_{02} = 0,3$, $p_{12} = 0,2$, $\alpha_1 = 2$, $\alpha_2 = 1,5$, $\lambda_1 = 0,5$, $\lambda_2 = 0,1$ e $\phi = 0,6$.

A Tabela 3.9 apresenta o resumo para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM. Os verdadeiros valores

são dados entre parênteses.

Tabela 3.9: Dados simulados. Média *a posteriori* (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.

	Parâmetro	Média	DP	HPD (95%)
	α_1 (2)	2.087	0.178	(1.745; 2.457)
Tempo 1	λ_1 (0,5)	0.674	0.078	(0.528; 0.837)
	p_{10} (0,5)	0.397	0.070	(0.263; 0.531)
	p_{11} (0,1)	0.084	0.033	(0.0254; 0.151)
	α_2 (1,5)	1.716	0.163	(1.406; 2.044)
Tempo 2	λ_2 (0,1)	0.108	0.024	(0.067; 0.159)
	p_{20} (0,3)	0.249	0.061	(0.134; 0.368)
	p_{21} (0,2)	0.162	0.047	(0.068; 0.261)
Cópula	ϕ (0,6)	0.668	0.204	(0.230; 0.977)

A Tabela 3.10 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar os modelos de sobrevivência bivariado baseado na cópula FGM, PS, Frank e Clayton, sendo que em todos os modelos assumimos marginais mistura padrão Weibull. O modelo baseado na cópula FGM supera os modelos baseados nas cópulas de Clayton, Frank e PS em todos os critérios considerados.

Tabela 3.10: Dados simulados. Critérios bayesianos.

Modelo	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
FGM	939,650	967,326	930,400	-465,243
PS	947,184	974,860	937,300	-469,120
Frank	941,447	969,124	932,400	-466,336
Clayton	940,140	967,816	931,100	-465,642

3.2.3.2 Diagnóstico de observações influentes

Para examinar o desempenho das medidas de diagnósticos propostas neste trabalho, utilizamos a mesma amostra previamente simulados. Seleccionamos os casos 5, 35, 95 (ambos os tempos de vida observados) e 142 (tempo 1 é observado e tempo 2 é censurado) para perturbação. Para criar observações artificialmente influentes no conjunto de dados, escolhemos um, dois ou três desses casos seleccionados. Para cada caso, perturbamos um ou ambos os tempos de vida da seguinte forma $\tilde{t}_i = t_i + 5S_t, i = 1, 2$, em que S_t é o desvio padrão dos t_i 's. Para o caso 5 foi perturbado apenas o tempo de vida t_1 , para o caso 35 o tempo de vida t_2 e, para os casos 95 e 142, ambos os tempos de vidas foram perturbados.

A Tabela 3.11 mostra que as inferências *a posteriori* são sensíveis à perturbação do(s) caso(s) seleccionado(s). Na Tabela 3.11, o conjunto de dados (a) denota os dados originais simulados sem perturbação, e os conjuntos de dados (b) to (i) denotam os conjuntos de dados com casos perturbados.

A Tabela 3.12 mostra os critérios bayesianos do ajuste de diferentes casos de conjunto de dados perturbado. Podemos observar que os dados originais simulados (conjunto de dados (a)) teve o melhor ajuste.

Vamos considerar a amostra da distribuição *a posteriori* dos parâmetros do modelo de sobrevivência FGM bivariado para obter uma estimativa das quatro medidas de divergência. Os resultados na Tabela 3.13 mostra, antes da perturbação (conjunto de dados (a)), que todos os casos seleccionados não são influencias, com pequenas medidas de divergência. Entretanto, após perturbações (conjunto de dados (b) a (i)) as quatro medidas aumentam, indicando que os casos são influentes. Os valores que ultrapassam os pontos de cortes apresentados na Subsecção 1.3.1 estão destacados com *.

A Figura 3.5 mostra os gráficos de índices das quatro medidas de divergência para o conjunto de dados (e). Claramente, podemos ver que as quatro medidas de divergência detectam o ponto influente.

Tabela 3.11: Dados simulados. Média *a posteriori* e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Nome dos dados	Identificação dos casos perturbados	α_1 Média [DP]	λ_1 Média [DP]	β_{01} Média [DP]	β_{11} Média [DP]	α_2 Média [DP]	λ_2 Média [DP]	β_{02} Média [DP]	β_{21} Média [DP]	ϕ Média [DP]
a	Amostra original	2,087 [0,178]	0,674 [0,078]	-0,427 [0,301]	-2,056 [0,550]	1,716 [0,163]	0,108 [0,024]	-1,133 [0,347]	-0,563 [0,484]	0,668 [0,204]
b	5	1,815 [0,156]	0,606 [0,081]	-0,684 [0,397]	-2,475 [1,202]	1,713 [0,165]	0,108 [0,024]	-1,139 [0,383]	-0,546 [0,487]	0,648 [0,216]
c	35	2,084 [0,186]	0,666 [0,082]	-0,451 [0,298]	-2,049 [0,637]	1,402 [0,124]	0,131 [0,025]	-1,567 [0,579]	-1,000 [1,084]	0,664 [0,199]
d	95	1,761 [0,143]	0,589 [0,078]	-0,755 [0,413]	-2,602 [1,177]	1,361 [0,121]	0,131 [0,024]	-1,876 [0,899]	-1,108 [1,395]	0,714 [0,194]
e	{5,35}	1,827 [0,154]	0,604 [0,077]	-0,662 [0,355]	-2,287 [0,819]	1,384 [0,128]	0,130 [0,025]	-1,753 [0,719]	-1,030 [1,303]	0,624 [0,219]
f	{35,95}	1,753 [0,143]	0,586 [0,080]	-0,744 [0,399]	-2,728 [1,238]	1,257 [0,098]	0,134 [0,025]	-2,409 [1,204]	-1,611 [1,666]	0,667 [0,202]
g	{5,35,95}	1,657 [0,127]	0,545 [0,075]	-0,916 [0,496]	-3,237 [1,514]	1,258 [0,102]	0,132 [0,024]	-2,615 [1,302]	-1,549 [1,777]	0,628 [0,219]
h	142	1,710 [0,135]	0,600 [0,078]	-0,719 [0,378]	-2,622 [1,250]	1,738 [0,155]	0,105 [0,023]	-1,112 [0,316]	-0,503 [0,441]	0,690 [0,206]
i	{95,142}	1,599 [0,117]	0,540 [0,072]	-1,012 [0,521]	-3,453 [1,495]	1,408 [0,115]	0,132 [0,025]	-1,538 [0,530]	-0,566 [0,688]	0,722 [0,192]

3.2.3.3 Dados reais

Nesta Seção ilustramos a metodologia proposta para o conjunto de dados de pacientes portadores de HIV (Subseção 1.4.2). Ao analisarmos este conjunto de dados, não detectamos nenhum ponto influente. Dessa forma, para criar uma observação influente para este conjunto de dados, perturbamos o caso 61 da seguinte forma: $\tilde{t}_1 = t_1 + 28$ e $\tilde{t}_2 = t_2 + 100$.

A Figura 3.6 mostra o gráfico de dispersão entre T_1 e T_2 (painel acima) e as curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita) dicotomizados pela idade do paciente juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.

Tabela 3.12: Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Conjunto de dados	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
a	939,650	967,326	930,400	-465,243
b	955,960	983,637	946,500	-474,896
c	958,307	985,983	947,800	-476,402
d	974,706	1002,382	962,600	-485,541
e	974,323	1002,000	963,000	-485,467
f	984,981	1012,658	973,000	-489,721
g	994,623	1022,299	982,600	-494,208
h	958,374	986,051	948,200	-475,905
i	986,177	1013,854	975,700	-491,553

Na Tabela 3.14 apresentamos os resumos *a posteriori* para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull e Exponencial baseado na cópula FGM. Além de avaliar a robustez do modelo relacionado às escolhas dos hiperparâmetros das distribuições *a priori*, foi realizado um estudo de sensibilidade no qual constatamos que as estimativas dos parâmetros *a posteriori* não apresentaram diferenças significativas e não alteraram os resultados da Tabela 3.14.

Na Tabela 3.15 apresentamos o valor das quatro medidas de divergência para o caso 61. Podemos observar que todas as medidas detectam o ponto influente, que também pode ser observado no gráfico de índices (Figura 3.7).

A Tabela 3.2.3 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar modelo de sobrevivência bivariado mistura padrão Weibull e Exponencial baseado na cópula FGM com os modelos baseados nas cópulas de Clayton, Frank e Positive Stable, como também com o caso assumindo independência entre os tempos.

Com base em todos os critérios bayesianos, há evidências positivas em favor

Tabela 3.13: Dados Simulados. Medidas de diagnósticos.

Nome dos dados	Identificação dos casos	Medidas de divergência			
		d_{K-L}	d_J	d_{L_1}	d_{χ^2}
a	5	0,013	0,027	0,065	0,027
	35	0,031	0,062	0,099	0,065
	95	0,021	0,042	0,082	0,044
	142	0,030	0,061	0,098	0,066
b	5	1,629*	3,839*	0,680*	61,106*
c	35	2,324*	5,703*	0,772*	299,144*
d	95	3,759*	7,598*	0,846*	267,393*
e	5	1,514*	3,692*	0,656*	69,668*
	35	2,162*	4,794*	0,752*	66,745*
f	35	0,526*	1,260*	0,411*	5,510*
	95	2,551*	5,338*	0,785*	62,277*
g	5	0,439*	1,005*	0,378*	2,784*
	35	0,593*	1,542*	0,435*	17,835*
	95	1,799*	4,494*	0,712*	124,172*
h	142	2,006*	4,149*	0,723*	26,839*
i	95	3,465*	7,854*	0,866*	360,396*
	142	0,845*	1,883*	0,510*	9,551*

da modelagem FGM, indicando que o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM (obteve um melhor ajuste) pode ser visto como um concorrente para os conhecidos modelos de sobrevivência bivariados induzido pelas cópulas PS, Frank e Clayton, comumente utilizados na literatura para ajuste dos dados de tempo de vida bivariado.

A Figura 3.8 mostra os gráficos de contorno da sobrevivência para o modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM de acordo com o sexo do paciente.

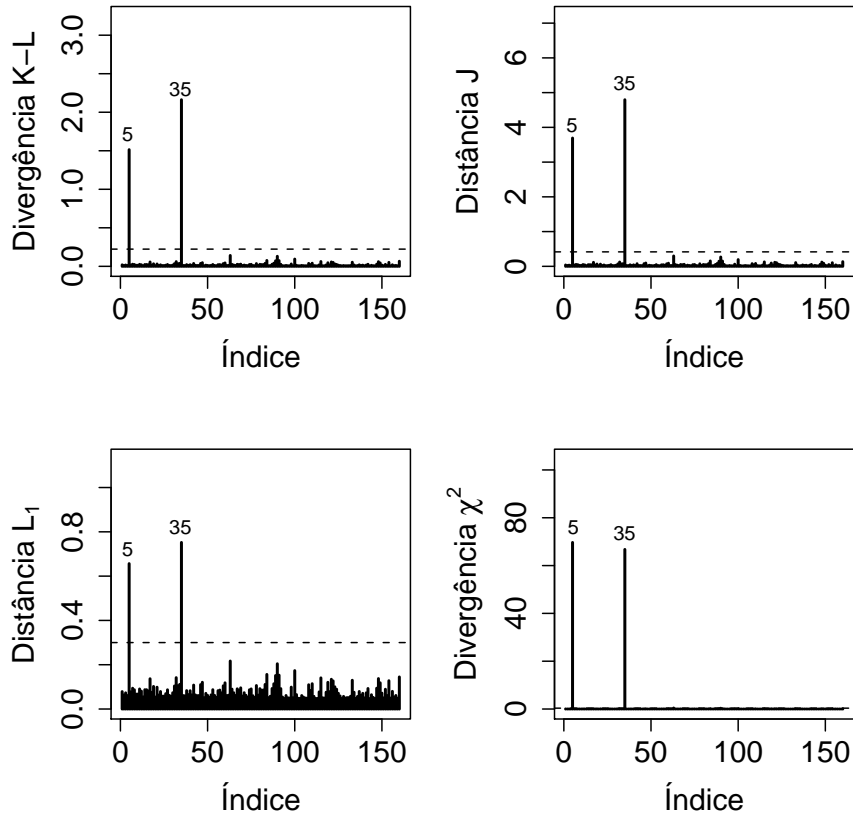


Figura 3.5: Gráficos de índices das medidas de divergência.

Tabela 3.14: Dados de HIV. Média *a posteriori* e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado mistura padrão Weibull e Exponencial baseado na cópula FGM.

	Parâmetro	Mistura padrão Weibull			Mistura padrão Exponencial		
		Média	DP	HPD (95%)	Média	DP	HPD (95%)
Tempo 1	α_1	1,332	0,127	(1,088; 1,569)	-	-	-
	λ_1	0,04	0,012	(0,020; 0,069)	0,071	0,011	(0,052; 0,095)
	β_{01}	-2,084	0,972	(-4,615; -0,697)	-3,995	1,01	(-5,660; -1,798)
	β_{11}	0,658	0,97	(-1,138; 2,999)	0,528	1,048	(-1,236; 4,229)
Tempo 2	α_2	1,457	0,14	(1,194; 1,747)	-	-	-
	λ_2	0,04	0,013	(0,019; 0,070)	0,094	0,015	(0,067; 0,125)
	β_{02}	-0,695	0,448	(-1,692; 0,081)	-1,189	0,647	(-2,694; -0,165)
	β_{12}	-0,759	0,514	(-1,717; 0,301)	-0,723	0,758	(-2,167; 0,828)
Cópula	ϕ	0,559	0,259	(0,021; 0,966)	0,599	0,264	(0,011; 0,980)

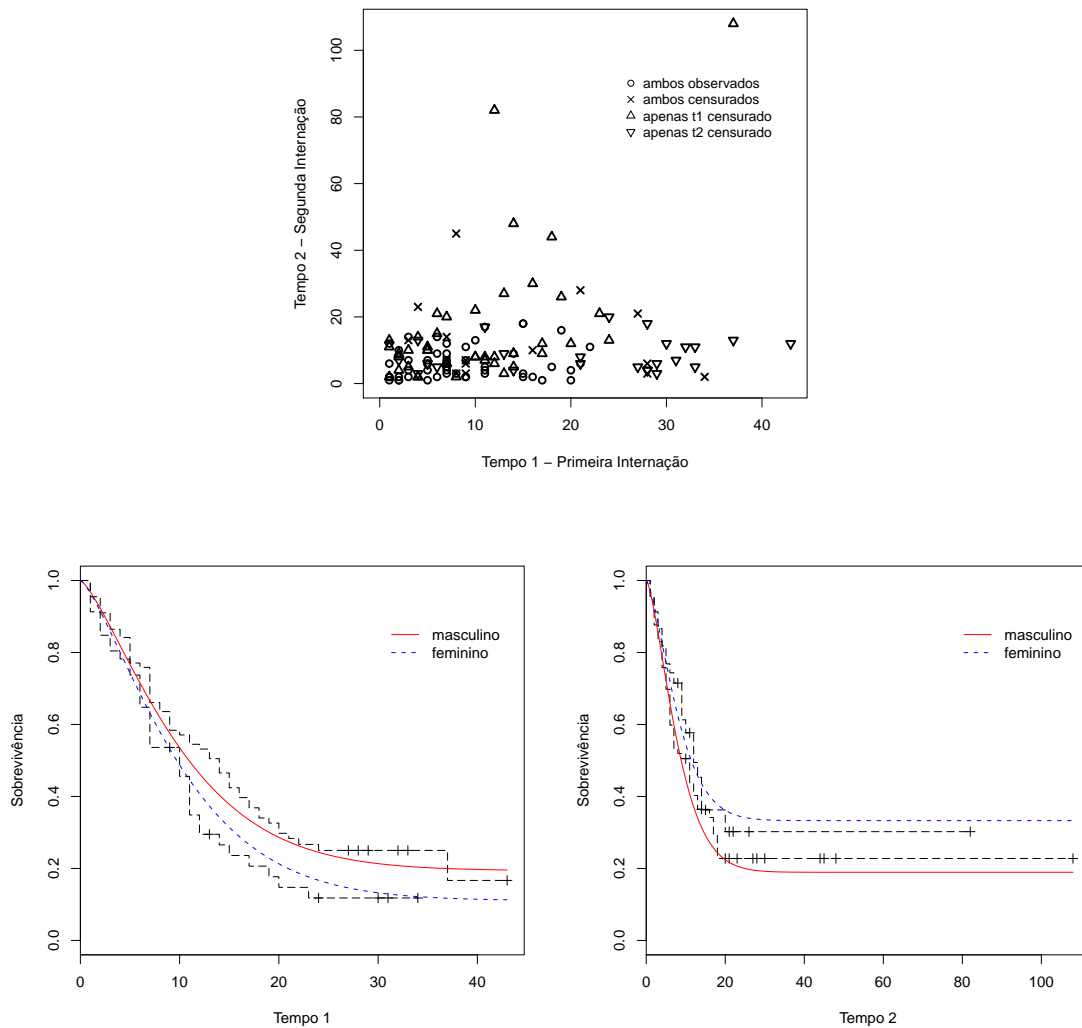


Figura 3.6: Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variável T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.

Tabela 3.15: Dados de HIV. Medidas de divergência.

Identificação dos casos	Medidas de divergência			
	d_{K-L}	d_J	d_{L_1}	d_{χ^2}
61	0,379	0,805	0,345	1,528

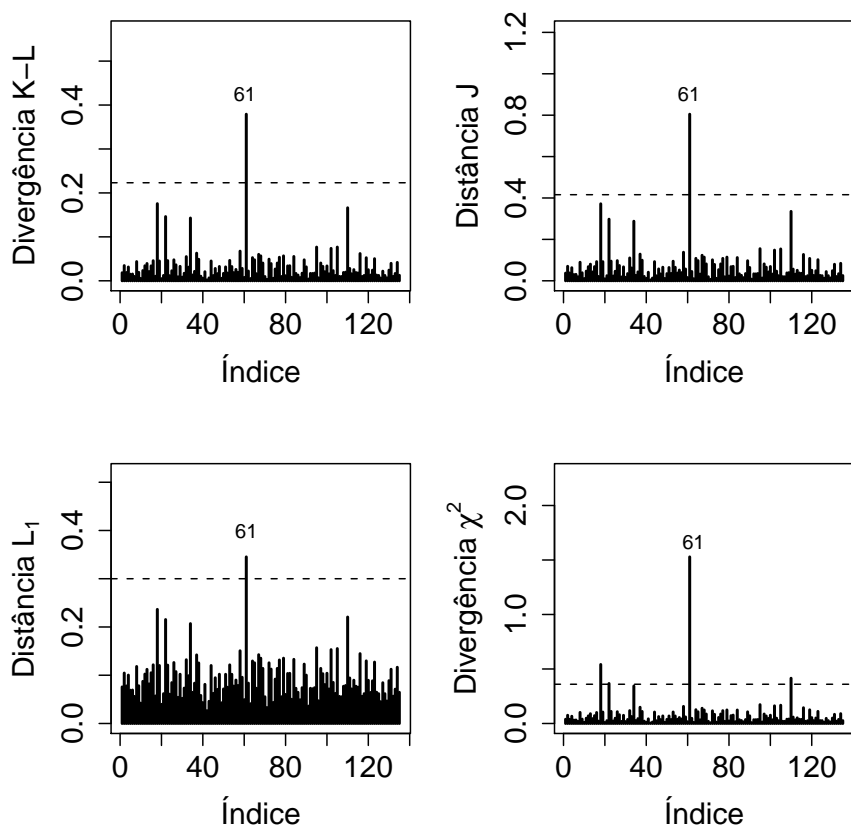


Figura 3.7: Gráficos de índices das medidas de divergência.

Tabela 3.16: Dados de HIV. Critérios bayesianos.

	Modelo	Critérios bayesianos			
		EAIC	EBIC	DIC	LPML
Weibull	FGM	1319,258	1345,406	1311	-655,054
	PS	1319,812	1345,959	1311	-655,541
	Frank	1322,804	1348,951	1316	-657,745
	Clayton	1325,07	1351,217	1317	-659,217
	Independência	1328,21	1351,45	1324	-662,22
Exponencial	FGM	1335,302	1355,639	1324	-663,661
	PS	1335,658	1355,995	1326	-664,236
	Frank	1335,854	1356,191	1324	-664,103
	Clayton	1338,163	1355,595	1329	-666,124
	Independência	1340,386	1360,723	1330	-666,679

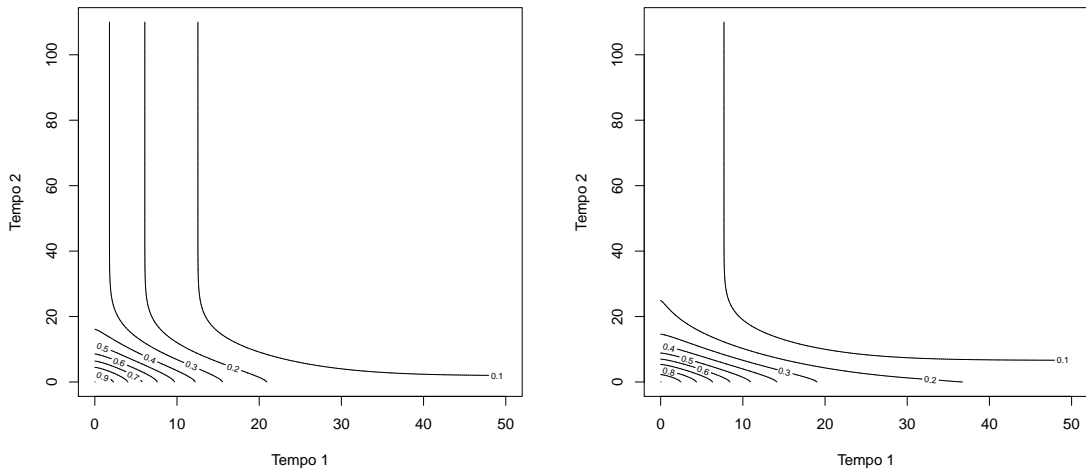


Figura 3.8: Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM.

Capítulo 4

Modelo de sobrevivência bivariado tempo de promoção baseado na cópula FGM

Conhecido na literatura como modelo de tempo de promoção (Yakovlev & Tsodikov, 1996; Chen *et al.*, 1999) assume que a variável latente M (Seção 3.1) segue uma distribuição de Poisson com parâmetro $\theta > 0$. A partir da função geradora de probabilidade (Feller, 1968) da distribuição de Poisson e do Teorema 3 temos que a função de sobrevivência de longa duração é dada por

$$S_{pop}(t) = A_p(S(t)) = \exp(-\theta[1 - S(t)]) = \exp[-\theta F(t)]. \quad (4.1)$$

O modelo (4.1) pode ser reescrito como

$$S_{pop}(t) = \exp\{-\theta\} + (1 - \exp\{-\theta\})S^*(t), \quad (4.2)$$

em que $S^*(t) = \frac{\exp\{-\theta F(t)\} - \exp\{-\theta\}}{1 - \exp\{-\theta\}}$, obtendo uma relação matemática com o modelo de mistura padrão (3.9).

A partir da equação (4.1) temos que a função densidade de longa duração é dada por

$$f_{pop}(t) = -\frac{dS_{pop}(t)}{dt} = \theta f(t) \exp\{-\theta F(t)\} \quad (4.3)$$

Do Teorema 4 obtemos que a fração de cura correspondente é dada por $p_0 = \exp(-\theta) > 0$.

Nesta Seção relacionamos o número médio de causas do evento de interesse a covariáveis por meio de $E[M_i] = \theta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$, para $i = 1, \dots, n$.

4.1 Modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM

Considere (T_1, T_2) os tempos de falhas pareados e, S_{pop_j} e f_{pop_j} , respectivamente, a função de sobrevivência de longa duração e a função densidade de longa duração de T_j , $j = 1, 2$.

Vamos assumir modelo de tempo de promoção para T_j , $j = 1, 2$. Dessa forma, a função de sobrevivência bivariado de (T_1, T_2) induzida pela cópula FGM (expressão 1.9) é dada por

$$S_{pop}(t_1, t_2) = S_{pop_1}(t_1)S_{pop_2}(t_2)(1 + \phi(1 - S_{pop_1}(t_1))(1 - S_{pop_2}(t_2))). \quad (4.4)$$

4.2 Inferência

Com as mesmas considerações da Seção 3.2.1, considerando a função de sobrevivência de longa duração $S_{pop}(t_1, t_2 | \phi, \gamma_1, \gamma_2)$ dado em (4.1), a contribuição do i -ésimo indivíduo para a log-verossimilhança de $\boldsymbol{\theta} = (\phi, \gamma_1, \gamma_2)$ é dado por

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) = & \delta_{i1}\delta_{i2} \log \left(\frac{\partial^2 S_{pop}(t_1, t_2 | \boldsymbol{\theta})}{\partial t_{i1} \partial t_{i2}} \right) + \delta_{i1}(1 - \delta_{i2}) \log \left(\frac{-\partial S_{pop}(t_1, t_2 | \boldsymbol{\theta})}{\partial t_{i1}} \right) \\ & + \delta_{i2}(1 - \delta_{i1}) \log \left(\frac{-\partial S_{pop}(t_1, t_2 | \boldsymbol{\theta})}{\partial t_{i2}} \right) + (1 - \delta_{i1})(1 - \delta_{i2}) \log S_{pop}(t_1, t_2 | \boldsymbol{\theta}). \end{aligned} \quad (4.5)$$

Consideramos as distribuições *a priori* de forma análoga à Subseção 2.2.1.

Como em Chen *et al.* (1999) incorporamos covariáveis para o modelo de longa duração (4.1) por meio do parâmetro taxa de cura θ como segue: Para cada indivíduo i , $i = 1, \dots, n$, e para cada tempo j , $j = 1, 2$, consideramos $\theta_{ji} \equiv \theta(\mathbf{x}_i' \boldsymbol{\beta}_j) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_j)$, em que $\mathbf{x}_i' = (x_{i1}, \dots, x_{iq})$ denota o vetor $q \times 1$ de covariáveis para o i -ésimo indivíduo e $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jq})'$ denota o correspondente vetor de coeficientes de regressão. Escolhemos as seguintes distribuições *a priori* independentes: $\beta_{0j} \sim Normal(\mu_{0j}; \sigma_{0j}^2)$, $\beta_{1j} \sim Normal(\mu_{1j}; \sigma_{1j}^2)$, para $j = 1, 2$.

4.2.1 Implementação Computacional

Para cada tempo T_j assumimos que a função de sobrevivência própria $S(t_j)$ dado em (4.1) tem uma distribuição Weibull dada em (2.4) com parâmetros α_j e λ_j , $j = 1, 2$. Escolhemos as seguintes distribuições a priori independentes, $\alpha_j \sim Gama(a_j; b_j)$ e $\lambda_j \sim Gama(c_j; d_j)$, para $j = 1, 2$.

A distribuição condicional para cada parâmetro é dada por

$$\pi(\phi|\mathcal{D}, \theta_{(-\phi)}) \propto (1 - \phi)^{r_1-1} (1 + \phi)^{r_2-1} \left(\prod_{i=1}^n \Delta_i \right),$$

$$\pi(\beta_{0j}|\mathcal{D}, \theta_{(-\beta_{0j})}) \propto \left(\prod_{i=1}^n \Delta_i S_{popj}(t_{ji}) \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} + \frac{\mu_{0j}}{n\sigma_{0j}^2} \right\} \beta_{0j} - \frac{\beta_{0j}^2}{2\sigma_{0j}^2} \right),$$

$$\pi(\beta_{1j}|\mathcal{D}, \theta_{(-\beta_{1j})}) \propto \left(\prod_{i=1}^n \Delta_i S_{popj}(t_{ji}) \right) \exp \left(\left\{ \sum_{i=1}^n \delta_{ji} x_i + \frac{\mu_{1j}}{n\sigma_{1j}^2} \right\} \beta_{1j} - \frac{\beta_{1j}^2}{2\sigma_{1j}^2} \right),$$

$$\pi(\lambda_j|\mathcal{D}, \theta_{(-\lambda_j)}) \propto \left(\prod_{i=1}^n \Delta_i S_j(t_{ji}) \right) \lambda_j^{n+c_j-1} \exp \left(\left[\sum_{i=1}^n \theta_{ji} S_j(t_{ji}) \right] - d_j \lambda_j \right)$$

e

$$\pi(r_j|\mathcal{D}, \theta_{(-r_j)}) \propto \left(\prod_{i=1}^n \Delta_i t_{ji}^{r_j-1} S_j(t_{ji}) \right) r_j^{n+a_j-1} \exp \left(\left[\sum_{i=1}^n \theta_{ji} S_j(t_{ji}) \right] - b_j r_j \right),$$

em que $\theta_{ji} = \exp(\beta_{0j} + \beta_{1j} x_i)$, $S_j(t_{ji}) = \exp(-\lambda_j t_{ji}^{r_j})$

e

$S_{popj}(t_{ji}) = \exp(-\theta_{ji}(1 - S_j(t_{ji})))$ para $j = 1, 2$ com $\mathcal{D} = (\mathbf{t}_1, \mathbf{t}_2, \mathbf{d}_1, \mathbf{d}_2, \mathbf{x})$,

$$\Delta_i = \nu_{1i}^{(1-\delta_{1i})(1-\delta_{2i})} \nu_{2i}^{\delta_{1i}\delta_{2i}} \nu_{3i}^{\delta_{1i}(1-\delta_{2i})} \nu_{4i}^{\delta_{2i}(1-\delta_{1i})},$$

$$\nu_{1i} = 1 + \phi \prod_{j=1}^2 (1 - S_{popj}(t_{ji})),$$

$$\nu_{2i} = 1 + \phi \left(1 + \prod_{j=1}^2 (1 - 2S_{popj}(t_{ji})) \right),$$

$$\nu_{3i} = 1 + \phi \left(1 - 2S_{pop1}(t_{ji}) - S_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right)$$

e

$$\nu_{4i} = 1 + \phi \left(1 - S_{pop1}(t_{ji}) - 2S_{pop2}(t_{ji}) + 2 \prod_{j=1}^2 S_{popj}(t_{ji}) \right).$$

4.3 Aplicações

Nesta Seção, resultados de estudos de simulação e um exemplo de dados reais são apresentados com o objetivo de ilustrar a performance da metodologia proposta.

4.3.1 Dados artificiais

Nesta Seção vamos considerar uma amostra artificial assumindo que a sobrevivência própria do modelo $S(t_j)$ dado em (4.1) tem distribuição Weibull (2.4) com parâmetros α_j e λ_j , $j = 1, 2$. As covariáveis x_i foram geradas de uma distribuição Bernoulli com parâmetro 0,5. Os dados bivariados artificiais (T_{i1}, T_{i2}) , $i = 1, \dots, n$, foram gerados assumindo $n = 160$. No primeiro passo, geramos os tempos T_j , $j = 1, 2$, de forma análoga à Subseção 3.2.2.1.

No passo seguinte, para cada nível de covariável k , $k = 0, 1$, fixamos valores para a taxa de cura $\pi_{kj} = \exp\{\theta_{kj}\}$, $j = 1, 2$. Um ponto de corte t_{kj}^* foi escolhido de forma que os tempos mais elevados representam a quantidade $\pi_{kj}\%$ de curados e, foram assumidos censurados. Os seguintes valores de taxa de cura foram fixados: $\pi_{01} = 0,4$, $\pi_{11} = 0,3$, $\pi_{02} = 0,3$ e $\pi_{12} = 0,2$. Foram considerados os seguintes valores para os outros parâmetros do modelo $\alpha_1 = 2$, $\lambda_1 = 0,5$, $\alpha_2 = 1,5$, $\lambda_2 = 0,2$ e $\phi = 0,6$.

A Tabela 4.1 apresenta o resumo para os parâmetros do modelo de sobrevivência de longa duração bivariado baseado na cópula FGM. Os verdadeiros valores são dados entre parênteses.

A Tabela 4.2 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar os modelos de sobrevivência bivariado baseado na cópula FGM e PS com marginais de tempo de promoção Weibull. O modelo baseado na cópula FGM supera o PS em todos os critérios considerados.

Tabela 4.1: Dados simulados. Média *a posteriori* (Média), desvio padrão (DP) e intervalo HPD (95%) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM.

	Parâmetro	Média	DP	HPD (95%)
Tempo 1	α_1 (2)	2,362	0,210	(1,972; 2,793)
	λ_1 (0,5)	0,365	0,162	(0,060; 0,696)
	θ_{01} (0,4)	0,364	0,054	(0,262; 0,470)
	θ_{11} (0,3)	0,315	0,051	(0,219; 0,413)
Tempo 2	α_2 (1,5)	1,678	0,149	(1,382; 1,974)
	λ_2 (0,2)	0,163	0,029	(0,112; 0,226)
	θ_{02} (0,3)	0,238	0,054	(0,128; 0,343)
	θ_{12} (0,2)	0,204	0,049	(0,110; 0,301)
Cópula	ϕ (0,6)	0,587	0,212	(0,151; 0,942)

Tabela 4.2: Dados simulados. Critérios bayesianos.

Modelo	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
FGM	839,750	867,426	830,300	-415,100
PS	843,506	871,183	834,300	-417,128

4.3.2 Diagnóstico de observações influentes

Para examinar o desempenho das medidas de diagnósticos propostas neste trabalho, geramos uma amostra de tamanho 160 com parâmetros fixos $\alpha_1 = 2$, $\lambda_1 = 0,5$, $\alpha_2 = 1,5$, $\lambda_2 = 0,2$ e $\phi = 0,6$. A porcentagem de censuras foi controlado da seguinte forma: Para cada nível de covariável k , $k = 0, 1$, fixamos valores para a taxa de cura $\pi_{kj} = \exp\{\theta_{kj}\}$, $j = 1, 2$. Um ponto de corte t_{kj}^* foi escolhido de forma que os tempos mais elevados representam a quantidade $\pi_{kj}\%$ de curados e, foram assumidos censurados.

Selecionamos os casos 90, 107 e 120 (ambos os tempos observados) e 42 (tempo de vida 1 censurado e tempo de vida 2 observado) para perturbação. Para criar

observações artificialmente influentes no conjunto de dados, escolhemos um, dois ou três desses casos selecionados. Para cada caso, perturbamos um ou ambos os tempos de vida da seguinte forma $\tilde{t}_i = t_i + 7S_t, i = 1, 2$, em que S_t é o desvio padrão dos t_i 's. Para o caso 107 foi perturbado apenas o tempo de vida t_1 , para o caso 90 o tempo de vida t_2 e, para os casos 42 e 120, ambos os tempos de vidas foram perturbados.

A Tabela 4.3 mostra que as inferências *a posteriori* são sensíveis à perturbação do(s) caso(s) selecionado(s). Na Tabela 4.3, o conjunto de dados (a) denota os dados originais simulados sem perturbação, e os conjuntos de dados (b) to (i) denotam os conjuntos de dados com casos perturbados.

Tabela 4.3: Dados simulados. Média *a posteriori* e desvio padrão (DP) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Nome dos dados	Identificação dos casos perturbados	β_{01} Média [DP]	β_{11} Média [DP]	α_1 Média [DP]	λ_1 Média [DP]	β_{02} Média [DP]	β_{12} Média [DP]	α_2 Média [DP]	λ_2 Média [DP]	ϕ Média [DP]
a	Amostra original	0,010 [0,149]	0,135 [0,191]	2,362 [0,210]	0,365 [0,162]	0,369 [0,162]	0,103 [0,178]	1,678 [0,149]	0,163 [0,029]	0,587 [0,212]
b	107	0,362 [0,231]	0,137 [0,197]	1,641 [0,136]	0,426 [0,108]	0,366 [0,162]	0,098 [0,186]	1,672 [0,152]	0,163 [0,030]	0,579 [0,218]
c	90	0,011 [0,150]	0,124 [0,195]	2,385 [0,207]	0,810 [0,113]	0,703 [0,245]	0,082 [0,173]	1,342 [0,112]	0,129 [0,032]	0,611 [0,212]
d	120	0,374 [0,250]	0,143 [0,195]	1,627 [0,129]	0,419 [0,110]	0,698 [0,247]	0,081 [0,179]	1,341 [0,109]	0,130 [0,033]	0,630 [0,207]
e	{90, 107}	0,359 [0,230]	0,137 [0,198]	1,641 [0,132]	0,427 [0,108]	0,681 [0,239]	0,086 [0,183]	1,343 [0,109]	0,133 [0,032]	0,599 [0,221]
f	{107, 120}	0,666 [0,334]	0,097 [0,197]	1,486 [0,123]	0,294 [0,095]	0,705 [0,266]	0,077 [0,186]	1,333 [0,111]	0,130 [0,033]	0,621 [0,213]
g	{90, 107, 120}	0,652 [0,305]	0,099 [0,198]	1,487 [0,119]	0,295 [0,093]	0,886 [0,279]	0,048 [0,184]	1,245 [0,099]	0,112 [0,032]	0,619 [0,217]
h	42	0,002 [0,149]	0,138 [0,194]	2,398 [0,212]	0,809 [0,112]	0,666 [0,243]	0,132 [0,181]	1,355 [0,109]	0,130 [0,033]	0,626 [0,209]
i	{42, 120}	0,349 [0,215]	0,145 [0,200]	1,639 [0,129]	0,429 [0,103]	0,858 [0,296]	0,108 [0,187]	1,250 [0,102]	0,110 [0,031]	0,659 [0,204]

A Tabela 4.4 mostra os critérios bayesianos do ajuste de diferentes casos de conjunto de dados perturbado. Podemos observar que para os dados da amostra original simulada (conjunto de dados (a)) obteve o melhor ajuste.

Tabela 4.4: Dados simulados. Critérios bayesianos ajustando o modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM para cada conjunto de dados simulados.

Conjunto de dados	Critérios bayesianos			
	EAIC	EBIC	DIC	LPML
a	839,750	867,426	830,300	-415,100
b	870,308	897,985	861,100	-434,932
c	857,951	885,628	848,400	-426,561
d	888,006	915,682	878,500	-445,262
e	888,387	916,063	878,900	-446,388
f	898,624	926,301	889,100	-449,301
g	908,782	936,458	899,200	-451,618
h	856,841	884,518	847,400	-425,651
i	897,509	925,185	888,100	-448,408

Vamos considerar a amostra da distribuição *a posteriori* dos parâmetros do modelo de sobrevivência FGM bivariado para obter uma estimativa das quatro medidas de divergência. Os resultados na Tabela 4.5 mostram, antes da perturbação (conjunto de dados (a)), que todos os casos selecionados não são influencias, com pequenas medidas de divergência. Entretanto, após perturbações (conjunto de dados (b) a (i)) as quatro medidas aumentam, indicando que os casos são influentes. Os valores que ultrapassam os pontos de cortes apresentados na Subseção 1.3.1 estão destacados com *.

A Figura 4.1 mostra o gráfico de índice das quatro medidas de divergência para o conjunto de dados (i). Claramente, podemos ver que as quatro medidas de divergência detectam o ponto influente.

Tabela 4.5: Dados simulados. Medidas de divergência.

Nome dos dados	Identificação dos casos	Medidas de divergência			
		d_{K-L}	d_J	d_{L_1}	d_{χ^2}
a	42	0,019	0,038	0,078	0,040
	90	0,010	0,020	0,056	0,020
	107	0,010	0,020	0,056	0,020
	120	0,010	0,020	0,056	0,020
b	107	4,896*	10,367*	0,924*	841,361*
c	90	2,660*	7,257*	0,825*	840,664*
d	120	6,715*	12,354*	0,956*	691,114*
e	90	2,047*	4,628*	0,744*	56,265*
	107	5,602*	12,207*	0,952*	2014,120*
f	107	0,789*	1,893*	0,511*	11,394*
	120	4,680*	11,731*	0,939*	2826,113*
g	90	0,544*	1,330*	0,426*	7,499*
	107	0,737*	1,702*	0,496*	6,910*
	120	1,572*	3,744*	0,669*	57,784*
h	42	2,184*	5,660*	0,768*	342,578*
i	42	0,602*	1,481*	0,441*	8,301*
	120	4,593*	8,921*	0,901*	254,399*

4.3.3 Dados reais

Nesta Seção, como ilustração da metodologia proposta em conjunto de dados reais, reanalisamos (Huster *et al.*, 1989; Sahu & Dey, 2000; Romeo *et al.*, 2006) o conjunto de dados de retinopatia diabética (Subseção 1.4.1). Ao analisarmos este conjunto de dados, não detectamos nenhum ponto influente. Dessa forma, para criar uma observação influente para este conjunto de dados, escolhemos o caso 100 e perturbamos ambos os tempos de vida da seguinte forma $\tilde{t}_u = t_u + 7S_t$, $u = 1, 2$, em que S_t é o desvio-padrão de t_u 's.

Na Tabela 4.6 apresentamos os resumos *a posteriori* para os parâmetros do mo-

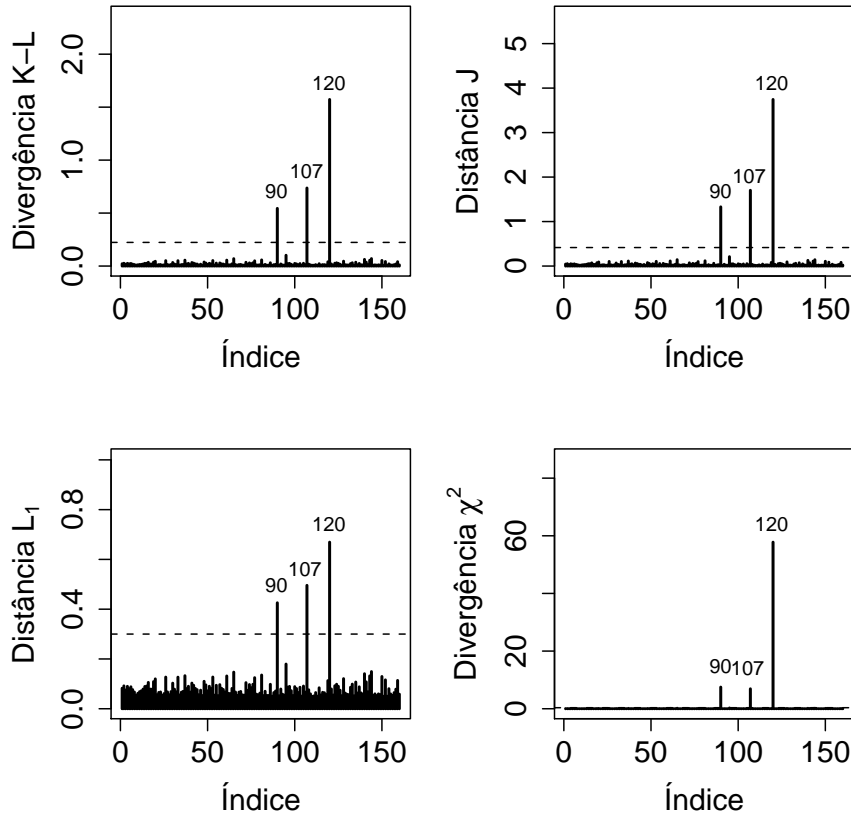


Figura 4.1: Gráficos de índices das medidas de divergência para o caso (i).

delo de sobrevivência bivariado tempo de promoção baseado na cópula FGM com marginais Weibull e Exponencial, considerando os dados originais como também o conjunto de dados com uma observação perturbada. Além de avaliar a robustez do modelo relacionado às escolhas dos hiperparâmetros das distribuições *a priori*, foi realizado um estudo de sensibilidade no qual constatamos que as estimativas dos parâmetros *a posteriori* não apresentaram diferenças significativas e não alteraram os resultados da Tabela 4.6.

Além disso, para ambas as amostras, os intervalos de credibilidade HPD para o parâmetro de locação das distribuições marginais Weibull contém o valor 1, levando a evidência em favor de uma distribuição Exponencial para as marginais.

Considerando modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM, computamos as medidas de divergência, mas não

Tabela 4.6: Dados de retinopatia diabética. Média *a posteriori*, desvio padrão (DP) e HPD (95%) para os parâmetros do modelo de sobrevivência bivariado tempo de promoção baseado na cópula FGM.

	Parâmetro	Tempo de promoção Weibull			Tempo de promoção Exponencial			
		Média	DP	HPD (95%)	Média	DP	HPD (95%)	
Dados originais	Tempo 1	α_1	1,057	0,138	(0,800; 1,330)	-	-	-
		λ_1	0,024	0,010	(0,010; 0,048)	0,026	0,009	(0,020; 0,044)
		β_{01}	-0,570	0,274	(-0,961; 0,187)	-0,523	0,300	(-1,020; 0,212)
		β_{11}	-0,469	0,200	(-0,999; -0,130)	-0,433	0,301	(-1,044 ; 0,114)
	Tempo 2	α_2	0,977	0,090	(0,804; 1,150)	-	-	-
		λ_2	0,019	0,006	(0,008; 0,033)	0,017	0,006	(0,006; 0,030)
		β_{02}	0,155	0,328	(-0,332; 1,035)	0,199	0,341	(-0,335; 1,023)
		β_{12}	0,363	0,176	(0,003; 0,669)	0,350	0,198	(0,031; 0,735)
	Cópula	ϕ	0,830	0,132	(0,499; 0,994)	0,813	0,156	(0,408; 0,993)
	Dados perturbados	Tempo 1	α_1	0,843	0,106	(0,638; 1,073)	-	-
λ_1			0,015	0,014	(0,002; 0,053)	0,015	0,007	(0,003; 0,031)
β_{01}			0,427	0,670	(-0,784; 1,774)	-0,103	0,046	(-0,791; 1,034)
β_{11}			0,239	0,270	(-0,212; 0,842)	-0,522	0,270	(-1,065; -0,011)
Tempo 2		α_2	0,943	0,097	(0,772; 1,143)	-	-	-
		λ_2	0,019	0,006	(0,009; 0,034)	0,014	0,005	(0,005; 0,025)
		β_{02}	-0,228	0,333	(-0,948; 0,174)	0,303	0,342	(-0,236; 1,072)
		β_{12}	0,308	0,255	(-0,206; 0,696)	0,348	0,181	(0,005; 0,740)
Cópula		ϕ	0,780	0,166	(0,385; 0,992)	0,809	0,148	(0,465; 0,993)

encontramos casos muito influentes. Nas Figuras 4.2 e 4.3 apresentamos os gráficos de índices das quatro medidas de divergência considerando os dados originais e perturbados, respectivamente. Podemos observar que para os dados originais (Figura 4.2), nenhum possível ponto influente foi detectado. No entanto, para a os dados perturbados (na Figura 4.3), as quatro medidas de divergência detectam o caso 100, cujos valores são apresentados na Tabela 4.7.

Tabela 4.7: Dados de retinopatia diabética. Medidas de divergência.

Identificação dos casos	Medidas de divergência			
	d_{K-L}	d_J	d_{L_1}	d_{χ^2}
100	1,019	2,289	0,547	15,082

A Figura 4.4 mostra o gráfico de dispersão entre T_1 e T_2 (painel acima) e as

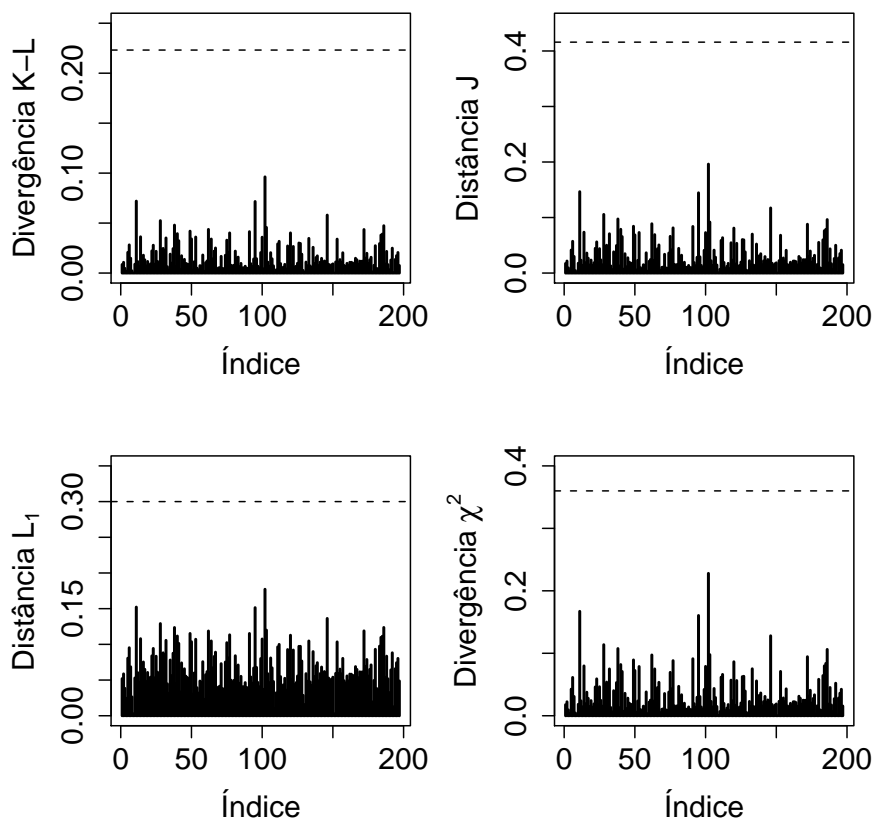


Figura 4.2: Dados de retinopatia diabética. Gráficos de índices das medidas de divergência.

curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), dicotomizadas pela idade do paciente juntamente com o ajuste do modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM.

A Tabela 4.8 apresenta os critérios de comparação de modelos discutidos na Seção 1.2 para comparar modelo de sobrevivência bivariado baseado na cópula FGM com marginais Weibull e Exponencial, com o modelo baseado na cópula de PS como também com os casos assumindo independência entre os tempos.

Com base em todos os critérios bayesianos, há evidências positivas em favor da modelagem FGM, indicando que o modelo de sobrevivência bivariado tempo de promoção baseado na cópula FGM pode ser visto como um concorrente para o modelo PS.

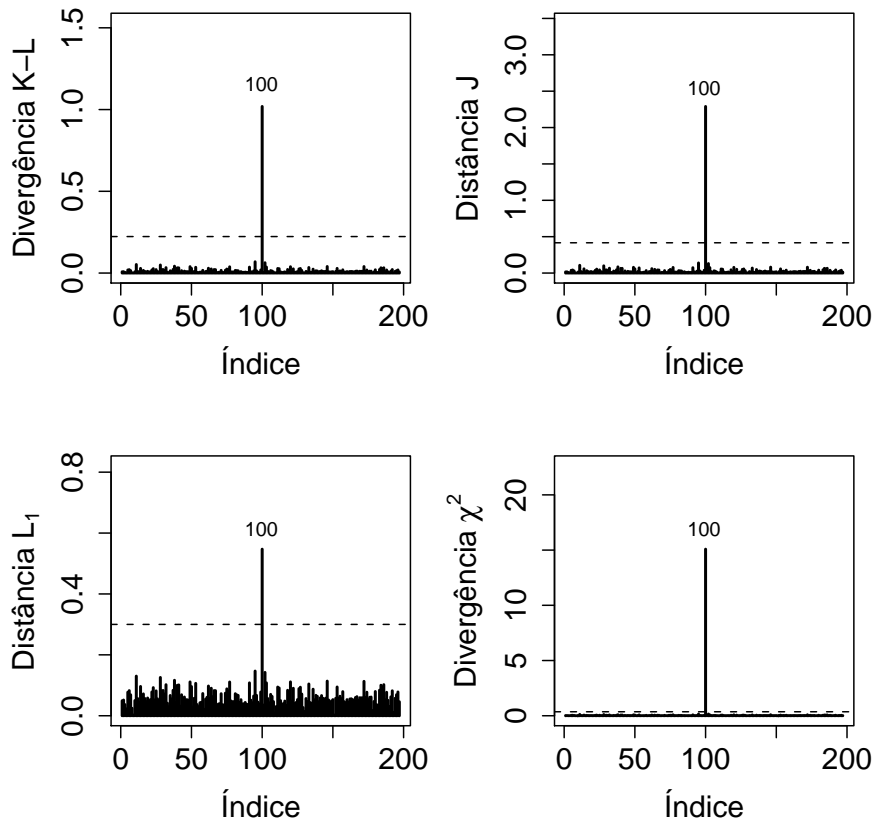


Figura 4.3: Dados de retinopatia diabética perturbado. Gráficos de índices das medidas de divergência.

Como esperado, todos os critérios apresentados evidência em favor da amostra original (sem o ponto influente) e, o melhor ajuste foi o caso assumindo tempo de promoção Exponencial para as marginais.

A Figura 4.5 mostra os gráficos de contorno da sobrevivência para o modelo de sobrevivência bivariado tempo de promoção Weibull baseado na cópula FGM de acordo com a idade do paciente.

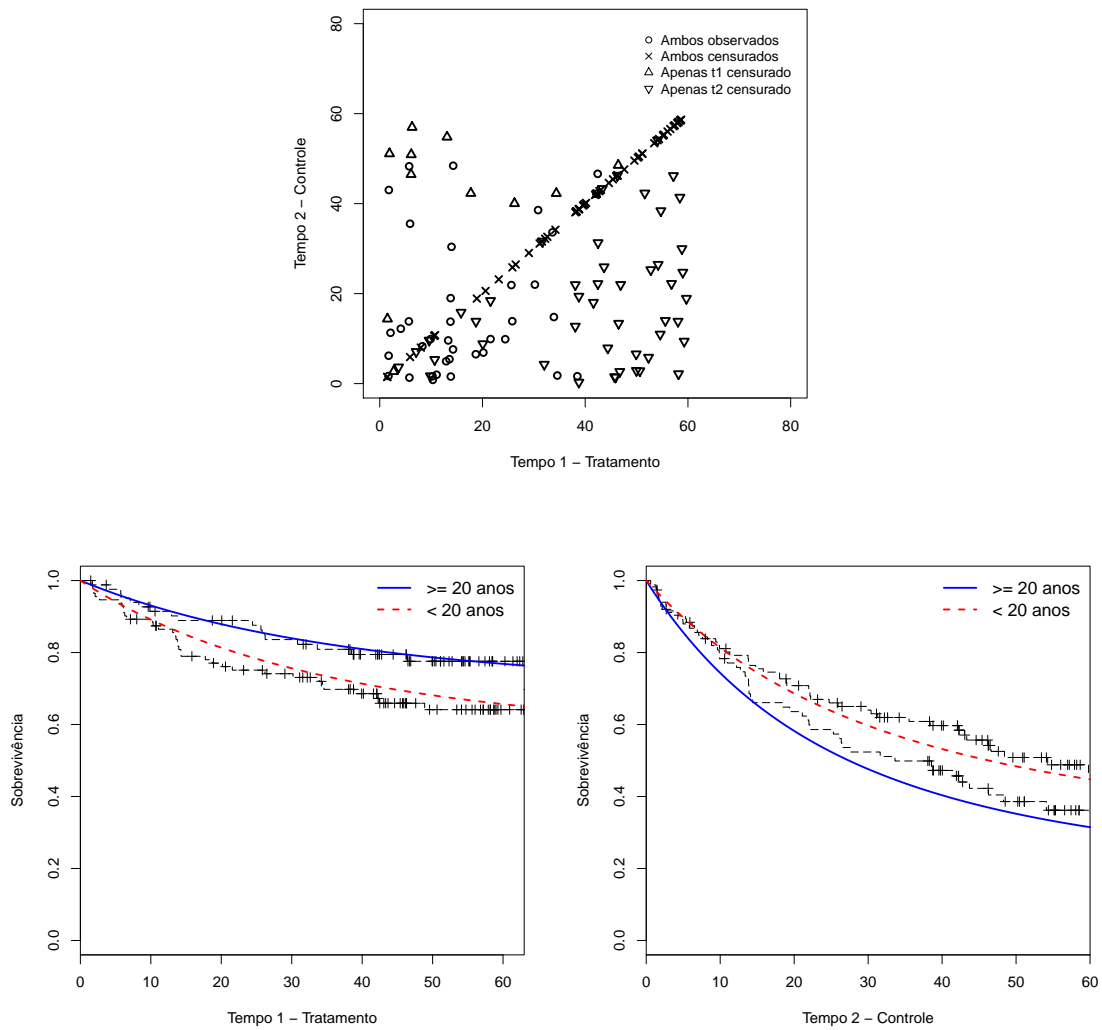


Figura 4.4: Dados de retinopatia diabética. Gráfico de dispersão de T_1 vs T_2 (painel acima). Curvas de Kaplan-Meier para as variáveis T_1 (painel abaixo à esquerda) e T_2 (painel abaixo à direita), juntamente com o ajuste do modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM.

Tabela 4.8: Dados de retinopatia diabética. Critérios bayesianos.

Modelo		Critérios bayesianos			
		EAIC	EBIC	DIC	LPML
Dados Originais	FGM Weibull	1668,667	1698,216	1657,000	-828,157
	FGM Exponencial	1665,295	1688,278	1657,000	-828,812
	PS Weibull	1672,354	1701,903	1662,000	-831,180
	PS Exponencial	1671,071	1700,619	1659,000	-830,188
	Independência Weibull	1682,922	1709,187	1673,000	-837,187
	Independência Exponencial	1677,419	1697,120	1671,000	-835,820
Dados Perturbados	FGM Weibull	1685,183	1714,732	1670,000	-838,286
	FGM Exponencial	1677,427	1700,409	1668,000	-835,423
	PS Weibull	1682,526	1712,074	1668,000	-836,560
	PS Exponencial	1679,256	1702,239	1671,000	-836,787
	Independência Weibull	1694,613	1720,879	1684,000	-843,252
	Independência Exponencial	1689,951	1709,650	1682,000	-842,850

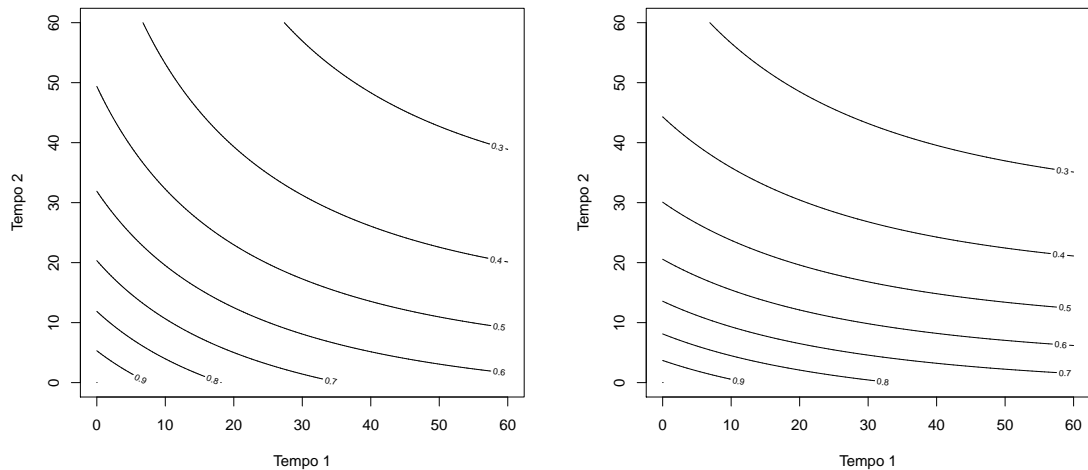


Figura 4.5: Gráfico de contornos para $x = 0$ (painel esquerdo) e $x = 1$ (painel direito) com o ajuste do modelo de sobrevivência bivariado tempo de promoção Exponencial baseado na cópula FGM.

Capítulo 5

Considerações finais e extensões

Neste trabalho observamos que a cópula FGM pode ser um modelo competitivo aos usuais quando os dados apresentam fraca dependência. Também verificamos por meio dos estudos de simulação e dos dados reais que as quatro medidas de divergência detectam os pontos influentes do modelo com um certo padrão de perturbação que foi adotado no trabalho.

Como estudos futuros estamos interessados em utilizar outras distribuições marginais, tais como o modelo de Cox e o modelo Exponencial por partes, como também os modelos propostos por Cancho *et al.* (2011a) e Rodrigues *et al.* (2010) que assumem, respectivamente, as distribuições Binomial Negativa e Poisson ponderada para o número de causas ou risco de ocorrência de um particular evento de interesse. Tais modelos são conhecidos como modelo de longa duração com dispersão e modelo de longa duração destrutivo, respectivamente. Também, utilizar novas funções de ligação para as frações de cura.

Além disso, realizar uma comparação dos resultados obtidos com o modelo baseado nos mecanismos de ativação com base em fatores latentes (Cooner *et al.*, 2007).

Uma das críticas sobre os modelos de cópulas é o fato de que estes desconsideram o comportamento marginal. Fernandez & Kolev (2007) mostram que o comportamento das densidades marginais (crescente, decrescente, unimodal etc.) influencia na estrutura de dependência descrita pela distribuição conjunta. Gonçalves (2008)

apresenta a função de dependência de Sibuya (Sibuya, 1960) como uma alternativa ao uso de cópulas, propondo um procedimento de como utilizá-la juntamente com as cópulas. Motivados por este procedimento, pretendemos implementá-lo em nossa análise.

Schultz (2011), sob um enfoque bayesiano, apresenta uma metodologia para inferência em modelos com cópulas FGM e suas generalizações. Generalizações da cópula FGM podem ser encontradas em Rodríguez-Lallena & Úbeda-Flores (2004), Huang & Kotz (1999) e Amblard & Girard (2009). Extensões deste trabalho para essas formas generalizadas da cópula FGM podem ser exploradas.

Alguns Programas

Neste apêndice apresentamos dois programas implementados no sistema *OpenBUGS* referentes ao Capítulo 3.

Modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula FGM com covariáveis utilizadas no parâmetro de fração de cura.

```
model
{
for (i in 1:N){
p1[i]←exp(beta01+beta11*x[i])/(1+exp(beta01+beta11*x[i]))
p2[i]←exp(beta02+beta12*x[i])/(1+exp(beta02+beta12*x[i]))
sp1[i]←p1[i]+(1-p1[i])*exp(-theta1*pow(t1[i],r1))
sp2[i]←p2[i]+(1-p2[i])*exp(-theta2*pow(t2[i],r2))
Fp1[i]←1-sp1[i]
Fp2[i]←1-sp2[i]
fp1[i]←(1-p1[i])*exp(-theta1*pow(t1[i],r1))*theta1*r1*pow(t1[i],r1-1)
fp2[i]←(1-p2[i])*exp(-theta2*pow(t2[i],r2))*theta2*r2*pow(t2[i],r2-1)
s[i]←1+phi*Fp1[i]*Fp2[i]
q1[i]←1+phi*(2*Fp1[i]-1)*(2*Fp2[i]-1)
q2[i]←1+phi*Fp2[i]*(2*Fp1[i]-1)
q3[i]←1+phi*Fp1[i]*(2*Fp2[i]-1)
}
```

Função de verossimilhança

```
L[i]←pow(fp1[i],d1[i])*pow(fp2[i],d2[i])*pow(sp1[i],1-d1[i])*pow(sp2[i],1-d2[i])*  
pow(s[i],(1-d1[i])*(1-d2[i]))*pow(q1[i],d1[i]*d2[i])*pow(q2[i],d1[i]*(1-d2[i]))*  
pow(q3[i],d2[i]*(1-d1[i]))
```

CPO

```
CPO[i]←1/L[i]
```

Logaritmo da função de verossimilhança

```
loglik[i]←log(L[i])
```

```
dummy[i]←0
```

```
dummy[i]∼dgeneric(loglik[i])
```

```
}
```

```
phi←2*alpha-1
```

Distribuições *a priori*

```
beta01∼dnorm(0,0.001)
```

```
beta02∼dnorm(0,0.001)
```

```
beta11∼dnorm(0,0.001)
```

```
beta12∼dnorm(0,0.001)
```

```
r1∼dgamma(1,0.001)
```

```
r2∼dgamma(1,0.001)
```

```
alpha∼dunif(0,1)
```

```
theta1∼dgamma(1,0.001)
```

```
theta2∼dgamma(1,0.001)
```

```
}
```

Modelo de sobrevivência bivariado mistura padrão Weibull baseado na cópula PS com covariáveis utilizadas no parâmetro de fração de cura.

```

model
{
for (i in 1:N){
p1[i]←exp(beta01+beta11*x[i])/(1+exp(beta01+beta11*x[i]))
p2[i]←exp(beta02+beta12*x[i])/(1+exp(beta02+beta12*x[i]))
sp1[i]←p1[i]+(1-p1[i])*exp(-theta1*pow(t1[i],r1))
sp2[i]←p2[i]+(1-p2[i])*exp(-theta2*pow(t2[i],r2))
sp1l[i]←(1-p1[i])*exp(-theta1*pow(t1[i],r1))*theta1*r1*pow(t1[i],r1-1)
sp2l[i]←(1-p2[i])*exp(-theta2*pow(t2[i],r2))*theta2*r2*pow(t2[i],r2-1)
v1[i]←pow(-log(sp1[i]),1/alpha)
v2[i]←pow(-log(sp2[i]),1/alpha)
s[i]←pow(v1[i]+v2[i],alpha)

```

Função de verossimilhança

$$L[i] \leftarrow \exp(-s[i]) \cdot \text{pow}(s[i], d1[i] + d2[i] - d1[i] \cdot d2[i]) \cdot \text{pow}(v1[i] \cdot \text{sp1l}[i], d1[i]) \cdot \text{pow}(v2[i] \cdot \text{sp2l}[i], d2[i]) \cdot \text{pow}(v1[i] + v2[i], -(d1[i] + d2[i])) \cdot \text{pow}(\text{sp1}[i] \cdot \log(\text{sp1}[i]), -d1[i]) \cdot \text{pow}(\text{sp2}[i] \cdot \log(\text{sp2}[i]), -d2[i]) \cdot \text{pow}(\alpha, -d1[i] \cdot d2[i]) \cdot \text{pow}(1 - \alpha + \alpha \cdot s[i], d1[i] \cdot d2[i]) \cdot \text{pow}(-1, d1[i] + d2[i] - 2 \cdot d1[i] \cdot d2[i])$$

CPO

$$\text{CPO}[i] \leftarrow 1/L[i]$$

Logaritmo da função de verossimilhança

$$\text{zeros}[i] \leftarrow 0$$

$$\text{phi}[i] \leftarrow \log(L[i])$$

```
zeros[i]~dgeneric(phi[i])  
}
```

Distribuições *a priori*

```
alpha~dbeta(1,1)  
r1~dgamma(1,0.001)  
r2~dgamma(1,0.001)  
beta01~dnorm(0,0.001)  
beta11~dnorm(0,0.001)  
beta02~dnorm(0,0.001)  
beta12~dnorm(0,0.001)  
theta1~dgamma(1,0.001)  
theta2~dgamma(1,0.001)  
}
```

Referências Bibliográficas

- Aarset, M. V. (1985). The null distribution for a test of constant versus “bathtub” failure rate. *Scandinavian Journal of Statistics*, **12**, 55-68.
- Aalen, O. & Hjort, N.L. (2002). Frailty models that yield proportional hazards. *Statistics & Probability Letters*, **58**, 335-342.
- Amblard, C. & Girard, S. (2009). A new extension of bivariate FGM copulas. *Metrika*, **70**, 1-17.
- Anjos, U. U., Ferreira, F. H., Kolev, N. V. & Mendes, B. V. M. (2004). *Modelando Dependências via Cópulas*. São Paulo: Associação Brasileira de Estatística.
- Balakrishnan, N. & Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, **25**, 2797-2816.
- Barlow, R. E. & Campo, R. (1975). *Total time on test processes and applications to failure data analysis*. In Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment, pages 451-481, eds. J. B. Fussell, R. E. Barlow, N. D. Singpurwalla, & Z. W. Birnbaum. CA: University of Berkeley.
- Berkson, J. & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501-515.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*, **11**, 15-53.

- Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. **64**, 616-618.
- BUGS team (2007). *An R / S-PLUS package containing OpenBUGS and its R / S-PLUS interface BRugs*. <http://mathstat.helsinki.fi/openbugs/>.
- Cancho, V., Ortega, E. & Paula, G. (2010). On estimation and influence diagnostics for log-Birnbaum-Saunders Student-t regression models: Full Bayesian analysis. *Journal of Statistical Planning and Inference*, **140**, 2486-2496.
- Cancho, V. G., Rodrigues, J. & de Castro, M. (2011). A flexive model for survival with a cure rate: A Bayesian approach. *Journal of Applied Statistics*, **38**, 57-70.
- Cancho, V. G., Dey, D. K., Lachos, V. H. & Andrade, M. G. (2011). Bayesian non-linear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis*, **55**, 588-602.
- Candiani, T. M. S., Pinto, J., Cardoso, C. A. A., Carvalho, I. R., Dias, A. C. M., Carneiro, M. & Goulart, E. A. (2007). Impact of highly active antiretroviral therapy (HAART) on the incidence of opportunistic infections, hospitalizations and mortality among children and adolescents living with HIV/AIDS in Belo Horizonte. *Cadernos de Saúde Pública*, **23**, S414-S423.
- Carlin, B. P. & Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Boca Raton: Chapman and Hall.
- Chen M. H., Ibrahim J. G. & Sinha D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909-919.
- Chen, M. H., Shao, Q. M. & Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Hoboken, New York: John Wiley & Sons.

- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327-335.
- Cho, H., Ibrahim, J. G., Sinha, D. & Shu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, **65**, 116-124.
- Clayton, D. G. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.
- Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **47**, 467-485.
- Colosimo, E. A. & Giolo, S. R. (2006). *Análise de Sobrevida Aplicada*. São Paulo: Edgard Blücher.
- Conway, D. A. (1983). *Farlie-Gumbel-Morgenstern distributions*. In: *Encyclopedia of Statistical Sciences* **3**. S. Kotz and N. L. Johnson editors, 28-31. New York: John Wiley & Sons.
- Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Boca Raton: Chapman and Hall.
- Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society Series B*, **48**, 133-169.
- Cooner, F., Banerjee, S., Carlin, B. & Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**, 560-572.
- Cowless, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, **91**, 883-904.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society Series B*, **30**, 248-275.

- 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.
- Dey, D. & Birmiwal, L. (1994). Robust Bayesian analysis using divergence measures. *Statistics & Probability Letters*, **20**, 287-294.
- Duchateau, L. & Janssen, P. (2008). *The Frailty Model*. Springer: New York.
- Embrechts, P., Linskog, F. & Mcniel, A. (2003). *Modelling dependence with copulas and applications to risks management*. Disponível em <http://www.math.ethz.ch/~embrechts/papers.html>. 10/05/2011.
- Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307-323.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. 1, 3rd edn. New York: Wiley.
- Fernandez, M. & Kolev, N. (2007). Bivariate density characterization by geometry of marginals. *Forthcoming in Economic Quality Control*, **22**, 3-18.
- Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, **19**, 194-226.
- Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd. Boca Raton: Chapman and Hall.
- Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics*, **4**, 147-167.
- Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
- Gelman, A., Carlin, J. & Rubin, D. (2006). *Bayesian Data Analysis*. Boca Ranton: Chapman and Hall.

- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.
- Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, **74**, 549-555.
- Genest G. & Mackay J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, **40**, 280-283.
- Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, **75**, 667-672.
- Goethals, K., Janssen, P. & Duchateau, L. (2008). Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, **35**, 1071-1079.
- Goldman, A. I. (1984). Survivorship analysis when cure is a possibility: A monte carlo study. *Statistics in Medicine*, **3**, 153-163.
- Gonçalves, M. (2008). *Um estudo sobre funções de dependência e medidas de risco*. Tese de doutorado, Instituto de Matemática e Estatística, USP, São Paulo, Brasil.
- Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Gustafson, P., Aeschliman, D. & Levy, A. R. (2003). A simple approach to fitting Bayesian survival models. *Lifetime Data Analysis*, **9**, 5-19.
- Hanagal, D. (2011). *Modeling Survival Data Using Frailty Models*. Boca Raton: Chapman and Hall.
- Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika*, **73**, 671-678.
- Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387-396.

- Hougaard, P. (1989). Fitting a multivariate failure time distribution. *IEEE Transactions on Reliability*, **38**, 444-448.
- Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, **1**, 255-273.
- Huang, J. S. and Kotz, S. (1999). Modifications of the Farlie-Gumbel-Morgenstern distributions: A tough hill to climb. *Metrika*, **49**, 135-145.
- Huster, W. J., Brookmeyer, R. & Self, S. G. (1989). Modeling paired survival data with covariates. *Biometrics*, **45**, 145-156.
- Hutchinson, T. P. & Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasising Applications*, Adelaide: Rumsby Scientific Publishing.
- Ibrahim, J. G., Chen, M-H. & Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.
- Jaworski, P. (2010). *Copula Theory and Its Applications*. In Proceedings of the Workshop Held in Warsaw, 25-26 September 2009, Volume 198, Lecture Notes in Statistics, eds. P. Jaworski, F. Durante, W. Härdle & T. Rychlik. Heidelberg: Springer.
- Kalbfleisch J. D. & Prentice R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. New York: John Wiley & Sons.
- Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.
- Kim, S., Chen, M. H., Dey, D. K. & Gamerman, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, **13**, 17-35.
- Kimeldorf, G. & Sampson, A. R. (1975). Uniform representations of bivariate distributions. *Communications in Statistics - Theory and Methods*, **4**, 617-627.
- Kolev, N., dos Anjos, U. & Mendes, B. V. M. (2006). Copulas: A review and recent developments. *Stochastic Models*, **22**, 617-660.

- Labra, F. V., Aoki, R. & Bolfarine, H. (2005). Local influence in null intercept measurement error regression under a Student-t model. *Journal of Applied Statistics*, **32**, 723-740.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. New York: Wiley & Sons.
- Longini, I. M. & Halloran, M. E. (1996). A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*, **45**, 165-173.
- Louzada-Neto F., Mazucheli J. & Achcar J. A. (2002). *Introdução à Análise de Sobrevida e Confiabilidade*, III Jornada Regional de Estatística e II Semana de Estatística. Universidade Estadual de Maringá, Maringá - PR.
- Maller, R. A. & Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley & Sons.
- Manatunga, A. K. & Oakes, D. (1999). Parametric analysis of matched pair survival data. *Lifetime Data Analysis*, **5**, 371-387.
- McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, **84**, 473-478.
- Meester, S. & MacKay, J. (1994). A parametric model for cluster correlated categorical data. *Biometrics*, **50**, 954-963.
- Nelsen, R. (1986). Properties of a one-parametric family of bivariate distributions with specified marginals. *Communications in Statistics-Theory and Methods*, **15**, 3277-3285.
- Nelsen, R. (2006). *An Introduction to Copulas*, 2nd edn. New York: Springer.
- Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B*, **44**, 414-422.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**, 487-493.

- Oakes, D. (2001). *On Frailty Models and Copulas Recent Advances in Statistical Methods. Proceedings of Statistics 2001 Canada: The 4th Conference in Applied Statistics Montreal, 218-224*. Canada, 6-8 July 2001.
- Ortega, E., Bolfarine, H. & Paula, G. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics & Data Analysis*, **42**, 165-186.
- Ortega, E. M. M., Cancho, V. G. & Bolfarine, H. (2006). Influence diagnostics in exponentiated-Weibull regression models with censored data. *Statistics and Operations Research Transactions*, **30**, 171-192.
- Peng, F. & Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics - La Revue Canadienne de Statistique*, **23**, 199-213.
- Peng, Y., Dear, K. B. G. & Denham, J. W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, **17**, 813-830.
- Perdoná, G. S. C. & Louzada-Neto, F. (2011). A general hazard model for lifetime data in the presence of cure rate. *Journal of Applied Statistics*, **38**, 1395-1405.
- Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, **46**, 863-867.
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). *Output analysis and diagnostics for MCMC*. <http://cran.r-project.org/web/packages/coda/index.html>.
- R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Rodrigues, J., de Castro, M., Cancho, V. G., & Louzada Neto, F. (2009). On the unification of long-survival models. *Statistics and Probabilities Letters*, **79**, 753-759.

- Rodrigues, J., Cancho, V., de Castro, M. & Balakrishnan, N. (2010). A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data. *Statistical Methods in Medical Research*. DOI: 10.1177/0962280210391443
- Rodríguez-Lallena, J. A. & Úbeda-Flores, M. (2004). A new class of bivariate copulas. *Statistics & Probability Letters*, **66**, 315-325.
- Romeo, J. S., Tanaka, N. I. & Pedroso de Lima, A. C. (2006). Bivariate survival modeling: a Bayesian approach based on copulas. *Lifetime Data Analysis*, **12**, 205-222.
- Ross, G. J. S. & Preece, D. A. (1985). The negative binomial distribution. *The Statistician*, **34**, 323-336.
- Saha, K. & Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, **61**, 179-185.
- Sahu, S. K. & Dey, D. K. (2000). A comparison of frailty and other models for bivariate survival data. *Lifetime Data Analysis*, **6**, 207-228.
- Salvadori, G., De Michele, C., Kottegoda, N. T. & Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Water Science and Technology Library **56**. Dordrecht (NL): Springer.
- Schultz, J. A. A. (2011). *FGM e suas generalizações sob um ponto de vista bayesiano*. Dissertação de mestrado, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, Campinas, Brasil.
- Self, S. G. & Prentice, R. L. (1986). *Incorporating random effects into multivariate relative risk regression models*. In Modern Statistical Methods in Chronic Disease Epidemiology, pages 167-177, eds. S. H. Moogaukar & R. L. Prentice. NY: New York.
- Shih, J. H. & Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384-1399.

- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, 195-210.
- Sinha, D. & Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, **92**, 1195-1212.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, **64**, 583-639.
- Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2007). *OpenBUGS user manual, version 3.0.2*. MRC Biostatistics Unit: Cambridge. <http://mathstat.helsinki.fi/openbugs/>.
- The Diabetic Retinopathy Study Research Group (1976). Preliminary report on the effect of photo-coagulation therapy. *American Journal of Ophthalmology*, **81**, 383-396.
- Therneau, T. M. & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.
- Tomazella, V., Louzada-Neto, F. & da Silva, G. L. (2006). Bayesian modeling of recurrent events data with an additive gamma frailty distribution and a homogeneous Poisson process. *Journal of Statistical Theory and Applications*, **5**, 417-429.
- Trivedi, P. K. & Zimmer, D. M. (2005). Copula modelling: an introduction for practitioners. *Foundations and Trends in Econometrics*, **1**, 1-111.
- Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**, 1063-1078.
- Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439-454.

- Vidal, I. & Castro, L. M. (2010). Influential observations in the independent Student-t measurement error model with weak nondifferential error. *Chilean Journal of Statistics*, **1**, 17-34.
- Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society Series B*, 739-750.
- Wienke, A., Lichtenstein, P. & Yashin, A. I. (2003). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, **59**, 1178-1183.
- Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton: Chapman & Hall/CRC.
- Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific: Singapore.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of permanent employment in Japan. *Journal of the American Statistical Association*, **87**, 284-292.
- Yin, G. & Ibrahim, J. G. (2005a). A general class of Bayesian survival models with zero and nonzero cure fractions. *Biometrics*, **61**, 403-412.
- Yin, G. & Ibrahim, J. G. (2005b). Cure rate models: a unified approach. *The Canadian Journal of Statistics*, **33**, 559-570.
- Zhang, S., Zhang, Y., Chaloner, K. & Stapleton, J. T. (2010). A copula model for bivariate hybrid censored survival data with application to the MACS study. *Lifetime Data Analysis*, **16**, 231-249.