

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Modelos de sobrevivência bivariados induzidos por fragilidade

Elizabeth Chipa Bedia

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Elizabeth Chipa Bedia

Modelos de sobrevivência bivariados induzidos por fragilidade

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutora em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística.
VERSÃO REVISADA

Área de Concentração: Estatística

Orientador: Prof. Dr. Vicente Garibay Cancho

USP – São Carlos
Setembro de 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

B412m Bedia, Elizabeth Chipa
 Modelos de sobrevivência bivariados induzido por
 fragilidade / Elizabeth Chipa Bedia; orientador
 Vicente Garibay Cancho. -- São Carlos, 2022.
 116 p.

 Tese (Doutorado - Programa Interinstitucional de
 Pós-graduação em Estatística) -- Instituto de Ciências
 Matemáticas e de Computação, Universidade de São
 Paulo, 2022.

 1. Modelo de fragilidade compartilhada. 2.
 Riscos semicompetitivos. 3. Processos de doença-
 morte. 4. Modelos de longa duração. 5. Modelos de
 fragilidade discreta. I. Cancho, Vicente Garibay,
 orient. II. Título.

Elizabeth Chipa Bedia

Bivariate survival models induced by frailty

Thesis submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Vicente Garibay Cancho

USP – São Carlos
September 2022

Dedico essa tese a todas as pessoas que me apoiaram durante esse processo. A meus pais que sempre me incentivaram a estudar e seguir em frente. A meus irmãos, que sempre acreditaram em mim. A meu querido esposo, por todo o amor, apoio e compreensão em dias difíceis. A meus amigos, por nunca me abandonarem nessa jornada. A meu orientador e professores, por terem me guiado neste caminho, que com base em suas experiências e sabedoria, conseguiram direcionar meus conhecimentos. Dedico esse título de doutora a todos vocês com muito amor e com a certeza de que estarão presentes em todas minhas próximas conquistas!

AGRADECIMENTOS

Em primeiro lugar quero agradecer a meus pais, Paulina e Timoteo, pilares fundamentais na minha vida, por todo amor e dedicação, e por todo o sacrifício feito para que eu pudesse estudar e me formar. A meus irmãos, Luis Miguel, Miguel Angel e Gerson Alaín, por compartilhar momentos de alegria e pelo constante apoio, moral.

A meu amado companheiro de vida, Joel Coacalle, quem com cada dificuldade me fez lembrar sobre a minha capacidade e quem foi o principal impulso para alcançar esse objetivo que parecia tão distante e que finalmente consegui alcançar.

Ao meu orientador, Prof. Dr. Vicente Garibay Cancho, não somente pelos ensinamentos e sugestões que foram fundamentais na minha formação acadêmica, mas também pela paciência, confiança e conselhos.

Aos todos os professores do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs) do Instituto de Ciências Matemáticas e de Computação (ICMC-USP) e do Departamento de Estatística (DEs-UFSCar), que de forma direta ou indireta contribuíram para minha formação acadêmica.

A todos os amigos feitos durante o Doutorado, em especial a Yuri, Fiorela, Cristel e Ana Paula, pelas reflexões, críticas e sugestões recebidas. A meus queridos amigos Carmen, José, Juan José, Mónica e Ramon que nunca estiveram ausentes e sempre disponibilizaram a amizade e o carinho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

BEDIA, E. CH. **Modelos de sobrevivência bivariados induzidos por fragilidade**. 2022. 116 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Modelos de fragilidade foram desenvolvidos para quantificar tanto a heterogeneidade quanto a associação em dados multivariados de tempos de eventos. As distribuições de fragilidade utilizadas em muitos estudos incluem as distribuições gama, Inversa Gaussiana (IG), ou a Positiva estável (PE). Estas distribuições geralmente são escolhidas devido à simplicidade analítica e computacional ou por alguma propriedade atrativa do modelo. A escolha da distribuição da fragilidade é de fundamental importância para assim chegar a uma boa descrição da estrutura de dependência presente nos dados. Uma alternativa para o problema da escolha do modelo de fragilidade seria escolher apenas uma família de distribuições de fragilidade e usá-la como modelo geral. Neste trabalho, estudamos dados de sobrevivência bivariados com estrutura de riscos semicompetitivos (FINE; JIANG; CHAPPELL, 2001) e dados bivariados de longa duração. Para incorporar uma estrutura de dependência entre os tempos de eventos propomos a família de distribuições Power variance function (PVF) como modelo de fragilidade compartilhada a qual inclui as distribuições antes mencionadas. Dados com estrutura de riscos semicompetitivos surge como uma variante da estrutura de riscos competitivos. Na estrutura de riscos semicompetitivos, usualmente, dois eventos são considerados, a saber, um terminal e um não terminal. Sendo que, o evento terminal censura o evento não terminal, mas não vice-versa. Geralmente, os dois eventos estão correlacionados. Então a dependência entre o tempo de falha do evento terminal e o não terminal é incorporada através da fragilidade PVF compartilhada entre as taxas de transição condicional do modelo de doença-morte que é equivalente a um problema de riscos semicompetitivos (XU; KALBFLEISCH; TAI, 2010). Para os dados bivariados de longa duração, que caracterizam-se por possuir uma fração de indivíduos não suscetíveis ao evento de interesse após um longo tempo, foram consideradas situações em que existem dois tipos de causas não observáveis, onde cada causa está relacionada com tempos de ocorrência de um evento de interesse. Para modelar a dependência entre os dois tempos introduzimos uma variável de fragilidade PVF. Para ambos os modelos, um estudo de simulação é apresentado para avaliar o desempenho do método de máxima verossimilhança na estimativa de parâmetros. Finalmente, dados de câncer de cólon são usados na aplicação do modelo com estrutura de riscos semicompetitivos e dados de churn de clientes brasileiros em uma instituição financeira são usados na aplicação do modelos de longa duração.

Palavras-chave: Modelo de fragilidade compartilhada, Riscos semicompetitivos, Processos de doença-morte, Modelos de longa duração, Modelos de fragilidade discreta..

ABSTRACT

BEDIA, E. CH. **Bivariate survival models induced by frailty**. 2022. 116 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Frailty models have been developed to quantify both heterogeneity as well as association in multivariate time-to-event data. The frailty distributions used in many studies include the gamma, inverse Gaussian (IG) or stable positive (SP) distributions. These distributions are usually chosen due to analytical and computational simplicity or due to some attractive property of the model. The choice of the frailty distribution is of fundamental importance in order to arrive at a good description of the dependence structure present in the data. An alternative to the problem of choosing the frailty model would be to choose only one family of frailty distributions and use it as a general model. In this work, we studied bivariate survival data with a semicompeting risk structure (FINE; JIANG; CHAPPELL, 2001), and long-term bivariate data. In order to incorporate a dependence structure between the times of events, we propose the family of distributions Power variance function (PVF) as a shared frailty model which includes the above mentioned distributions. Data with a semicompeting risk structure arises as a variant of the competing risk structure. In the semicompeting risk framework, usually, two events are considered, namely, a terminal and a non-terminal. The terminal event censors the non-terminal event, but not vice versa. Generally, the two events are correlated. So the dependence between the terminal and non-terminal failure time is incorporated through the PVF shared frailty between the conditional transition rates of the illness-death model (XU; KALBFLEISCH; TAI, 2010), that is equivalent to a semicompeting risks problem. For long-term bivariate data, which are characterized by having a fraction of individuals non-susceptible to the event of interest after a long time, were considered situations in which there are two types of unobservable causes, where each cause is related to occurrence times of an event of interest. To model the dependence between the two times we introduce a PVF frailty variable. For both models, a simulation study is presented to evaluate the performance of the maximum likelihood method in the parameters estimation. Finally, colon cancer data are used in the application of the model with a semicompeting risk structure and Brazilian customer churn data in a financial institution are used in the application of the long-term models.

Keywords: Shared frailty model, Semicompeting risks, Illness-death process, Long-term model, Discrete fragility models..

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 – A figura ilustra possíveis observações em dados bivariados (painel esquerdo), dados de riscos semicompetitivos (painel central) e dados de riscos concorrentes (painel direito). Um ponto indica que T_1 e T_2 foram observados e uma seta indica a censura do tempo de sobrevivência na direção apontada. | 25 |
| Figura 2 – Representação de riscos semicompetitivos (esquerda) e riscos competitivos (direita). | 31 |
| Figura 3 – Diagrama de dispersão de T_1 e T_2 para o conjunto de dados de câncer de cólon (panel esquerdo). Modelo de Doença-Morte para dados com estrutura de riscos semicompetitivos (panel direito). | 52 |
| Figura 4 – Função densidade conjunta de (T_1, T_2) , em $T_1 \leq T_2$. Com $\gamma = -0.5$ (painel esquerdo), $\gamma = 0$ (painel central) e $\gamma = 0.5$ (painel direito) | 62 |
| Figura 5 – Função densidade para T_2 , quando $T_1 > T_2$ ($T_1 = \infty$). Com $\theta = 0.5$ (painel esquerdo), $\theta = 1$ (painel central) e $\theta = 2$ (painel direito) | 63 |
| Figura 6 – Medida de associação local entre os tempos T_1 e T_2 , para $\gamma = -0.5$ (painel esquerdo), $\gamma = 0$ (painel central) e $\gamma = 0.5$ (painel direito) | 64 |
| Figura 7 – Função de verossimilhança perfilada de γ . Modelo geral (painel esquerdo), modelo restrito (painel direito) | 70 |
| Figura 8 – razão de risco explicativa condicional estimado, EHR , para cada tratamento. | 73 |
| Figura 9 – Taxas de transição estimadas para a recorrência, $\hat{\lambda}_1(t)$, (painel esquerdo) e para a morte, $\hat{\lambda}_2(t)$, (painel direito), estratificado pela variável explicativa tratamento. | 73 |
| Figura 10 – Taxas de transição marginal estimadas para óbito, dado que um paciente teve recorrência do tumor em t_1 , $\hat{\lambda}_1(t_2 t_1)$, estratificado pela variável explicativa tratamento, para $t_1 = 1, 3, 6$ e 9 anos. | 74 |
| Figura 11 – Probabilidades de transição marginal estimadas para a recorrência, $\hat{P}_1(t)$, (painel superior esquerdo) e para a morte, $\hat{P}_2(t)$, (painel superior direito) no instante t , estratificado pela variável explicativa tratamento. Comparação entre probabilidade de transição marginal para recorrência (azul) e morte (preto) (painel inferior esquerdo). Probabilidade de transição marginal estimada para a morte, dado que um paciente estava no estado de recorrência no instante t_1 , $\hat{P}_{12}(t t_1)$, (painel inferior direito). | 75 |

| | |
|--|----|
| Figura 12 – Função de densidade dada em (5.10) e seu gráfico de contorno do novo modelo para os parâmetros $\theta = 0$ (painel esquerdo), $\theta = 1$ (painel central), $\theta = 4$ (painel direito), $\rho_1 = 1$, $\rho_2 = 0.4$ e função de sobrevivência $S_{01}(t_1) = e^{-t_1^2}$ e $S_{01}(t_2) = e^{-t_2^3}$ | 81 |
| Figura 13 – Medida de associação local entre T_1 e T_2 , $\vartheta^*(t_1, t_2)$, do novo modelo para os parâmetros $\gamma = -0.5$ e $\theta = 1.5$ (painel esquerdo), $\gamma = 0$ e $\theta = 1$ (painel central), $\gamma = 0.5$ e $\theta = 0.5$ (painel direito), $\rho_1 = 1$, $\rho_2 = 0.4$ e função de sobrevivência $F_{01}(t_1) = 1 - e^{-t_1^2}$ e $F_{02}(t_2) = 1 - e^{-t_2^3}$ | 81 |
| Figura 14 – Gráfico de dispersão (painel esquerdo) e Estimativas de Kaplan-Meier das funções de sobrevivência do produto 1 e 2 (painel direito) | 88 |
| Figura 15 – QQ-plot dos resíduos quantílicos randomizados normalizados com linha de identidade para a função de sobrevivência marginal | 89 |
| Figura 16 – A probabilidade de não cancelar os produtos 1 (painel esquerdo) e 2 (painel direito) de seis clientes de uma banco brasileiro, dado que não cancelaram após o terceiro ano | 93 |
| Figura 17 – Estimativas de Kaplan-Meier das funções de risco acumulado e do modelo proposto para o produto 1 (painel esquerdo) e 2 (painel direito) | 93 |
| Figura 18 – Uma situação de riscos semicompetitivos em que a falha do evento terminal consiste em dois eventos separados, $2a$ e $2b$. As taxas de transição para cada transição estão incluídas. | 97 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 1 – Possíveis observações para dados de riscos semicompetitivos. | 31 |
| Tabela 2 – Taxa de transição marginal e Função de sobrevivência marginal para os modelos Inversa Gaussiana (IG) e Gama | 61 |
| Tabela 3 – Resultados da simulação para o modelo geral ($EHR > 1$) | 68 |
| Tabela 4 – Resultados da simulação para o modelo restrito ($EHR = 1$) | 69 |
| Tabela 5 – Análise dos dados de câncer de cólon, assumindo os modelos de fragilidade PVF, Inverse Gaussian (IG) e Gamma. | 71 |
| Tabela 6 – Log-Verossimilhança, AIC e BIC para os modelos ajustados. | 72 |
| Tabela 7 – média dos EMVs, média do Erro padrão (MEP), Desvios padrão (DP), viés, raízes dos erros quadráticos médios (REQM) e probabilidades de cobertura (PC) dos parâmetros do novo modelo quando $\gamma = 0.3$ e 0.7 | 86 |
| Tabela 8 – Média dos EMVs, Média do Erro padrão (MEP), Desvios padrão (DP), viés, raízes dos erros quadráticos médios (REQM) e probabilidades de cobertura (PC) dos parâmetros do novo modelo quando $\gamma = -0.5$ e 0.5 | 87 |
| Tabela 9 – EMV dos parâmetros para o modelo de taxa de cura bivariada para conjunto de dados do banco, considerando os modelos de fragilidade Gamma, IG e PVF. | 91 |
| Tabela 10 – Critérios de comparação de modelos, Log-Verossimilhança, AIC e BIC, para os dados do banco | 92 |
| Tabela 11 – Estimativas de máxima verossimilhança da proporção de não abandono dos produtos 1 e 2 do sistema financeiro | 92 |
| Tabela 12 – Possíveis observações para dados de riscos semicompetitivos. | 106 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|--|
| 5-FU | Fluorouracil |
| AIC | Akaike information criterion |
| BCH | Bounded cumulative hazard |
| BIC | Bayesian information criterion |
| DP | Desvios padrão |
| EMV | Estimadores de máxima verossimilhança |
| EP | Erro padrão |
| i.i.d | independentes e identicamente distribuídas |
| IC(95%) | Intervalo de 95% de confiança |
| IG | Inversa Gaussiana |
| Lev | Levamisole |
| MEP | média do erro padrão |
| PC | probabilidades de cobertura |
| Pc | Poisson composta |
| PE | Positiva estável |
| PVF | Power variance function |
| REQM | raízes dos erros quadráticos médios |

LISTA DE SÍMBOLOS

$F(t)$ — Função de distribuição acumulada da variável aleatória T

$\mathbb{P}[\cdot]$ — Função de probabilidade

$S(t)$ — Função de sobrevivência da variável aleatória T

I_A — Função indicadora do conjunto A

$\mathcal{C}_\theta(\cdot)$ — Cópula de parâmetro θ

$\phi_\theta(\cdot)$ — Função geradora de cópula

$\lambda_0(t)$ — Função de risco de base

$\Lambda_0(\cdot)$ — Função de risco de base acumulada

$\mathbb{E}[Z]$ — Valor esperado da variável Z

$\mathcal{L}_Z(\cdot)$ — Transformada de Laplace associada á variável Z

$g(z)$ — Função de densidade da variável de fragilidade Z

$\psi_Z(\cdot)$ — Função geradora de probabilidade da variável Z

$\vartheta^*(t_1, t_2)$ — Medida de associação local entre os tempos T_1 e T_2

$\stackrel{D}{=}$ — tem a mesma distribuição que

p_0 — Proporção não susceptível ao evento de interesse

SUMÁRIO

| | | |
|---------|--|----|
| 1 | INTRODUÇÃO | 23 |
| 2 | CONCEITOS PRELIMINARES | 29 |
| 2.1 | Riscos competitivos | 29 |
| 2.2 | Risco semicompetitivos | 30 |
| 2.3 | Funções de cópula | 31 |
| 2.3.1 | <i>Cópula de Clayton</i> | 32 |
| 2.3.2 | <i>Cópula de Gumbel ou positiva estável</i> | 33 |
| 3 | MODELOS DE FRAGILIDADE | 35 |
| 3.1 | Modelo de fragilidade univariado | 35 |
| 3.2 | Modelo de fragilidade discreta | 38 |
| 3.3 | Modelo de fragilidade multivariada | 39 |
| 3.3.1 | <i>Medida de dependência local</i> | 41 |
| 3.4 | Principais modelos de fragilidade | 41 |
| 3.4.1 | <i>Modelo de fragilidade Gama</i> | 41 |
| 3.4.2 | <i>Modelo de fragilidade Inversa Gaussiana (IG)</i> | 43 |
| 3.4.3 | <i>Modelo de fragilidade Positiva estável (PE)</i> | 44 |
| 3.4.4 | <i>Modelo de fragilidade PVF: Versão (WIENKE, 2010)</i> | 46 |
| 3.4.5 | <i>Modelo de fragilidade PVF: Versão (HOUGAARD; HOUGAARD, 2000)</i> | 47 |
| 3.4.6 | <i>Modelo de fragilidade Poisson composta (Pc)</i> | 48 |
| 3.4.6.1 | <i>Modelo Pc a partir do modelo PVF: Versão (WIENKE, 2010)</i> | 48 |
| 3.4.6.2 | <i>Modelo Pc a partir do modelo PVF: Versão (HOUGAARD; HOUGAARD, 2000)</i> | 49 |
| 4 | MODELO DE FRAGILIDADE PARA DADOS DE RISCOS SEMI-COMPETITIVOS | 51 |
| 4.1 | Introdução | 51 |
| 4.2 | Modelo de riscos semicompetitivos | 53 |
| 4.2.1 | <i>Uma abordagem de fragilidade</i> | 54 |
| 4.2.2 | <i>O modelo de doença-morte com fragilidade compartilhada PVF</i> | 59 |
| 4.2.3 | <i>Função de densidade de probabilidade e dependência local</i> | 62 |
| 4.3 | Inferência para o novo modelo doença-morte | 64 |

| | | |
|-------|--|------------|
| 4.4 | Estudo de simulação | 66 |
| 4.5 | Aplicação: dados de câncer de cólon | 67 |
| 5 | MODELO DE SOBREVIVÊNCIA BIVARIADO INDUZIDO POR FRAGILIDADE DISCRETA | 77 |
| 5.1 | Introdução | 77 |
| 5.2 | Descrição do modelo | 78 |
| 5.3 | Inferência | 83 |
| 5.4 | Estudo de simulação | 84 |
| 5.5 | Aplicação: dados de churn de clientes brasileiros | 88 |
| 6 | CONSIDERAÇÕES FINAIS | 95 |
| 6.1 | Conclusões | 95 |
| 6.2 | Sugestões para pesquisa futura | 96 |
| | REFERÊNCIAS | 99 |
| | APÊNDICE A FUNÇÃO DE VEROSSIMILHANÇA | 105 |
| A.1 | Função densidade de probabilidade conjunta | 105 |
| A.2 | Função densidade de probabilidade na linha $T_1 = \infty$ | 105 |
| A.3 | Construção da função de verossimilhança | 106 |
| | APÊNDICE B PROBABILIDADE DE TRANSIÇÃO MARGINAL | 109 |
| | APÊNDICE C FUNÇÃO DE SOBREVIVÊNCIA CONJUNTA CONDICIONAL | 111 |
| C.1 | Função de sobrevivência conjunta condicional em ω | 111 |
| | APÊNDICE D FUNÇÃO DE VEROSSIMILHANÇA | 113 |
| D.1 | Função verossimilhança para dados completos | 113 |
| D.2 | Função verossimilhança para dados observados | 115 |
| D.2.1 | <i>Prova do Teorema 5.3.1</i> | 115 |

INTRODUÇÃO

Na análise de sobrevivência, o objetivo é modelar o tempo até a ocorrência de um evento de interesse, chamado de tempo de falha. Existem situações em que há mais de um tempo de falha e em muitos casos existe uma correlação não observável entre os tempos de falha de interesse. Por exemplo, para um caso bivariado, se os eventos de interesse são a recidiva de uma doença e a morte de um paciente, não podemos afirmar que esses eventos sejam independentes, já que a recidiva da doença pode influenciar no óbito do paciente. Além disso, fatores como estilo de vida, tabagismo, ocupação e fatores de risco genético podem influenciar na ocorrência dos eventos de interesse, tais fatores muitas vezes são desconhecidos e, portanto, não podem ser incluídos explicitamente na análise. Esses fatores variam em cada indivíduo o que implica que existe uma heterogeneidade não observada nos indivíduos da população. Nestes casos, uma possibilidade é usar modelos de fragilidade que se caracterizam pela introdução de uma variável aleatória não observada na modelagem a qual fornece informações sobre a variabilidade no tempo de vida e a correlação ou associação entre os tempos de eventos multivariados. Dessa forma, os modelos de fragilidade permitem a modelagem de risco em populações heterogêneas considerando efeitos aleatórios tanto de associação quanto de variabilidade, de modo que indivíduos com maior fragilidade ou vulnerabilidade tendem a apresentar o evento mais rapidamente do que outros. Estes efeitos aleatórios, denominados de fragilidade, representam as informações que não podem ou não foram observadas e também têm o objetivo de explicar a correlação entre os tempos multivariados. Os modelos de fragilidade foram inicialmente investigados por [Vaupel, Manton e Stallard \(1979\)](#), e seus trabalhos aplicados em bioestatística foram estendidos por [Hougaard e Hougaard \(2000\)](#), [Duchateau e Janssen \(2008\)](#), e [Wienke \(2010\)](#). Em muitas aplicações, os modelos de fragilidade assumem um modelo de risco proporcional condicionado à fragilidade, que atua de forma multiplicativa na função de risco de base (ou basal), assim, esta abordagem é considerada como uma extensão do conhecido modelo de Cox ([COX, 1972](#)), o qual assume implicitamente uma população homogênea a ser estudada.

Na literatura, as distribuições de fragilidade utilizadas em muitos estudos incluem as distribuições gama (VAUPEL; MANTON; STALLARD, 1979), inversa Gaussiana (IG) (HOU-GAARD, 1984) log-normal (SANTOS; DAVIES; FRANCIS, 1995) ou a positiva estável (PE) (HOUGAARD, 1986b). Um problema importante na área de modelos de fragilidade é a escolha da distribuição de fragilidade. O motivo das escolhas destas distribuições muitas vezes se deve à simplicidade analítica e computacional, como é o caso da distribuição gama, que pode ser conveniente para o estudo. Outra razão frequente se deve a propriedades específicas do modelo proposto, que podem ser atrativas, como é o caso da distribuição Positiva estável. No entanto, essas propriedades só são uma vantagem desde que sejam suportadas pelas informações dos dados, pois são os dados que devem orientar a escolha da distribuição de fragilidade e não uma propriedade atrativa do modelo matemático (DUCHATEAU; JANSSEN, 2008). A escolha da distribuição da fragilidade é de importância crucial para se chegar a uma boa descrição da estrutura de dependência presente nos dados além disso a fragilidade especificada incorretamente pode levar a resultados tendenciosos. Nos modelos de fragilidade, a dependência entre observações correlacionadas muda ao longo do tempo e a distribuição da fragilidade determina como a dependência se modifica (DUCHATEAU; JANSSEN, 2008). Uma alternativa para o problema da escolha do modelo de fragilidade seria escolher apenas uma família de distribuições de fragilidade e usá-la como modelo geral. Como apontado por Wienke (2010) uma família generalizada de distribuições de fragilidade que inclui as distribuições anteriores é a família de distribuição “power variance function” (PVF) introduzida por Tweedie (1984) e derivada independentemente por Hougaard (1986a). De acordo com Wienke (2010), por um lado, a família triparamétrica PVF com parâmetros $\mu > 0$, $\sigma > 0$ e $0 < \gamma \leq 1$, é facilmente tratável uma vez que se tem a transformada de Laplace de forma fechada e conseqüentemente não é difícil obter as funções de risco e sobrevivência. Por outro lado, o modelo é flexível no sentido de que contém os modelos de fragilidade gama, IG e PE como casos particulares e, portanto, pode ser usado para julgar se o ajuste desses modelos é satisfatório. A distribuição Poisson composta (Pc) foi introduzida por Aalen (1988) como uma distribuição de fragilidade. Essa distribuição pode ser construída como soma de variáveis aleatórias independente e identicamente com distribuição gama, em que o número de termos da soma é uma variável aleatória Poisson. Fazendo uma conveniente parametrização no modelo Pc Wienke (2010) mostra que sua transformada de Laplace é a mesma do modelo PVF com exceção do intervalo do parâmetro γ , que pode ser negativo no modelo Pc. Esta família de distribuições PVF é apresentada também em Hougaard e Hougaard (2000), onde a transformada de Laplace depende apenas do parâmetro de especificação de distribuição, γ , e conseqüentemente a variância corresponde a $1 - \gamma$, isto sob a suposição usada para evitar o problema de identificabilidade.

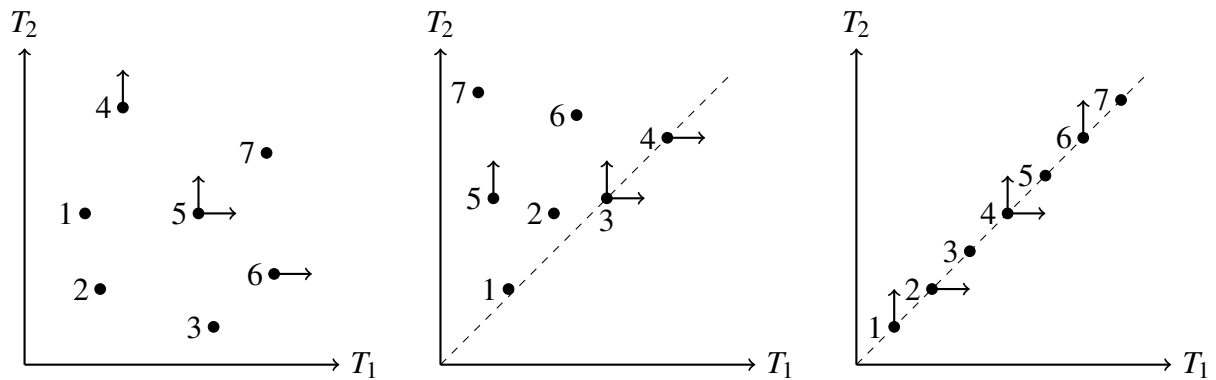


Figura 1 – A figura ilustra possíveis observações em dados bivariados (painel esquerdo), dados de riscos semicompetitivos (painel central) e dados de riscos concorrentes (painel direito). Um ponto indica que T_1 e T_2 foram observados e uma seta indica a censura do tempo de sobrevivência na direção apontada.

Na tese consideramos a análise de dados de sobrevivência bivariados com estruturas diferentes. A primeira, corresponde a dados com estrutura de riscos semicompetitivos introduzido inicialmente por [Fine, Jiang e Chappell \(2001\)](#), sendo considerada uma extensão de dados com estrutura de riscos competitivos como mostrado na Figura 1 (painel direito). Este tipo de dados são comuns em muitos estudos médicos, por exemplo, a morte pode ser o evento de interesse, o evento pode acontecer devido ao tratamento, devido à própria doença ou devido a outra causa. Desta forma, cada observação consiste em tempo de falha e causa de falha, ver por exemplo ([HOUGAARD, 2012](#)). No entanto, nos últimos anos, com o avanço da medicina, os tratamentos para vários tipos de câncer evoluíram bastante, o que ocasionou uma diminuição na mortalidade de pacientes e um aumento na proporção de pacientes com recidiva ou recorrência da doença. Desse modo, surge o interesse no estudo dos dois eventos: a morte e a recidiva, essa estrutura de dados é de riscos semicompetitivos como mostrada na Figura 1 (painel central). Mais especificamente, a estrutura de riscos semicompetitivos é caracterizada quando o interesse é o estudo da ocorrência de dois eventos, sendo um deles não terminal ou intermediário (recorrência) e outro terminal (óbito). Notar que o evento terminal impede a ocorrência do evento não terminal, porém, este último não impede a ocorrência do evento terminal. Assim, tem-se um par de variáveis aleatórias observadas no mesmo paciente, e é razoável considerar de uma estrutura de dependência entre as variáveis. Desta forma, modelamos os dados de riscos semicompetitivos através de processos de doença-morte com fragilidade compartilhada introduzidos por [Xu, Kalbfleisch e Tai \(2010\)](#), onde consideramos a família de distribuições PVF, proposta por [Wienke \(2010\)](#) e descrita na Seção 3.4.4, como nosso modelo de fragilidade a qual é incorporada para poder estudar a dependência não observada entre os tempos de falha e a heterogeneidade não observada dos indivíduos em estudo.

A segunda, dados bivariados com estrutura de longa duração ([BERKSON; GAGE, 1952; YAKOVLEV; TSODIKOV; ASSELAIN, 1996; TSODIKOV; IBRAHIM; YAKOVLEV, 2003; RODRIGUES *et al.*, 2009](#)), a qual é caracterizada por ter unidades amostrais com fator de risco

zero, implicado que existe uma proporção de unidades amostrais não suscetíveis ao evento de interesse. Em modelos de sobrevivência clássicos não toma em consideração essa proporção, mas nem sempre acontece que todas as unidades experimentais falharam em algum momento. Se consideramos, por exemplo, dados de clientes de uma determinada empresa, onde o evento de interesse é o cancelamento ou abandono de clientes, existe uma proporção de clientes que não abandonaram os produtos oferecidos pela empresa mesmo após um longo tempo. Isso ocorre porque para as empresas é importante manter seus clientes, pois pode ser mais caro atrair novos clientes do que fidelizá-los. Nestas situações é melhor considerar um modelo que leva em conta esta proporção de clientes fiéis à empresa. Na literatura, várias extensões de modelos de sobrevivência univariados de longa duração foram desenvolvidos na literatura. No entanto, pouca atenção tem sido dada à pesquisa em modelos de sobrevivência multivariada de longa duração. Nesta tese é proposto um modelo de sobrevivência bivariado de longa duração, cujo número de causas de risco para k diferentes tipos de eventos de interesse é modelado através de um modelo de fragilidade discreta Poisson($\rho_k \omega$), sendo $\rho > 0$ e ω uma variável de fragilidade PVF (proposta por Hougaard e Hougaard (2000) e descrita na Seção 3.4.5), sendo esta última a responsável de incorporar uma estrutura de dependência entre os k tipos de eventos.

Em ambos casos estudaremos as suas propriedades estatísticas e observaremos a flexibilidade do modelo, assim como a estrutura de dependência para cada submodelo (gama, IG e Pc). Consideramos a modelagem dos tempos de falha desde a perspectiva paramétrica, considerando as taxas de risco de base a partir da distribuição Weibull. Para a inferência, a estimativa de máxima verossimilhança é realizada para ajustar o modelo a um conjunto de dados. Primeiro, é fornecido um estudo de simulação para avaliar o desempenho do método de máxima verossimilhança na estimativa de parâmetros. Em seguida, veremos a aplicabilidade dos modelos primeiro ao conjunto de dados de câncer de cólon a qual tem uma estrutura de riscos semicompetitivos e segundo a um conjunto de dados de clientes de um sistema financeiro cuja estrutura de dados é do tipo longa duração.

O trabalho está organizado da seguinte forma: No Capítulo 2, é apresentado conceitos básicos onde descrevemos brevemente riscos competitivos e semicompetitivos. Algumas funções de copulas são apresentadas com fins de comparação aos modelos de fragilidade associados. No Capítulo 3 apresentamos uma revisão teórica de modelos de fragilidade univariados, discreta e multivariada onde incluímos também as principais distribuições usadas na literatura. No Capítulo 4 apresentamos o modelo de dados de riscos semicompetitivos vista como um processo de doença-morte onde a dependência é incorporada através modelos de fragilidade compartilhada entre as taxas de risco em cada indivíduo. Em seguida, na Seção 4.4 apresentamos um estudo de simulação para o modelo com estrutura de dados semicompetitivos, para avaliar o desempenho das estimativas de máxima verossimilhança dos parâmetros do modelo de doença-morte. A seguir, na Seção 4.5 é exibida a aplicação do modelo a um conjunto de dados reais de um ensaio clínico de câncer de cólon. No Capítulo 5, apresentamos o modelo bivariado de longa duração

induzida por fragilidade discreta onde é descrita as suas principais propriedades. Na Seção 5.3 é apresentado procedimentos inferenciais para o modelo proposto. Em seguida, um estudo de simulação é apresentado na Seção 5.4 e para finalizar o capítulo, na Seção 5.5 apresentamos a aplicação do modelos a um conjunto de dados de clientes de um sistema financeiro. Finalmente, Capítulo 6 são discutidas algumas conclusões sobre o desenvolvimento dos modelos de dados de risco semicompetitivos. Também são apresentados algumas propostas para trabalhos futuros.

CONCEITOS PRELIMINARES

Nesta seção apresentamos definições e ferramentas fundamentais para o desenvolvimento ao longo deste trabalho. Na Seção 2.1 são apresentadas os principais conceitos básicos em riscos competitivos. Em seguida, na Seção 2.2 é explicada a estrutura dos riscos semicompetitivos. Finalmente, na Seção 2.3 é feita uma revisão da definição de cópula, incluindo as cópulas Arquimedianas para dados de sobrevivência bivariados.

2.1 Riscos competitivos

Em riscos competitivos, observa-se o par (T, K) , onde T é o tempo de falha e K é a causa da falha. A estrutura básica de probabilidade é, portanto, uma distribuição bivariada na qual uma componente, K , é discreta e a outra, T , é contínua. Ao estudar o tempo até a falha de uma causa específica, as falhas de outras causas são eventos competitivos. Como pode haver muitas causas de falha, pode-se considerar a observação T como o menor de vários tempos de falha latentes T_1, T_2, \dots, T_N , isto é, $T = \min\{T_1, T_2, \dots, T_N\}$ e $K = j$ se $T = T_j$, com $j \in 1, 2, \dots, N$. A distribuição conjunta de (T, K) é especificada por *funções de sub-distribuição* ou *funções de incidência acumulada* (CROWDER, 2012), como também são chamadas,

$$F_j(t) = \mathbb{P}(T \leq t, K = j), \quad t > 0, \quad j \in \{1, \dots, N\} \quad (2.1)$$

Por diferenciação, as *funções de sub-densidade* tornam-se

$$f_j(t) = \frac{d}{dt} F_j(t)$$

A *distribuição marginal* de T é dada por

$$F(t) = \mathbb{P}(T \leq t) = \sum_{j=1}^N F_j(t)$$

Assim, a *função de sobrevivência* é dada por, $S(t) = 1 - F(t)$, e pode ser interpretada como a probabilidade de não ter falhado por qualquer causa até o tempo t .

A distribuição de (T, K) também pode ser especificada pela *funções de risco de causa específica*, que é dado por

$$\lambda_j(t) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, K = j | T > t]}{\Delta} = \frac{f_j(t)}{S(t)} \quad (2.2)$$

A interpretação da função de risco de causa específica é a taxa de falha de causa j no instante t na presença de outras causas de falha, considerando que o tempo de vida T é maior que t . A função geral de risco de T torna-se então

$$\lambda(t) = \sum_{j=1}^N \lambda_j(t)$$

Podemos escrever função de incidência cumulativa em termos do risco de causa específica, isto é,

$$\lambda(t) = \int_0^t \lambda_j(s) S(s) ds. \quad (2.3)$$

As funções de risco de causa específica acumulada são definidas como

$$\Lambda_j(t) = \int_0^t \lambda_j(s) ds.$$

Desta forma, usando este resultado, a função de risco acumulado de T é $\Lambda(t) = \sum_{j=1}^N \Lambda_j(t)$, logo temos a seguinte relação

$$S(t) = e^{-\Lambda(t)} = e^{-\sum_{j=1}^N \Lambda_j(t)} \quad (2.4)$$

2.2 Risco semicompetitivos

O problema dos riscos semicompetitivos foi introduzido pela primeira vez por [Fine, Jiang e Chappell \(2001\)](#) como uma variação de riscos competitivos. O problema refere-se a uma situação em que um indivíduo pode experimentar dois tipos de eventos, um não terminal e outro terminal. Assim, o evento terminal é considerado uma censura ao evento não terminal, isto é, o evento não terminal não pode ocorrer depois que o evento terminal ocorreu. No entanto, se o evento não terminal ocorreu primeiro, o evento terminal ainda pode ocorrer. Essa última afirmação difere no caso de riscos competitivos, pois, por exemplo, no caso bivariado, o primeiro tempo de falha que ocorre censura o outro (ver Figura 2). Um evento não terminal pode, por exemplo, ser recaída ou recorrência de alguma doença, enquanto um evento terminal pode ser morte. Um paciente que apresenta uma recaída após o tratamento pode morrer, enquanto um paciente que está morrendo não pode ter uma recaída após a morte. Uma censura independente

dos tempos de falha geralmente é incluída no problema de riscos semicompetitivos, nesta tese será considerado a censura à direita.

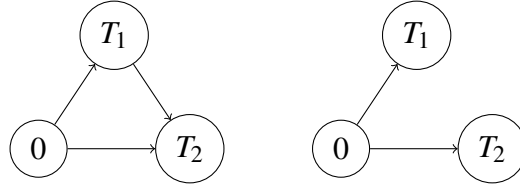


Figura 2 – Representação de riscos semicompetitivos (esquerda) e riscos competitivos (direita).

Sejam T_1 e T_2 , os tempos até a ocorrência dos eventos não terminal e terminal, respectivamente. Note que, para dados de risco semicompetitivos (T_1, T_2) podem ser observados sempre que T_1 é observado antes de T_2 , desta forma as observações estão restritas a $0 < t_1 \leq t_2$, assim como se mostra na figura 1 (painel central). Seja também C o tempo de censura para um indivíduo específico. A observação para cada indivíduo será $(Y_1, Y_2, \delta_1, \delta_2)$, o que pode ser descrito como

$$Y_2 = \min\{T_2, C\}, \quad \delta_2 = \begin{cases} 1, & T_2 \leq C \\ 0, & T_2 > C \end{cases},$$

$$Y_1 = \min\{T_1, Y_2\} \text{ e } \delta_1 = \begin{cases} 1, & T_1 \leq Y_2 \\ 0, & T_1 > Y_2 \end{cases},$$

em que Y_1 refere-se ao primeiro tempo de evento observado e δ_1 indica se este foi T_1 ou não. Y_2 refere-se ao tempo que foi observado após Y_1 que também pode ser o mesmo que Y_1 e δ_2 indica se T_2 foi censurado ou não. Existem 4 possíveis casos de observações que são apresentadas na Tabela 1.

Tabela 1 – Possíveis observações para dados de riscos semicompetitivos.

| Cenário | (Y_1, Y_2) | (δ_1, δ_2) |
|--|--------------|------------------------|
| Recorrência e censura antes da morte | (T_1, C) | $(1, 0)$ |
| Morte depois da recorrência | (T_1, T_2) | $(1, 1)$ |
| Morte sem recorrência | (T_2, T_2) | $(0, 1)$ |
| Censura antes da recorrência e a Morte | (C, C) | $(0, 0)$ |

2.3 Funções de cópula

Nesta seção apresentamos alguns conceitos básicos sobre cópulas e algumas funções de cópulas mais usadas na literatura. Informações complementares podem ser encontradas por exemplo em [Nelsen \(2007\)](#), [Trivedi, Zimmer *et al.* \(2007\)](#), [Durante e Sempi \(2010\)](#)

Definição 2.3.1 (Cópula). Uma cópula é uma distribuição multivariada cujas marginais seguem uma distribuição Uniforme(0, 1). Considere o vetor aleatório $\mathbf{U} = (U_1, \dots, U_n) \in \mathbb{I}^n$ com cópula

n -dimensional dada por

$$\mathcal{C}_\theta(u_1, \dots, u_n) = \mathbb{P}[U_1 \leq u_1, \dots, U_n \leq u_n; \theta], \quad (u_1, \dots, u_n) \in \mathbb{I}^n, \quad (2.5)$$

onde θ é o parâmetro associado a função cópula e $\mathbb{I} = [0, 1]$.

Teorema 2.3.2 (Teorema de Sklar). Seja F uma função de distribuição conjunta com marginais $F_1(t_1), \dots, F_n(t_n)$. Então, existe uma cópula n -dimensional tal que

$$F(t_1, \dots, t_n; \theta) = \mathcal{C}(F_1(t_1), \dots, F_n(t_n); \theta) \quad (2.6)$$

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

A prova deste teorema pode ser encontrado em [Nelsen \(2007\)](#). No Teorema 2.3.2 podemos ver que a cópula conecta as distribuições univariadas formando uma distribuição multivariada ou também que uma distribuição multivariada pode ser decomposta nas marginais univariadas e na estrutura de dependência dada pela cópula. Uma importante classe de cópulas conhecida como cópulas Arquimedianas encontram uma ampla gama de aplicações devido a facilidade com que elas podem ser construídas e a grande variedade de famílias de cópulas que pertencem a essa classe.

Definição 2.3.3 (Cópula Arquimediana). A cópula $\mathcal{C}_\theta(\cdot)$ é dita Arquimediana se

$$\mathcal{C}_\theta(u, v) = \phi_\theta^{-1}[\phi_\theta(u) + \phi_\theta(v)], \quad u, v \in [0, 1], \quad \theta \in \Theta.$$

Em que $\phi_\theta(\cdot) : [0, 1] \rightarrow [0, \infty)$ é uma função contínua, estritamente decrescente e convexa, tal que $\phi_\theta(1) = 0$, $\phi_\theta'(\cdot) < 0$ e $\phi_\theta''(\cdot) \geq 0$. A função $\phi_\theta(\cdot)$ é chamada de função geradora de cópula.

A seguir apresentamos as cópulas Arquimedianas bivariadas de Clayton e de Gumbel ou positiva estável.

2.3.1 Cópula de Clayton

A cópula de Clayton ([CLAYTON, 1978](#)) é uma cópula arquimediana simétrica com função geradora de cópula $\phi_\theta(t) = \theta^{-1}(t^{-\theta} - 1)$. Da Definição 2.3.3, a cópula de Clayton tem a forma

$$\mathcal{C}_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \quad \theta \in \mathbb{R}^+ \quad (2.7)$$

Veja que $\theta \rightarrow 0$ implica independência, isto é $\mathcal{C}_\theta(u, v) = uv$. Mas $\theta \rightarrow \infty$ indica dependência positiva. A cópula de Clayton não contempla dependência negativa. No contexto de análise de sobrevivência, suponha que T_1 e T_2 são tempos de falha e que suas respectivas funções de sobrevivência marginal são dadas por $u = S_1(t_1)$ e $v = S_2(t_2)$, então a função de sobrevivência conjunta é dada por

$$S_\theta(t_1, t_2) = (S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1)^{-\frac{1}{\theta}}. \quad (2.8)$$

2.3.2 Cópula de Gumbel ou positiva estável

A cópula positiva estável, também conhecida como família de Gumbel-Hougaard, foi considerada por Hougaard (1986a), Hougaard (1986b). Tem a forma

$$\mathcal{C}_\theta(u, v) = \exp \left\{ - \left[(-\log(u))^{-\theta} + (-\log(v))^{-\theta} \right]^\theta \right\}, \quad \theta \in (0, 1), \quad (2.9)$$

e função geradora dada por $\phi_\theta(t) = (-\log(t))^{-\theta}$. Veja que quando $\theta \rightarrow 1$ implica independência. Como a cópula de Clayton a cópula de Gumbel não permite dependência negativa, mas em contraste à de Clayton, apresenta uma forte dependência na cauda direita e dependência relativamente fraca na cauda esquerda. No contexto de análise de sobrevivência, suponha que T_1 e T_2 são tempos de falha e que suas respectivas funções de sobrevivência marginal são dadas por $u = S_1(t_1)$ e $v = S_2(t_2)$, então a função de sobrevivência conjunta é dada por

$$S_\theta(t_1, t_2) = \exp \left\{ - \left[(-\log(S_1(t_1)))^{-\theta} + (-\log(S_2(t_2)))^{-\theta} \right]^\theta \right\}, \quad \theta \in (0, 1), \quad (2.10)$$

Estas funções de cópula foram apresentadas com finalidades comparativas com os modelos de fragilidade apresentadas na Seção 3, a continuação.

MODELOS DE FRAGILIDADE

Iniciamos esta seção coletando alguns resultados cujo foco é a transformada de Laplace da fragilidade, uma vez que as funções de sobrevivência e de risco não condicional podem ser facilmente obtidas usando essa transformada. Desta forma, a função de verossimilhança também pode ser obtida através da transformada de Laplace. Esses resultados são apresentados em um contexto univariado e multivariado e também são válidos para todas as distribuições específicas que serão descritas no final desta seção.

3.1 Modelo de fragilidade univariado

Em análise de sobrevivência, tal como apontado por [Wienke \(2007, cap. 3\)](#), é impossível incluir todos os fatores de risco associados à ocorrência do evento de interesse, talvez porque tenhamos pouca ou nenhuma informação no nível individual. Esses fatores de risco desconhecidos levam a uma heterogeneidade na população em estudo, que pode ser descrita por uma variável que é chamada de fragilidade. O modelo de fragilidade clássico e aplicado com mais frequência assume uma estrutura de riscos proporcionais, condicionada ao efeito aleatório (fragilidade). Mais especificamente, a função de risco de um indivíduo depende de uma variável aleatória Z não observável e independente do tempo. Na estrutura de riscos multiplicativos que focaremos ao longo desta tese, Z atua multiplicativamente na função de risco de base, isto é

$$\lambda(t|Z) = Z\lambda_0(t). \quad (3.1)$$

Aqui, Z é considerado uma variável aleatória não negativa, $\lambda_0(t)$ é o risco de base comum a todos os indivíduos. Para garantir que o modelo seja identificável, as distribuições de fragilidade são padronizadas para $\mathbb{E}[Z] = 1$, se o valor esperado da distribuição de fragilidade existir. A variância $\text{Var}[Z] = \theta > 0$ (se existir) é interpretável como uma medida de heterogeneidade entre a população no risco de base ([WIENKE, 2010](#)). Quando θ é pequeno, os valores de Z estão

localizados perto de um. Se θ é grande, então os valores de Z são mais dispersos, induzindo maior heterogeneidade nos riscos individuais $Z\lambda_0(t)$. As covariáveis podem ser incorporadas ao modelo (3.1) da seguinte forma

$$\lambda(t|Z, \mathbf{x}) = Z\lambda_0(t) \exp(\boldsymbol{\varphi}'\mathbf{x}), \quad (3.2)$$

sendo $\mathbf{x} = (x_1, \dots, x_p)'$ e $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_p)'$ as covariáveis e parâmetros de regressão, respectivamente. Consequentemente, um modelo de fragilidade é uma generalização do conhecido modelo de riscos proporcionais, ou modelo de Cox (COX, 1972). O modelo de riscos proporcionais é obtido se a distribuição da fragilidade é degenerada em $Z = 1$ para todos os indivíduos (WIENKE, 2010). Note que, de modo interpretativo, indivíduos com maior fragilidade podem ser considerados mais frágeis e, portanto, espera-se que eles experimentem o evento de interesse mais precocemente do que outros indivíduos semelhantes em relação às covariáveis observadas. Por motivos de simplicidade, os resultados abaixo são restritos ao modelo (3.1). Seja $S(t|Z)$ a função de sobrevivência de um indivíduo condicional à fragilidade Z . É fácil mostrar que

$$S(t|Z) = e^{-\int_0^t \lambda(u|Z) du} = e^{-Z\Lambda_0(t)} \quad (3.3)$$

onde $\Lambda_0(t) = \int_0^t \lambda_0(u) du$, denota a função de risco de base acumulada. Desde que Z é uma variável aleatória latente, os dados para este modelo condicional (3.3) (ou individual) não são observáveis, portanto, é necessário considerar o modelo marginal (ou populacional), em que o termo fragilidade está integrado. A função de sobrevivência da população é obtida a partir da função de sobrevivência condicional $S(t|Z)$, integrando a fragilidade, isto é,

$$S(t) = \int_0^\infty S(t|Z)g(z)dz = \int_0^\infty e^{-Z\Lambda_0(t)}g(z)dz = \mathbb{E}[e^{-Z\Lambda_0(t)}],$$

em que $g(z)$ é a função densidade da fragilidade Z . Note que a última expressão é a transformada de Laplace de Z , $\mathcal{L}_Z(s) = \mathbb{E}[e^{-Zs}]$, avaliado em $s = \Lambda_0(t)$, de modo que podemos escrever

$$S(t) = \mathcal{L}_Z(\Lambda_0(t)). \quad (3.4)$$

A conexão com a transformada de Laplace foi primeiramente apontada e explorada por Hougaard (1984), Hougaard (1986b). A transformada de Laplace tem um papel importante no estudo de modelos de fragilidade. As derivadas da transformada de Laplace podem ser usadas para obter resultados gerais sobre a distribuição de sobrevivência não condicional. A derivada r -ésima da transformada de Laplace é dada por

$$\mathcal{L}_Z^{(r)}(s) = (-1)^r \mathbb{E}[Z^r e^{-Zs}], \quad (3.5)$$

e portanto o r -ésimo momento da fragilidade é dado por

$$\mathbb{E}[Z^r] = (-1)^r \mathcal{L}_Z^{(r)}(0). \quad (3.6)$$

A função densidade do tempo de falha é dada por

$$f(t) = -\frac{dS(t)}{dt} = -\lambda_0(t)\mathcal{L}'_Z(\Lambda_0(t)). \quad (3.7)$$

A função de risco é então obtida como a razão entre a densidade e a função de sobrevivência

$$\lambda(t) = -\lambda_0(t)\frac{\mathcal{L}'_Z(\Lambda_0(t))}{\mathcal{L}_Z(\Lambda_0(t))} \quad (3.8)$$

De (3.6), a esperança e a variância também podem ser caracterizadas pela transformada de Laplace da fragilidade

$$\mathbb{E}[Z] = -\mathcal{L}'(0) \quad (3.9)$$

$$\text{Var}[Z] = \mathcal{L}''(0) - (-\mathcal{L}'(0))^2 \quad (3.10)$$

assumindo a existência das expressões anteriores. Por exemplo, no caso de uma distribuição positiva estável, os momentos não existem. Aqui, \mathcal{L}' e \mathcal{L}'' denotam a primeira e a segunda derivada da transformada de Laplace.

A distribuição da fragilidade descreve a fragilidade da população no início do seguimento. Supõe-se que a fragilidade seja fixa para cada indivíduo ao longo do tempo, mas a composição da população muda com o passar do tempo. Em média, os indivíduos mais frágeis experimentam o evento de interesse em um tempo relativamente menor. Devido a este fato, a distribuição da fragilidade na população de risco muda ao longo do tempo. O teorema a seguir aponta para esse fato e estabelece a ligação entre o modelo condicional e o não condicional.

Teorema 3.1.1 (Vaupel, Manton e Stallard (1979)). Assuma o modelo de fragilidade dada por (3.1). a função de risco populacional, $\lambda(t) = \frac{f(t)}{S(t)}$ é geralmente $\lambda(t) = \mathbb{E}[\lambda(t|Z)|T > t]$, ou mais especificamente,

$$\begin{aligned} \lambda(t) &= \int_0^\infty \lambda(t|Z)g(z|T > t)dz \\ &= \lambda_0(t) \int_0^\infty zg(z|T > t)dz \end{aligned} \quad (3.11)$$

onde $g(z|T > t)$ representa a função densidade de fragilidade dos sobreviventes por um período de tempo t .

O risco da população é assim interpretado como a média ponderada dos riscos individuais entre os sobreviventes (ver (3.11)). Os pesos são determinados pela distribuição da fragilidade. Indivíduos frágeis com altos valores de Z tendem a morrer primeiro. A prova deste teorema não será apresentada neste trabalho mas pode ser encontrada em Wienke (2010, cap. 3). No entanto, apresentamos a função densidade de fragilidade dos sobreviventes até o tempo t , que é dada por

$$g(z|T > t) = \frac{S(t|Z)g(z)}{S(t)} = \frac{\exp\{-Z\Lambda_0(t)\}g(z)}{\mathcal{L}_Z(\Lambda_0(t))} \quad (3.12)$$

Da mesma forma, usando (3.7), podemos derivar a densidade de fragilidade para indivíduos que morrem no momento t

$$g(z|T=t) = \frac{f(t|Z)g(z)}{f(t)} = \frac{-Z \exp\{-Z\Lambda_0(t)\} g(z)}{\mathcal{L}'_Z(\Lambda_0(t))}, \quad (3.13)$$

onde $f(t|Z) = -\frac{dS(t|Z)}{dt}$ é a função densidade condicional do tempo de falha. Considerando covariáveis no estudo, sob o modelo (3.2), a função de risco não condicional pode ser expressa como a esperança da função de risco condicionada a estar em risco no instante t e o vetor de covariáveis \mathbf{x} , isto é,

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\boldsymbol{\varphi}'\mathbf{x}) \mathbb{E}[Z|T > t, \mathbf{x}]. \quad (3.14)$$

3.2 Modelo de fragilidade discreta

Em geral as distribuições de fragilidade contínua não permitem assumir risco zero, indicando que existe um subgrupo de indivíduos não suscetíveis ao evento de interesse mesmo após um longo período de observação, isto devido a que a variável não assume valor igual a zero. A fragilidade Pc, é útil para a modelagem de dados de sobrevivência que contenha uma proporção de unidades experimentais para os quais o evento de interesse não ocorre mesmo após um período longo de observação (WIENKE, 2010). Outra abordagem, são os modelos de fragilidade discreta, cuja massa de probabilidade é definida nos inteiros não negativos $\{0, 1, \dots\}$.

Suponha que a distribuição de probabilidade de Z é especificada por $\mathbb{P}[Z = z] = p_z$, para $z = 0, 1, \dots$ com $\sum p_z = 1$ a função de sobrevivência não condicional pode ser obtida somando na Equação (3.3) sobre o suporte da distribuição de Z . Assim, o função de sobrevivência não condicional é dado por

$$S(t) = \sum_{z=0}^{\infty} S(t|Z) p_z = \sum_{z=0}^{\infty} [S_0(t)]^z p_z = \mathbb{E}[[S_0(t)]^z] = \psi_Z(S_0(t)), \quad (3.15)$$

onde $\psi_Z(\cdot)$ é a função geradora de probabilidade da variável de fragilidade Z e $S_0(t) = \exp(-\Lambda_0(t))$ é a função de sobrevivência de base. O modelo em (3.15), é o mesmo modelo de sobrevivência com fração de cura obtido nos cenários de riscos competitivos proposto por Tsodikov, Ibrahim e Yakovlev (2003) e Rodrigues *et al.* (2009). Desde que $\mathbb{P}[Z = 0] > 0$ a função de sobrevivência em (3.15) é impropria, isto é

$$\lim_{t \rightarrow \infty} S(t) = \mathbb{P}[Z = 0] = \psi_Z(0) = p_0 > 0 \quad (3.16)$$

Este resultado significa que após um longo período de tempo existe uma probabilidade $p_0 > 0$ de que uma unidade experimental não seja susceptível ao evento de interesse.

3.3 Modelo de fragilidade multivariada

Na análise de dados de sobrevivência, implicitamente, a maioria dos modelos e métodos estatísticos, especialmente o modelo de riscos proporcionais de Cox, foram desenvolvidos sob a suposição de que as observações de diferentes sujeitos são estatisticamente independentes umas das outras. Embora isso seja sensato em muitas aplicações, tornou-se óbvio que essa suposição não se aplica a certas situações (WIENKE, 2010). Por exemplo, algumas vezes os dados de sobrevivência são observados em grupos de indivíduos, assim como, indivíduos da mesma família, pacientes tratados no mesmo hospital ou dentes restaurados de um mesmo indivíduo. Em tais situações, é esperado que os tempos observados dentro de cada grupo apresentem certas semelhanças que não seriam observadas entre os diferentes grupos. Para um único indivíduo podemos observar o mesmo evento várias vezes. O número de repetições da medida pode ser aleatória em um período de acompanhamento (eventos recorrentes) ou previamente prefixada (medidas repetidas). Para uma mesma unidade experimental podemos observar mais de um evento de interesse. Nas situações dadas anteriormente, é razoável supor a existência de associação entre os tempos de sobrevivência, os quais caracterizam os dados de sobrevivência multivariados. A abordagem mais popular são os modelos de fragilidade em que um ou mais efeitos aleatórios são incluídos na função de risco, para levar em conta a dependência entre as observações.

A primeira abordagem importante é o conceito de fragilidade compartilhada. Neste modelo a função risco para cada indivíduo no mesmo grupo, é exatamente o mesmo que no modelo padrão de fragilidade univariada, isto é

$$\lambda(t|Z_k) = Z_k \lambda_0(t). \quad (3.17)$$

onde Z_k é a fragilidade comum para todos os indivíduos no grupo k e $\lambda_0(t)$ é a função de risco de base comum para todos os indivíduos. Este modelo induz correlação entre os tempos de eventos dos indivíduos no mesmo grupo, isto é, o modelo assume que todos os tempos de eventos no mesmo grupo k são independentes dadas as variáveis de fragilidade Z_k . Em outras palavras, trata-se de um modelo de independência condicional onde a fragilidade Z_k é comum a todos os indivíduos de um grupo k e, portanto, responsável por criar dependência entre os tempos dos eventos. Se $\text{Var}(Z_k) = 0$ isso implica independência entre os tempos de eventos no grupo k ; caso contrário, há dependência positiva eles (WIENKE, 2010). Por outro lado, assume-se que há independência entre as observações de diferentes grupos. Para mais detalhes veja o livro de Duchateau e Janssen (2008, Cap. 4)

Suponha que $j = 1, 2, \dots, n_k$ em que n_k é o número de indivíduos no grupo k , logo a função de sobrevivência condicional conjunta (dado Z_k) para o grupo k é dado por

$$S(t_1, \dots, t_{n_k}|Z_k) = \exp\left(-Z_k \sum_{j=1}^{n_k} \Lambda_0(t_j)\right), \quad (3.18)$$

com $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ é a função de risco de base acumulada. O valor da fragilidade Z_k é constante ao longo do tempo. A função de sobrevivência conjunta não condicional é obtida a partir da função de sobrevivência condicional conjunta integrando a fragilidade em relação à distribuição da fragilidade, isto é

$$S(t_1, \dots, t_{n_k}) = \int_0^\infty \exp\left(-Z_k \sum_{j=1}^{n_k} \Lambda_0(t_j)\right) g(z_k) dz_k = \mathcal{L}_{Z_k}\left(\sum_{j=1}^{n_k} \Lambda_0(t_j)\right), \quad (3.19)$$

onde $g(z_k)$ e $\mathcal{L}_{Z_k}(\cdot)$ são a função densidade de probabilidade e a transformada de Laplace associada à fragilidade Z_k do grupo k . Usando a propriedade da transformada de Laplace em (3.5) e a função de sobrevivência conjunta (3.19), podemos obter a função densidade conjunta para um grupo k de tamanho n_k , isto é

$$\begin{aligned} f(t_1, \dots, t_{n_k}) &= (-1)^{n_k} \frac{\partial^{n_k}}{\partial t_1 \dots \partial t_{n_k}} S(t_1, \dots, t_{n_k}) \\ &= (-1)^{n_k} \prod_{j=1}^{n_k} \lambda_0(t_j) \mathcal{L}_{Z_k}^{(n_k)}\left(\sum_{j=1}^{n_k} \Lambda_0(t_j)\right) \end{aligned} \quad (3.20)$$

Supondo que os tempos de sobrevivência estão sujeitos a censura à direita e não informativa, a função de verossimilhança completa para o k -ésimo grupo de tamanho n_k , é dada por

$$\left(\prod_j^{n_k} [\lambda_0(t_{kj})]^{\delta_{kj}}\right) Z_k^{d_k} \exp\left\{-Z_k \sum_{j=1}^{n_k} \Lambda_0(t_j)\right\} \quad (3.21)$$

onde $d_k = \sum_{j=1}^{n_k} \delta_{kj}$ e δ_{kj} é o indicador de censura do j -ésimo indivíduo no k -ésimo grupo. A função de verossimilhança marginal para o k -ésimo grupo resulta integrando a verossimilhança completa em relação a fragilidade, isto é

$$\left(\prod_j^{n_k} [\lambda_0(t_{kj})]^{\delta_{kj}}\right) (-1)^{d_k} \mathcal{L}_{Z_k}^{(d_k)}\left(\sum_{j=1}^{n_k} \Lambda_0(t_j)\right) \quad (3.22)$$

As funções de sobrevivência não condicionais univariadas também são de interesse, pois mostram como a função de sobrevivência evolui ao longo do tempo para um indivíduo escolhido aleatoriamente, sem levar em consideração o agrupamento dos dados. Obviamente, como os indivíduos mais frágeis morrem mais cedo, a população muda automaticamente com o tempo, e isso tem um impacto importante na função de sobrevivência da população (DUCHATEAU; JANSSEN, 2008). A função de sobrevivência não condicionais do j -ésimo indivíduo no grupo k pode ser expressa em termos da transformada de Laplace (WIENKE, 2010), isto é

$$S(t_{jk}) = \mathcal{L}_{Z_k}(\Lambda_0(t_{jk})), \quad (3.23)$$

Alternativamente, a função de sobrevivência conjunta também pode ser dada em termos das funções de sobrevivência marginal dos indivíduos no grupo k de tamanho n_k . Logo, de (3.4)

segue a relação

$$\Lambda_0(t_{jk}) = \mathcal{L}_{Z_k}^{-1}(S_j(t_{jk})), \quad j = 1, \dots, n_k. \quad (3.24)$$

A função de sobrevivência conjunta não condicional (3.19) pode ser reescrita como

$$S(t_1, \dots, t_{n_k}) = \mathcal{L}_{Z_k} \left(\mathcal{L}_{Z_k}^{-1}(S_1(t_1)) + \dots + \mathcal{L}_{Z_k}^{-1}(S_{n_k}(t_{n_k})) \right). \quad (3.25)$$

Esta é a representação de cópula da função de sobrevivência conjunta com base no modelo de fragilidade, uma vez que vincula a função de sobrevivência conjunta às funções de sobrevivência marginais. Observe que as funções de sobrevivência das marginais nesta representação de cópula (3.25) são derivadas do modelo de fragilidade e, portanto, são uma função da distribuição de fragilidade (ver (3.4)). Nos modelos de cópula discutidos na Seção 2.3, as funções de sobrevivência marginais também são estimadas marginalmente, o que leva a uma função de sobrevivência conjunta diferente da função de sobrevivência conjunta obtida a partir do modelo de fragilidade Duchateau e Janssen (2008, Cap. 4).

3.3.1 Medida de dependência local

Uma medida de associação local entre os tempos T_1 e T_2 foi introduzido por Clayton (1978) e sugerido por Oakes (1982), que denotamos por $\vartheta^*(t_1, t_2)$ e é definida como

$$\vartheta^*(t_1, t_2) = \frac{\lambda(t_2|T_1 = t_1)}{\lambda(t_2|T_1 > t_1)} = \frac{S(t_1, t_2) \frac{\partial^2}{\partial t_1 \partial t_2} S(t_1, t_2)}{\left(\frac{\partial}{\partial t_1} S(t_1, t_2) \right) \left(\frac{\partial}{\partial t_2} S(t_1, t_2) \right)} \quad (3.26)$$

Esta medida nos permite investigar como a dependência muda com o tempo. De acordo com Clayton (1978), $\vartheta^*(t_1, t_2) = 1$, implica independência e $\vartheta^*(t_1, t_2) > 1$ associação positiva entre os tempos T_1 e T_2 .

3.4 Principais modelos de fragilidade

Nesta seção descrevemos as principais distribuições utilizadas em modelos de fragilidade como Gama, Inversa Gaussiana (IG), Positiva Estável (PE) e Poisson Composta. Essas distribuições por sua vez pertencem a uma família de distribuições denominada PVF (Power Variance Function) que também será descrita nesta seção na versão proposta por (HOUGAARD; HOUGAARD, 2000) e na versão proposta por (WIENKE, 2010).

3.4.1 Modelo de fragilidade Gama

A distribuição gama tem sido amplamente aplicada como modelo de fragilidade, por exemplo, Vaupel, Manton e Stallard (1979), Santos, Davies e Francis (1995), Hougaard e

Hougaard (2000), Duchateau e Janssen (2008), Hougaard (2012). É fácil derivar de forma fechada as expressões de sobrevivência não condicional, densidade e função de risco, devido à simplicidade da transformação de Laplace. Essa também é a razão pela qual esta distribuição foi usada na maioria das aplicações publicadas até o momento. A densidade de uma variável aleatória Z com distribuição gama é dada por:

$$g(z) = \frac{1}{\Gamma(\kappa)} \sigma^\kappa z^{\kappa-1} e^{-\sigma z}, \quad z > 0 \quad (3.27)$$

com parâmetro de forma $\kappa > 0$ e parâmetro de escala $\sigma > 0$. A transformada de Laplace é

$$\mathcal{L}_Z(s) = \left(1 + \frac{s}{\sigma}\right)^{-\kappa}. \quad (3.28)$$

Usando (3.9) e (3.10) e considerando o Laplaciano em (3.28) pode ser mostrado que $\mathbb{E}[Z] = \frac{\kappa}{\sigma}$ e $\text{Var}[Z] = \frac{\kappa}{\sigma^2}$. Para garantir que o modelo seja identificável, a restrição $\kappa = \sigma$ é usada para a distribuição gama, o que resulta em $\mathbb{E}[Z] = 1$. Denote por $\theta := \frac{1}{\sigma}$ a variância da variável fragilidade. Assim $Z \sim \Gamma(1/\theta, 1/\theta)$, desta forma o correspondente Laplaciano da fragilidade pode ser reescrito como

$$\mathcal{L}_Z(s) = (1 + \theta s)^{-\frac{1}{\theta}}. \quad (3.29)$$

Isso implica que a função sobrevivência e a função de risco não condicional no modelo de fragilidade gama são dados por

$$S(t) = (1 + \theta \Lambda_0(t))^{-\frac{1}{\theta}}. \quad (3.30)$$

e

$$\lambda(t) = \frac{\lambda_0(t)}{1 + \theta \Lambda_0(t)} \quad (3.31)$$

A densidade da distribuição de fragilidade entre os sobreviventes no momento t pode ser escrita a partir de (3.12), na seguinte forma

$$g(z|T > t) = \frac{\left(\frac{1}{\theta} + \Lambda_0(t)\right)^{\frac{1}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)} z^{\frac{1}{\theta}-1} \exp\left\{-z \left(\frac{1}{\theta} + \Lambda_0(t)\right)\right\} \quad (3.32)$$

que é a densidade de uma variável de fragilidade com distribuição gama com parâmetro de forma $\frac{1}{\theta}$ e parâmetro de escala $\frac{1}{\theta} + \Lambda_0(t)$. A media e variância da fragilidade entre os sobreviventes no momento t , são dadas por $\mathbb{E}[Z|T > t] = (1 + \theta \Lambda_0(t))^{-1}$ e $\text{Var}[Z|T > t] = \theta(1 + \theta)(1 + \theta \Lambda_0(t))^{-2}$, respectivamente. Para mais detalhe veja Wienke (2010, Seção 3.3)

A função de sobrevivência conjunta pode ser reescrita como uma cópula, considerando o caso bivariado, (T_1, T_2) , desde que $\mathcal{L}_Z^{-1}(s) = \theta^{-1}(s^{-\theta} - 1)$, logo a partir de (3.25), temos que

$$S(t_1, t_2) = \left((S_1(t_1))^{-\theta} + (S_2(t_2))^{-\theta} - 1 \right)^{-\frac{1}{\theta}}, \quad (3.33)$$

onde $S_1(t_1)$ e $S_2(t_2)$, são as funções de sobrevivência marginal dos tempos de falha T_1 e T_2 , respectivamente. Essa expressão é a cópula Arquimediana que corresponde a uma distribuição de fragilidade gama e é chamada cópula de Clayton (CLAYTON, 1978), que é dada em (2.8). Embora a forma funcional da cópula em (2.8) e (3.33) seja a mesma, há uma diferença na maneira como as funções de sobrevivência da população são modeladas (DUCHATEAU; JANSSEN, 2008).

Por outro lado, considerando a fragilidade gama, a medida de dependência local $\vartheta^*(t_1, t_2)$ definida em (3.26), não depende de (t_1, t_2) , ou seja, a dependência é constante ao longo do tempo, isto é

$$\vartheta^*(t_1, t_2) = 1 + \theta \quad (3.34)$$

Então, para valores grandes do parâmetro da fragilidade, maior é a associação entre T_1 e T_2 . Para maior informação veja Duchateau e Janssen (2008, Seção 4.2).

3.4.2 Modelo de fragilidade Inversa Gaussiana (IG)

A distribuição Inversa Gaussiana (IG) foi introduzida como uma alternativa à distribuição gama por Hougaard (1984). Semelhante ao modelo de fragilidade gama, existem expressões simples de forma fechada para as funções incondicionais de sobrevivência e risco, o que torna o modelo atraente. A função densidade de probabilidade de uma variável aleatória distribuída gaussiana inversa com parâmetros $\mu > 0$ e $\sigma > 0$ é dada por

$$g(z) = \sqrt{\frac{\sigma}{2\pi z^3}} \exp\left(-\frac{\sigma}{2\mu^2 z}(z - \mu)^2\right), \quad z > 0 \quad (3.35)$$

A transformada de Laplace é dada por

$$\mathcal{L}_Z(s) = \exp\left\{\frac{\sigma}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 s}{\sigma}}\right)\right\} \quad (3.36)$$

Logo de (3.9) e (3.10) obtemos $\mathbb{E}[Z] = \mu$ e $\text{Var}[Z] = \frac{\mu^3}{\sigma} := \theta$. Tomando $\mathbb{E}[Z] = \mu = 1$ e $\text{Var}[Z] = \frac{1}{\sigma}$ resulta na seguinte transformada Laplace simplificada:

$$\mathcal{L}_Z(s) = \exp\left\{\frac{1}{\theta} \left(1 - \sqrt{1 + 2\theta s}\right)\right\} \quad (3.37)$$

Isso implica que a função sobrevivência e a função de risco não condicional no modelo de fragilidade IG são dados por

$$S(t) = \exp\left\{\frac{1}{\theta} \left(1 - \sqrt{1 + 2\theta \Lambda_0(t)}\right)\right\} \quad (3.38)$$

e

$$\lambda(t) = \frac{\lambda_0(t)}{\sqrt{1 + 2\theta \Lambda_0(t)}} \quad (3.39)$$

A densidade da distribuição de fragilidade entre os sobreviventes no instante t pode ser escrita a partir de (3.12), na seguinte forma

$$g(z|T > t) = \frac{1}{\sqrt{2\pi^2 z^3}} \exp\left(-\frac{\left(z - (1 + 2\theta\Lambda_0(t))^{-\frac{1}{2}}\right)^2}{\frac{2\theta z}{1 + 2\theta z\Lambda_0(t)}}\right) \quad (3.40)$$

que é a densidade de uma variável de fragilidade com distribuição IG com uma media e variância da fragilidade entre os sobreviventes no instante t , dado por $\mathbb{E}[Z|T > t] = (1 + 2\theta\Lambda_0(t))^{-\frac{1}{2}}$ e $\text{Var}[Z|T > t] = \theta(1 + 2\theta\Lambda_0(t))^{-2}$, respectivamente. Para mais detalhe veja [Wienke \(2010, Seção 3.5\)](#).

A função de sobrevivência conjunta pode ser reescrita como uma cópula, considerando o caso bivariado, (T_1, T_2) , desde que $\mathcal{L}_Z^{-1}(s) = \frac{\log s}{2}(\theta \log s - 2)$, logo a partir de (3.25), temos que

$$\begin{aligned} S(t_1, t_2) &= \mathcal{L}_Z(\mathcal{L}_Z^{-1}(S_1(t_1)) + \mathcal{L}_Z^{-1}(S_2(t_2))) \\ &= \exp\left\{\frac{1}{\theta} - \left[\frac{1}{\theta^2} + \log S_1(t_1)(\log S_1(t_1) - \frac{2}{\theta}) + \log S_2(t_2)(\log S_2(t_2) - \frac{2}{\theta})\right]^{1/2}\right\}, \end{aligned} \quad (3.41)$$

onde $S_1(t_1)$ e $S_2(t_2)$, são as funções de sobrevivência marginal dos tempos de falha T_1 e T_2 , respectivamente. Note que esta cópula pertence à família de cópulas Arquimedianas com função geradora de cópulas $\phi_\theta(s) = \mathcal{L}_Z^{-1}(s) = \frac{\log s}{2}(\theta \log s - 2)$, $\theta > 0$.

A medida de dependência local $\vartheta^*(t_1, t_2)$ em (3.26), pode ser obtido usando (3.41), isto é,

$$\vartheta^*(t_1, t_2) = 1 + \frac{\theta}{1 - \theta \log S(t_1, t_2)} \quad (3.42)$$

O valor mais alto é obtido no início e é igual a $1 + \theta$ e vai para 1 quando a função de sobrevivência vai para zero. Para maior informação veja [Duchateau e Janssen \(2008, Seção 4.3\)](#).

3.4.3 Modelo de fragilidade Positiva estável (PE)

O uso de distribuições positivas estáveis na análise de dados de sobrevivência multivariada remonta a ([HOUGAARD, 1986a](#)). Uma distribuição é positiva estável se a soma normalizada de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d) tiver a mesma distribuição que um fator de escala multiplicado por uma única variável aleatória. Em termos matemáticos, uma distribuição estável tem a propriedade de que, para Z_1, \dots, Z_n variáveis aleatórias independentes e identicamente distribuídas e para cada n existe uma constante normalizadora $c(n)$ de modo que

$$c(n)Z_1 \stackrel{D}{=} Z_1 + \dots + Z_n,$$

onde $\stackrel{D}{=}$ denota “tem a mesma distribuição que”. A constante $c(n)$ toma a forma $n^{\frac{1}{\gamma}}$ onde o parâmetro γ é do intervalo $(0, 2]$ e denominado expoente característico. Para $\gamma = 2$, a distribuição normal é obtida. Para garantir uma distribuição nos números positivos, restringimos aqui o caso em que $\gamma \in (0, 1)$. A função densidade de probabilidade de uma variável aleatória uniparamétrica com distribuição estável positiva é dada por [Feller \(1971\)](#)

$$g(z; \gamma) = \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\Gamma(k\gamma + 1)}{k!} z^{-k\gamma-1} \sin(k\gamma\pi), \quad z \geq 0, \quad 0 < \gamma \leq 1. \quad (3.43)$$

Essa expressão é uma série de potências, convergindo rapidamente para grandes valores de z e para valores pequenos de z . No caso especial de $\gamma = 1$, a distribuição da fragilidade se degenera na massa pontual $Z = 1$ ([WIENKE, 2010](#)). Embora a função densidade de probabilidade possa ser representada apenas por séries infinitas, a transformada de Laplace tem uma forma muito simples e é dado por

$$\mathcal{L}_Z(s) = \exp\{-s^\gamma\} \quad (3.44)$$

Todos os momentos dessa distribuição são infinitos. Consequentemente, a esperança da fragilidade é infinita e a variância não existe. Como apontado por [Duchateau e Janssen \(2008\)](#), este fato tornou esta distribuição atrativa, pois, apenas funções de densidade com média infinita têm a propriedade de que o parâmetro de heterogeneidade é independente da informação da covariável como mostrado por [Hougaard \(1986a\)](#). Usando a transformada de Laplace (3.44), a função de sobrevivência e a função de risco não condicional são dadas por

$$S(t) = \exp\{-(\Lambda_0(t))^\gamma\} \quad (3.45)$$

e

$$\lambda(t) = \gamma\lambda_0(t)(\Lambda_0(t))^{\gamma-1} \quad (3.46)$$

A função de sobrevivência conjunta bivariada de (T_1, T_2) , pode ser reescrita como uma cópula, desde que $\mathcal{L}_Z^{-1}(s) = -(\log s)^{1/\gamma}$. Logo, a partir de (3.25), temos que

$$S(t_1, t_2) = \exp\left\{-\left[(-\log(S_1(t_1)))^{1/\gamma} + (-\log(S_2(t_2)))^{1/\gamma}\right]^\gamma\right\} \quad (3.47)$$

Essa expressão é a cópula Arquimediana que corresponde a uma distribuição de fragilidade positiva estável e é chamada cópula de Gumbel, que é dada em (2.10).

Por outro lado, a medida de dependência local $\vartheta^*(t_1, t_2)$ definida em (3.26), para o modelo PE, é dada por

$$\vartheta^*(t_1, t_2) = 1 + \frac{\gamma - 1}{\gamma \log S(t_1, t_2)} \quad (3.48)$$

Para maior informação veja o livro de [Duchateau e Janssen \(2008, Seção 4.4\)](#).

3.4.4 Modelo de fragilidade PVF: Versão (WIENKE, 2010)

Nesta seção apresentamos a família das distribuições PVF (Power variance function) na versão apresentada por Wienke (2010) no livro intitulado “Frailty models in survival analysis”. Esta família de distribuições que inclui as distribuições Gama, IG e PE foi sugerida por Tweedie (1984) e derivada independentemente por Hougaard (1986b). A família triparamétrica denotada por $PVF(\gamma, \mu, \sigma)$ com $\mu > 0$, $\sigma > 0$ e $0 < \gamma \leq 1$ tem função densidade de probabilidade dada por

$$g(z) = e^{-\sigma(1-\gamma)\left(\frac{z}{\mu} - \frac{1}{\gamma}\right)} \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\sigma(1-\gamma))^{k(1-\gamma)} \mu^{k\gamma} \Gamma(k\gamma+1)}{\gamma^k k!} z^{-k\gamma-1} \sin(k\gamma\pi). \quad (3.49)$$

A transformada de Laplace é dada por

$$\mathcal{L}_Z(s) = \exp \left\{ \frac{\sigma(1-\gamma)}{\gamma} \left[1 - \left(1 + \frac{\mu s}{\sigma(1-\gamma)} \right)^\gamma \right] \right\}. \quad (3.50)$$

A esperança e variância da variável aleatória Z com distribuição PVF são obtidas com a ajuda da primeira e da segunda derivadas da transformação de Laplace, isto é, $\mathbb{E}[Z] = \mu$ e $\text{Var}[Z] = \mu^2/\sigma$, respectivamente. A distribuição PVF se reduz à distribuição gama quando $\gamma \rightarrow 0$ e à distribuição IG quando $\gamma = 0.5$. A distribuição PE pode ser obtida como um caso especial da distribuição PVF. Para mostrar esse fato, são necessárias algumas considerações assintóticas. Reescrevemos o logaritmo da transformada de Laplace (3.50), para mais detalhes veja Wienke (2010). Para o caso em que $\gamma < 0$, a distribuição PC pode ser obtida, isto será visto na Seção 3.4.6.

Para garantir a identificabilidade do modelo, assumimos $\mathbb{E}[Z] = \mu = 1$ e considerando $\text{Var}(Z) = \frac{1}{\sigma} := \theta$, a transformada de Laplace (3.50) se torna

$$\mathcal{L}_Z(s) = \exp \left\{ \frac{(1-\gamma)}{\theta\gamma} \left[1 - \left(1 + \frac{\theta s}{(1-\gamma)} \right)^\gamma \right] \right\}. \quad (3.51)$$

Isso implica que a função sobrevivência e a função de risco não condicional no modelo de fragilidade PVF são dados respectivamente por

$$S(t) = \exp \left\{ \frac{(1-\gamma)}{\theta\gamma} \left[1 - \left(1 + \frac{\theta\Lambda_0(t)}{(1-\gamma)} \right)^\gamma \right] \right\} \quad (3.52)$$

e

$$\lambda(t) = \lambda_0(t) \left(1 + \frac{\theta\Lambda_0(t)}{\gamma-1} \right)^{1-\gamma}. \quad (3.53)$$

Para calcular a densidade da distribuição de fragilidade entre os sobreviventes num momento específico, restringimos novamente ao caso $\mu = 1$ e eliminamos as covariáveis observadas para manter as fórmulas simples:

$$g(z|T > t) = e^{\frac{-(1-\gamma)}{\theta A^\gamma} \left(\frac{z}{A^{\gamma-1}} - \frac{1}{\gamma} \right)} \quad (3.54)$$

$$\times \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(1-\gamma)^{k(1-\gamma)} A^{k\gamma(\gamma-1)}}{(\theta A^\gamma)^{k(1-\gamma)} \gamma^k} \frac{\Gamma(k\gamma+1)}{k!} z^{-k\gamma-1} \sin(k\gamma\pi). \quad (3.55)$$

que é a densidade de uma variável de fragilidade com distribuição PVF com parâmetros $A = 1 + \frac{\theta\Lambda_0(t)}{1-\gamma}$ (em vez de μ), $\theta A^{-\gamma}$ (em vez de θ) e γ .

A função de sobrevivência conjunta pode ser reescrita como uma cópula, considerando o caso bivariado, (T_1, T_2) . A partir de (3.25), temos que

$$S(t_1, t_2) = \exp \left\{ \frac{(1-\gamma)}{\theta\gamma} \left[1 - \left(b_1^{1/\gamma} + b_2^{1/\gamma} - 1 \right)^\gamma \right] \right\}, \quad (3.56)$$

onde $b_i = 1 + \theta \left(1 - \frac{1}{1-\gamma} \right) \log(S_i(t_i))$, $i = 1, 2$. A medida de dependência local $\vartheta^*(t_1, t_2)$ em (3.26), pode ser obtida usando (3.56), isto é,

$$\vartheta^*(t_1, t_2) = 1 + \theta \left(1 + \theta \left(1 - \frac{1}{1-\gamma} \right) \log(S(t_1, t_2)) \right)^{-1} \quad (3.57)$$

3.4.5 Modelo de fragilidade PVF: Versão (HOUGAARD; HOUGAARD, 2000)

Nesta seção apresentamos a família das distribuições PVF (Power variance function) na versão apresentada por Hougaard e Hougaard (2000) no livro intitulado “Analysis of multivariate survival data”. Nesta versão da família de distribuições PVF denotado por $PVF(\gamma, \mu, \sigma)$, a variância da variável aleatória é uma função de potência da sua média. Sob essa parametrização, a família de distribuições PVF inclui como casos especiais as distribuições gamma para $\gamma = 0$, PE para $\sigma = 0$ e $\mu = \gamma$ e IG para $\gamma = 0.5$, entre outros. O espaço paramétrico da distribuição PVF consiste em $\gamma \leq 1, \mu > 0$ com $\sigma \geq 0$ para $\gamma > 0$ e $\sigma > 0$ para $\gamma \leq 0$.

Para $\gamma \geq 0$ a distribuição PVF está concentrada nos números positivos e sua função de densidade é dada por (veja Hougaard e Hougaard (2000, Pág. 504))

$$g(z) = e^{-\sigma z + \frac{\mu\sigma^\gamma}{\gamma}} \left(-\frac{1}{\pi z} \right) \sum_{k=1}^{\infty} \frac{\Gamma(k\gamma + 1)}{k!} \left(-z^{-\gamma} \frac{\mu}{\gamma} \right)^k \sin(\gamma k\pi), \quad y > 0. \quad (3.58)$$

Para $\gamma < 0$ a distribuição PVF está concentrada nos números positivos ou zero, isto implica que alguns grupos têm risco zero. O termo Γ na densidade não é necessariamente definido e, portanto, podemos usar a expressão alternativa para a função densidade da PVF como (veja Hougaard e Hougaard (2000, Pág. 504))

$$g(z) = e^{-\sigma z + \frac{\mu\sigma^\gamma}{\gamma}} \frac{1}{z} \sum_{k=1}^{\infty} \frac{(-\mu z^\gamma / \gamma)^k}{k! \Gamma(-k\gamma)}, \quad y > 0. \quad (3.59)$$

Essa expressão é válida para todos os valores de γ , exceto 0 e 1, com a convenção de que quando a função Γ no denominador é indefinida (o que acontece quando $k\gamma$ é um inteiro positivo), o termo inteiro na soma é zero.

A transformada de Laplace é dada por

$$\mathcal{L}_Z(s) = \exp \left\{ \frac{-\mu}{\gamma} [(\sigma + s)^\gamma - \sigma^\gamma] \right\}. \quad (3.60)$$

Seja $Z \sim \text{PVF}(\gamma, \mu, \sigma)$, onde $\gamma \leq 1$, $\mu > 0$ e $\sigma \geq 0$, então

$$\mathbb{E}[Z] = \mu \sigma^{\gamma-1} \text{ e } \text{Var}[Z] = \mu(1 - \gamma) \sigma^{\gamma-2}.$$

Quando $\sigma > 0$, todos os momentos positivos da distribuição existem. Para evitar problema de identificabilidade, assumimos que $\mu = \sigma = 1$ assim $\mathbb{E}[Z] = 1$ e $\text{Var}(Z) = 1 - \gamma$. Logo a transformada de Laplace da distribuição de PVF com essa restrição é

$$\mathcal{L}_Z(s) = \exp \left\{ -\frac{1}{\gamma} [(1 + s)^\gamma - 1] \right\}. \quad (3.61)$$

Este modelo estende os modelos IG, PE e, portanto, é útil para testar qualquer um dos modelos. Também pode ser usado como uma maneira flexível de descrever a dependência.

3.4.6 Modelo de fragilidade Poisson composta (Pc)

A distribuição composta de Poisson foi introduzida por [Aalen \(1988\)](#), [Aalen et al. \(1992\)](#) como uma distribuição de fragilidade. Uma propriedade interessante do modelo é que ele permite um subgrupo de fragilidade zero, que nunca experimenta o evento de interesse. Esse modelo é relevante para a medicina e a demografia, por exemplo, ao considerar a mortalidade por causa específica ou a ocorrência de uma doença. Apesar do fato de que a densidade da parte contínua é dada apenas como uma série infinita que deve ser calculada numericamente, a distribuição é matematicamente conveniente. Também pode ser visto como uma escolha natural. A distribuição pode ser construída da seguinte forma: A fragilidade Z dos indivíduos da população em estudo é

$$Z = \begin{cases} X_1 + X_2 + \dots + X_N, & \text{se } N > 0; \\ 0, & \text{se } N = 0 \end{cases} \quad (3.62)$$

onde $N \sim \text{Poisson}(\rho)$ com $\rho > 0$ e as variáveis X_1, X_2, \dots são i.i.d. com $X_i \sim \Gamma(\kappa, \nu)$, $\kappa > 0$ e $\nu > 0$. Além disso, N é tomado para ser independente de X_1, X_2, \dots . Usando o laplaciano da distribuição gama $\mathcal{L}_X(s) = (1 + \frac{s}{\nu})^{-\kappa}$ e a função geradora de probabilidades da distribuição Poisson, $\psi_N(s) = e^{-\rho(1-e^{-s})}$, pode ser mostrado que

$$\mathcal{L}_Z(s) = \exp \left\{ -\rho \left[1 - \left(1 + \frac{s}{\nu} \right)^{-\kappa} \right] \right\}. \quad (3.63)$$

3.4.6.1 Modelo Pc a partir do modelo PVF: Versão ([WIENKE, 2010](#))

Fazendo a parametrização $\rho = -\frac{\kappa \nu^\gamma}{\gamma}$, $\nu = \nu$ e $\kappa = -\gamma$, o Laplaciano é dado por

$$\mathcal{L}_Z(s) = \exp \left\{ -\frac{\kappa \nu^\gamma}{\gamma} \left[1 - \left(1 + \frac{s}{\nu} \right)^\gamma \right] \right\}. \quad (3.64)$$

A partir de (3.64) pode se mostrar que $\mathbb{E}[Z] = \kappa v^{\gamma-1}$ e $\text{Var}[Z] = \kappa(1-\gamma)v^{\gamma-2}$. Logo, para $\mathbb{E}[Z] = 1$ e $\text{Var}[Z] = \theta = \frac{1-\gamma}{v}$ o Laplaciano da distribuição Pc se torna em (3.51), que é igual ao Laplaciano da distribuição PVF. Portanto, como as transformadas de Laplace da distribuição PVF e a distribuição Pc são iguais (exceto para a faixa do parâmetro γ , que pode ser negativo no modelo Pc), as propriedades obtidas da transformação de Laplace são as mesmas. Consequentemente, a função de sobrevivência da população e as funções de risco do modelo de fragilidade Pc coincidem com as respectivas funções no modelo de fragilidade do PVF. Assim, a função de risco para a distribuição Pc, $\lambda(t)$, corresponde à expressão (3.53). Deve-se notar que a integral de (3.53) em $[0, \infty)$ é finita se $\gamma < 0$. Consequentemente, a função de sobrevivência dada em (3.52) é imprópria dado que uma fração dos indivíduos tem fragilidade Z e nunca experimentarão o evento em estudo. O tamanho da proporção não susceptível ao evento de interesse, p_0 , é dado por

$$p_0 = \lim_{t \rightarrow \infty} S(t) = \lim_{t \rightarrow \infty} \exp \left\{ \frac{(1-\gamma)}{\theta\gamma} \left(1 - \left(1 + \frac{\theta\Lambda_0(t)}{(1-\gamma)} \right)^\gamma \right) \right\} = \exp \left\{ \frac{(1-\gamma)}{\theta\gamma} \right\} \quad (3.65)$$

3.4.6.2 Modelo Pc a partir do modelo PVF: Versão (HOUGAARD; HOUGAARD, 2000)

Quando $\gamma < 0$ em (3.60) a distribuição é também chamada de Poisson composta. Considerando a parametrização $\kappa = -\gamma$, $v = \sigma$ e $\rho = -\frac{\mu\sigma^\gamma}{\gamma}$ em (3.63) podemos obter a transformada de Laplace da distribuição PVF dada em (3.60). Aqui, para $\gamma < 0$ há uma probabilidade p_0 da variável aleatória ser zero, e é dada por

$$p_0 = \exp \left(\frac{\mu\sigma^\gamma}{\gamma} \right)$$

MODELO DE FRAGILIDADE PARA DADOS DE RISCOS SEMICOMPETITIVOS

4.1 Introdução

Em algumas situações, os sujeitos em estudo podem vivenciar um dos vários eventos denominados ‘terminais’, e onde a ocorrência de um evento impossibilita a ocorrência posterior de qualquer outro (HANEUSE; LEE, 2016). Esta é a estrutura clássica de “riscos competitivos” (LAU; COLE; GANGE, 2009), onde o tempo até o evento de falha e a causa da falha/indicador de censura são registrados para cada observação. No entanto, em muitos ensaios clínicos de câncer, a ocorrência de um evento intermediário não terminal, como a recorrência do tumor, é frequentemente o evento de interesse. Essa recorrência do tumor (evento não terminal) pode não impedir a morte do sujeito (evento terminal), com o evento terminal censurando dependentemente o evento não terminal, mas não vice-versa. Este é o cenário de riscos semicompetitivos (FINE; JIANG; CHAPPELL, 2001).

Para fins motivacionais, considere nosso conjunto de dados gerado a partir de um ensaio clínico randomizado, multicêntrico e controlado simultaneamente (MOERTEL *et al.*, 1990; MOERTEL *et al.*, 1995) para determinar a eficácia de dois regimes de terapia adjuvante (somente levamisol e levamisol mais fluorouracil) na melhoria das taxas de cura cirúrgica para o câncer de cólon em estágio III. Do ponto de vista clínico, é necessária uma avaliação adequada da distribuição de sobrevida de pacientes com recorrência tumoral intermediária após um regime de tratamento específico pós-resseção para uma previsão informada do risco de recorrência para outros pacientes e estratégias de tratamento adaptativas subsequentes. No entanto, isso é um desafio, dada a forte correlação positiva (veja a Figura 3, painel esquerdo) entre T_1 (tempo até a recorrência do tumor) e T_2 (tempo até a morte), com $T_1 \leq T_2$ (dados observados apenas acima da diagonal), o que torna inválidas as suposições típicas de censura independente em análise de sobrevivência (OAKES, 1993) e riscos competitivos (consistindo apenas em eventos terminais).

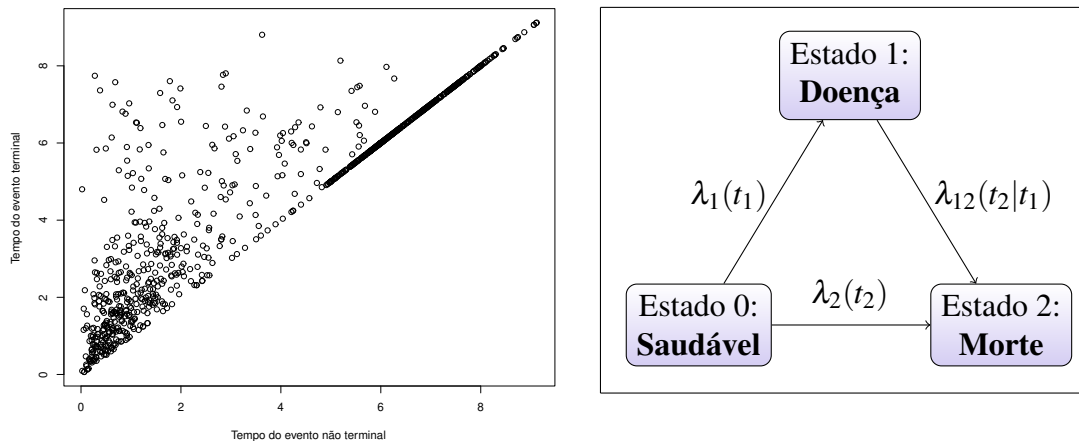


Figura 3 – Diagrama de dispersão de T_1 e T_2 para o conjunto de dados de câncer de cólon (panel esquerdo). Modelo de Doença-Morte para dados com estrutura de riscos semicompetitivos (panel direito).

Existem duas abordagens principais para analisar dados de riscos semicompetitivos. A primeira abordagem considera uma distribuição conjunta para (T_1, T_2) especificada por meio de um modelo de cópula na parte superior da diagonal. Dentro da estrutura de cópula, Fine, Jiang e Chappell (2001) considerou uma cópula de Clayton (CLAYTON, 1978), Wang (2003) considerou uma configuração de cópula mais geral, enquanto Lakhali, Rivest e Abdous (2008) consideraram uma cópula Arquimediana para estimar o parâmetro de dependência. A formulação relacionada considerando os efeitos variantes no tempo de um tratamento na distribuição marginal de um evento não terminal foi considerada em Peng e Fine (2007) e Hsieh e Huang (2012). Além dos modelos de transformação semiparamétricos (CHEN, 2012), abordagens mais recentes (ZHOU *et al.*, 2016) abordam a censura dependente selecionando primeiro um modelo de cópula adequado por meio de uma abordagem de diagnóstico exploratória (BANDEEN-ROCHE *et al.*, 2005) e em seguida, desenvolver um procedimento de inferência para estimar simultaneamente a função de sobrevivência marginal da recorrência do câncer e um parâmetro de associação no modelo de cópula. Recentemente, Peng, Xiang e Wang (2018) consideraram uma extensão semiparamétrica para lidar com dados de riscos semicompetitivos agrupados. Esta formulação de variável latente através de uma estrutura conjunta (cópula) leva a uma interpretação hipotética da distribuição marginal dos eventos não terminais, e também complica a análise de covariância (XU; KALBFLEISCH; TAI, 2010). A segunda abordagem traz o problema de riscos semicompetitivos na estrutura clássica do compartimento doença-morte (HANEUSE; LEE, 2016), que, na realidade, é um caso especial de modelos multiestado (ANDERSEN; KEIDING, 2002). O conceito de avaliação de riscos semicompetitivos se encaixa naturalmente no paradigma doença-morte bem estabelecido (XU; KALBFLEISCH; TAI, 2010), onde um paciente pode transitar para o evento terminal diretamente ou via evento não terminal, com o modelo completamente especificado através das funções de intensidade de transição para as três transições distintas. Aqui, a dependência entre T_1 e T_2 é introduzida por meio de uma fragilidade compartilhada ou termo de efeitos aleatórios. Enquanto Xu, Kalbfleisch e Tai (2010) consideraram uma fragilidade

Gamma, Jiang e Haneuse (2017) estenderam isso para uma classe de modelos de fragilidade de transformação, permitindo uma gama mais ampla de possíveis distribuições de fragilidade. Dentro desse quadro doença-morte, existem outros clássicos (KIM; KIM; KIM, 2019; HA *et al.*, 2020; LEE; GILSANZ; HANEUSE, 2021) e Bayesianos (HAN *et al.*, 2014; LEE *et al.*, 2015; LEE *et al.*, 2016) sendo estas formulações baseadas em fragilidade de riscos semicompetitivos que lidam com várias complicações de dados, como censura de intervalo, falha intermediária, redução de viés, etc.

Nesta tese, estudamos dados de riscos semicompetitivos sob a formulação de doença-morte com fragilidade compartilhada. Consideramos como nossa escolha de fragilidade a distribuição PVF, proposta por Wienke (2010), a qual é descrita na Seção 3.4.4. Esta distribuição é facilmente tratável devido às expressões de forma fechada da função de sobrevivência marginal e taxa de risco marginal. Além disso, é um modelo flexível no sentido de que contém as densidades gama, IG e PE como casos específicos. Desta forma ao considerar a família de fragilidade PVF podemos verificar se o ajuste dos submodelos (que são membros da família) é satisfatório.

4.2 Modelo de riscos semicompetitivos

Como mencionamos anteriormente, dados de riscos semicompetitivos podem ser convenientemente modelados usando modelos de doença-morte, que é um modelo multi-estado (MEIRA-MACHADO *et al.*, 2009). No contexto de nosso estudo do câncer de cólon, ele pode modelar uma situação na qual o indivíduo pode estar num estado inicial (após a ressecção do tumor) “Saudável” (estado 0) do qual ele pode transitar diretamente para o estado de “morte” (estado 2) ou pode transitar para um “estado de recorrência da doença” (estado 1) primeiro e depois para estado de “morte” (estado 2). Essa situação pode ser observada na Figura 3 (painel direito). O modelo de doença-morte é completamente especificado pelas funções de transição ou risco para as três transições possíveis: risco de causa específica da recorrência da doença, $\lambda_1(t_1)$; risco de causa específica da morte, $\lambda_2(t_2)$; e risco de morte condicional a um período de recorrência da doença, $\lambda_{12}(t_2|t_1)$, $0 < t_1 < t_2$. Como em Xu, Kalbfleisch e Tai (2010), as taxas de transição são definidas da seguinte forma:

Definição 4.2.1. As taxas de risco ou de transição entre os estados 0 (saudável), 1 (recorrência) e 2 (morte) são definido por

$$\lambda_1(t_1) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_1 \in [t_1, t_1 + \Delta) | T_1 \geq t_1, T_2 \geq t_1]}{\Delta}, \quad t_1 > 0 \quad (4.1)$$

$$\lambda_2(t_2) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta) | T_1 \geq t_2, T_2 \geq t_2]}{\Delta}, \quad t_2 > 0 \quad (4.2)$$

$$\lambda_{12}(t_2|t_1) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta) | T_1 = t_1, T_2 \geq t_2]}{\Delta}, \quad 0 < t_1 \leq t_2. \quad (4.3)$$

Na Definição 4.2.1, $\lambda_1(t_1)$ e $\lambda_2(t_2)$ representam as taxas de transição do estado 0 (saudável) para o estado 1 (recorrência) e taxa de transição do estado 0 (saudável) para o estado 2

(morte), respectivamente, $\lambda_{12}(t_2|t_1)$ representa a taxa de transição do estado 1 (recorrência) para o estado 2 (morte). Observe que a taxa de transição para o evento terminal após a ocorrência do evento não terminal no instante, $T_1 = t_1$, $\lambda_{12}(t_2|t_1)$, em geral pode depender de t_1 e t_2 . No entanto, em um modelo de Markov, $\lambda_{12}(t_2|t_1)$, depende apenas de t_2 , devido à propriedade de Markov, que faz com que $\lambda_{12}(t_2|t_1) = \lambda_{12}(t_2)$. O modelo de Markov é usado com mais frequência devido à sua simplicidade (MEIRA-MACHADO *et al.*, 2009).

As taxas (4.1)-(4.3), desempenham um papel muito importante pois definem a distribuição conjunta no quadrante superior do suporte de (T_1, T_2) . De acordo com Xu, Kalbfleisch e Tai (2010), para o modelo de probabilidade de (T_1, T_2) considere uma função de densidade conjunta $f(t_1, t_2)$ absolutamente contínua no quadrante superior $t_2 \geq t_1$, de modo que

$$\int_0^{\infty} \int_{t_1}^{\infty} f(t_1, t_2) dt_2 dt_1 = \mathbb{P}[T_1 < \infty] \leq 1$$

Se um sujeito experimenta o evento terminal antes do evento não terminal acontecer, definimos $T_1 = \infty$. Assim, o evento não terminal está sempre definido mesmo que o evento terminal tenha acontecido antes. Desta forma, o complemento da probabilidade é distribuído ao longo da linha em $t_1 = \infty$ com densidade contínua $f_{\infty}(t_2)$, $t_2 > 0$ e de modo que

$$\int_0^{\infty} \int_{t_1}^{\infty} f(t_1, t_2) dt_2 dt_1 + \int_0^{\infty} f_{\infty}(t_2) dt_2 = 1$$

Note que, nesta formulação não está sendo especificada qualquer probabilidade positiva no quadrante $t_2 < t_1 < \infty$, o que é razoável dado o processo considerado.

Observe que o evento não terminal, que pode ser a recorrência da doença, pode ocorrer antes do evento terminal, como a morte, mas não vice-versa. Então, é claro que a recorrência da doença pode influenciar na ocorrência da morte do paciente observado, ou seja, um sujeito com recorrência da doença pode morrer mais rápido do que outro sem recorrência. Isto significa que existe uma dependência não observada entre os tempos T_1 e T_2 associados a esses eventos. Uma maneira de incorporar uma estrutura de dependência entre os eventos não terminal e terminal é usar o modelo de fragilidade compartilhada (ou efeito aleatório) descrita na Seção 3.3 e que pode ser encontrada em Wienke (2010), no qual os tempos T_1 e T_2 são considerados condicionalmente independentes (dada a variável fragilidade). Na seguinte seção é descrito o modelo de fragilidade no contexto de modelos de doença-morte.

4.2.1 Uma abordagem de fragilidade

Seja $Z > 0$ uma variável aleatória que denota a fragilidade compartilhada com $\mathbb{E}[Z] = 1$. Definimos funções de transição condicionais análogas a $\lambda_1(t_1)$, $\lambda_2(t_2)$ e $\lambda_{12}(t_2|t_1)$ da seguinte

forma:

$$\lambda_1(t_1|Z) = Z\lambda_{01}(t_1), \quad t_1 > 0, \quad (4.4)$$

$$\lambda_2(t_2|Z) = Z\lambda_{02}(t_2), \quad t_2 > 0, \quad (4.5)$$

$$\lambda_{12}(t_2|t_1, Z) = Z\lambda_{03}(t_2), \quad 0 < t_1 \leq t_2, \quad (4.6)$$

em que $\lambda_{0i}(\cdot)$, $i = 1, 2, 3$ são as funções de risco base (ou basal), que podem ser consideradas de forma paramétrica ou não paramétrica, e Z é o efeito aleatório ou fragilidade de um indivíduo específico. Neste modelo, as três taxas de transição condicionais em cada indivíduo têm a mesma fragilidade. No entanto, dado que os indivíduos sob observação não têm a mesma predisposição para a ocorrência do evento de interesse (heterogeneidade), a fragilidade é diferente para cada sujeito.

A taxa de transição condicional para o evento terminal, dado que ocorreu um evento não terminal, $\lambda_{12}(t_2|t_1, Z)$, em (4.6), é assumida como um modelo de Markov, ou seja, $\lambda_{03}(\cdot)$, que não depende de t_1 . Deste modo, a *razão de risco explicativo condicional* (CLAYTON, 1978; XU; KALBFLEISCH; TAI, 2010), que é uma caracterização da dependência entre T_1 e T_2 , é dada por

$$\mathbf{EHR} = \frac{\lambda_{12}(t_2|t_1, Z)}{\lambda_2(t_2|Z)} = \frac{\lambda_{03}(t_2)}{\lambda_{02}(t_2)}, \quad t_2 > t_1.$$

Assim, para $\lambda_{02}(t_2) \neq \lambda_{03}(t_2)$ (modelo geral), a recorrência da doença afeta o risco de morte. Portanto, a dependência de T_2 em T_1 é descrita tanto pela razão de risco explicativa condicional (dado Z), $\lambda_{03}(t_2)/\lambda_{02}(t_2)$, bem como pela fragilidade Z , através do seu parâmetro desconhecido, que denotamos por $\theta > 0$. Quando $\lambda_{02}(t_2) = \lambda_{03}(t_2)$, (modelo restrito), a dependência de T_2 e T_1 é completamente especificada por Z .

A partir de (4.4)-(4.6) é possível encontrar as respectivas funções de sobrevivência condicionais. No contexto do modelo de Doença-Morte, $S(t|Z)$ pode ser interpretada como “a probabilidade de permanecer no estado 0 até o tempo t , dada a fragilidade Z ”. Da mesma forma como em riscos competitivos, equivalente a (2.4), podemos derivar a função de sobrevivência condicional geral, $S(t|Z)$, para riscos semicompetitivos no modelo de doença-morte, usando as taxas de transição condicionais definidas em (4.4) e (4.5), isto é, para $0 < t_1 \leq t_2$

$$\begin{aligned} S(t|Z) &= \mathbb{P}[T_1 \geq t, T_2 \geq t|Z] \\ &= \mathbb{P}[\text{Permanecer no estado 0 até o tempo } t|Z] \\ &= e^{-Z(\Lambda_{01}(t)+\Lambda_{02}(t))}, \end{aligned} \quad (4.7)$$

Se existe uma transição para o estado 1, a probabilidade condicional de permanecer no

estado 1 até o tempo t_2 é dado por

$$\begin{aligned}
S_{12}(t_2|t_1, Z) &= \mathbb{P}[T_2 \geq t_2 | T_1 = t_1, Z] \\
&= e^{-Z \int_{t_1}^{t_2} \lambda_{03}(s) ds} \\
&= e^{-Z \left(\int_0^{t_2} \lambda_{03}(s) ds - \int_0^{t_1} \lambda_{03}(s) ds \right)} \\
&= e^{-Z \Lambda_{03}(t_1, t_2)}, \tag{4.8}
\end{aligned}$$

onde $\Lambda_{0i}(t) = \int_0^t \lambda_{0i}(s) ds$ para $i = 1, 2, 3$, são as funções de risco de base acumulado e $\Lambda_{03}(u, t) = \Lambda_{03}(t) - \Lambda_{03}(u)$. Os resultados (4.7) e (4.8) serão úteis para o estudo de simulação, apresentados posteriormente na Seção 4.4.

Por outro lado, a função de sobrevivência bivariada condicional, dado Z , pode ser obtida da seguinte forma.

$$\begin{aligned}
S(t_1, t_2 | Z) &= \mathbb{P}[T_1 \geq t_1, T_2 \geq t_2 | Z] = \mathbb{P}[T_2 \geq t_2 | T_1 \geq t_1, Z] \mathbb{P}[T_1 \geq t_1 | Z] \\
&= \mathbb{P}[T_2 \geq t_1 | Z] \mathbb{P}[T_2 \geq t_2 | T_1 = t_1, Z] \mathbb{P}[T_1 \geq t_1 | Z]
\end{aligned}$$

Observe que, a $\mathbb{P}[T_2 \geq t_2 | T_1 \geq t_1, Z]$ é diferente se $t_1 \leq t_2$ ou $t_1 > t_2$, este último associado com a transição direta para o estado do evento terminal, sem ocorrência do evento intermediário não terminal). Assim, pode-se verificar que,

$$S(t_1, t_2 | Z) = \begin{cases} \mathbb{P}[T_2 \geq t_1 | Z] \mathbb{P}[T_2 \geq t_2, | T_1 = t_1, Z] \mathbb{P}[T_1 \geq t_1 | Z], & t_1 \leq t_2 \\ \mathbb{P}[T_2 \geq t_2 | T_1 = \infty, Z] \mathbb{P}[T_1 \geq t_1 | Z], & t_1 > t_2 \end{cases}$$

Note que, no caso $t_1 > t_2$, $\mathbb{P}[T_2 \geq t_2 | T_1 = \infty, Z] = \mathbb{P}[T_2 \geq t_2 | Z]$. Logo,

$$S(t_1, t_2 | Z) = \begin{cases} S_2(t_1 | Z) S_{12}(t_2 | t_1, Z) S_1(t_1 | Z), & t_1 \leq t_2 \\ S_2(t_2 | Z) S_1(t_1 | Z), & t_1 > t_2 \end{cases}$$

Assim, considerando as funções de transição (4.4)-(4.6), temos que

$$S(t_1, t_2 | Z) = \begin{cases} e^{-Z \left(\int_0^{t_1} \lambda_{02}(s) ds \right)} e^{-Z \left(\int_{t_1}^{t_2} \lambda_{03}(s) ds \right)} e^{-Z \left(\int_0^{t_1} \lambda_{01}(s) ds \right)}, & t_1 \leq t_2 \\ e^{-Z \left(\int_0^{t_2} \lambda_{02}(s) ds \right)} e^{-Z \left(\int_0^{t_1} \lambda_{01}(s) ds \right)}, & t_1 > t_2 \end{cases}$$

ou equivalentemente

$$S(t_1, t_2 | Z) = \begin{cases} e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1))}, & t_1 \leq t_2 \\ e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1))}, & t_1 > t_2 \end{cases} \tag{4.9}$$

A transformada de Laplace da distribuição associada à variável de fragilidade é crucial para a inferência, dado que as funções de sobrevivência e as taxas de transição não condicionais,

podem ser facilmente obtidas usando a transformada de Laplace. Portanto, a função de verossimilhança também pode ser obtida por meio da transformação de Laplace (WIENKE, 2010). No contexto de riscos semicompetitivos, a partir do modelo Doença-Morte com fragilidade compartilhada, é possível escrever as taxas de transição marginais e as funções de sobrevivência marginais em função da transformada de Laplace associadas à fragilidade, $\mathcal{L}_Z(s)$, a partir das taxas de transição condicionais (dado Z) nas equações (4.4)-(4.6). Este resultado é apresentado no Teorema 4.2.2 a seguir.

Teorema 4.2.2. Sejam as taxas de transição $\lambda_1(t_1|Z)$, $\lambda_2(t_2|Z)$ e $\lambda_{12}(t_2|t_1, Z)$, definidas em (4.4), (4.5) e (4.6), respectivamente, onde Z é uma variável aleatória contínua (fragilidade) com suporte positivo, média 1 e variâncias finita. Se a distribuição Z tem transformada de Laplace, $\mathcal{L}_Z(s)$, então

1. Para $0 < t_1 \leq t_2$, temos as seguintes funções de sobrevivência marginais

$$(1.a) \quad S_{12}(t_2|t_1) = \mathcal{L}_Z(\Lambda_{03}(t_1, t_2)).$$

$$(1.b) \quad S(t_1, t_2) = \mathcal{L}_Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2)).$$

2. As taxas de transição marginais são dadas por

$$(2.a) \quad \lambda_i(t_i) = -\lambda_{0i}(t_i) \frac{d}{ds} \log(\mathcal{L}_Z(s)), \quad s = \Lambda_{01}(t_i) + \Lambda_{02}(t_i), \quad i = 1, 2 \text{ e}$$

$$(2.b) \quad \lambda_{12}(t_2|t_1) = -\lambda_{03}(t_2) \frac{d}{ds} \log\left(\frac{d}{ds} \mathcal{L}_Z(s)\right), \quad s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2), \quad 0 < t_1 \leq t_2.$$

onde $\Lambda_{03}(t_1, t_2) = \Lambda_{03}(t_2) - \Lambda_{03}(t_1)$ e $\Lambda_{0i}(t_i) = \int_0^{t_i} \lambda_{0i}(t) dt$, $i = 1, 2, 3$. são as funções de distribuição base acumulada.

Prova:

(1) Para provar o Item (1.a), note que, a partir da função de sobrevivência condicional dada em (4.8), temos o seguinte resultado

$$\begin{aligned} S_{12}(t_2|t_1) &= \int_0^\infty S_{12}(t_2|t_1, z) g(z) dz \\ &= \int_0^\infty e^{-z\Lambda_{03}(t_1, t_2)} g(z) dz \\ &= \mathcal{L}_Z(\Lambda_{03}(t_1, t_2)), \end{aligned}$$

em que $g(z)$ representa a função densidade da fragilidade Z . De forma análoga e usando o resultado (4.9), pode se provar o Item (1.b)

(2) O resultado do item (2) são obtidos a partir da definição de taxas de transição 4.2.1 e usando as funções de sobrevivência condicionais (4.7) e (4.8), isto é

(2.a) Veja que de (4.1), na Definição 4.2.1, obtemos que

$$\begin{aligned}\lambda_1(t_1) &= \int_0^\infty \left(\lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_1 \in [t_1, t_1 + \Delta) | T_1 \geq t_1, T_2 \geq t_1, z]}{\Delta} \right) \frac{\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | z] g(z)}{\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1]} dz \\ &= \frac{1}{S(t_1)} \int_0^\infty \lambda_1(t_1 | z) \mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | z] g(z) dz,\end{aligned}\quad (4.10)$$

Logo, usando (4.7) obtemos que

$$S(t_1) = \int_0^\infty \mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | z] g(z) dz = \int_0^\infty e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1))} g(z) dz = \mathcal{L}_Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1)),$$

onde, $\mathcal{L}_Z(\cdot)$ é a transformada de Laplace de Z . Desta forma, em (4.10) obtemos

$$\lambda_1(t_1) = \frac{\lambda_{01}(t_1)}{\mathcal{L}_Z(s)} \int_0^\infty z e^{-zs} g(z) dz = -\frac{\lambda_{01}(t_1)}{\mathcal{L}_Z(s)} \frac{d}{ds} \mathcal{L}_Z(s), \quad (4.11)$$

onde, $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1)$. A última igualdade é obtida usando a propriedade de Transformada de Laplace. Analogamente, pode se provar que

$$\lambda_2(t_2) = -\frac{\lambda_{02}(t_2)}{\mathcal{L}_Z(s)} \frac{d}{ds} \mathcal{L}_Z(s), \quad (4.12)$$

(2.b.) Seguindo a mesma ideia da prova anterior, temos

$$\lambda_{12}(t_2 | t_1) = \frac{1}{\mathbb{P}[T_1 = t_1, T_2 \geq t_2]} \int_0^\infty \lambda_{12}(t_2 | t_1, z) \mathbb{P}[T_1 = t_1, T_2 \geq t_2 | z] g(z) dz \quad (4.13)$$

logo,

$$\begin{aligned}\mathbb{P}[T_1 = t_1, T_2 \geq t_2 | z] &= \mathbb{P}[T_2 \geq t_2 | T_1 = t_1, Z] \mathbb{P}[T_1 = t_1 | Z] \\ &= \mathbb{P}[T_2 \geq t_2 | T_1 = t_1, Z] \\ &\times \frac{\mathbb{P}[T_1 = t_1, T_2 \geq t_1 | Z]}{\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | Z]} \frac{\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | Z]}{\mathbb{P}[T_2 \geq t_1 | Z]}\end{aligned}\quad (4.14)$$

Note que, a partir de (4.8) e (4.7)

$$\mathbb{P}[T_2 \geq t_2 | T_1 = t_1, Z] = e^{-Z(\Lambda_{03}(t_2) - \Lambda_{03}(t_1))},$$

e

$$\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | Z] = e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1))},$$

respectivamente. Logo, usando a definição de $\lambda_1(t_1 | Z)$, não é difícil mostrar que

$$\frac{\mathbb{P}[T_1 = t_1, T_2 \geq t_1 | Z]}{\mathbb{P}[T_1 \geq t_1, T_2 \geq t_1 | Z]} = \lambda_1(t_1 | Z) = Z \lambda_{01}(t_1).$$

Assim, em (4.14), temos que

$$\mathbb{P}[T_1 = t_1, T_2 \geq t_2 | z] = Z \lambda_{01}(t_1) e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1))}. \quad (4.15)$$

Fazendo a integração em z , temos

$$\begin{aligned} \mathbb{P}[T_1 = t_1, T_2 \geq t_2] &= \int_0^\infty \mathbb{P}[T_1 = t_1, T_2 \geq t_2 | z] g(z) dz \\ &= \int_0^\infty Z \lambda_{01}(t_1) e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1))} g(z) dz \\ &= -\frac{\lambda_{01}(t_1)}{\mathcal{L}_Z(s)} \frac{d}{ds} \mathcal{L}_Z(s) \end{aligned} \quad (4.16)$$

onde, $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1)$. Finalmente, substituindo (4.15) e (4.16) em (4.13), obtemos

$$\begin{aligned} \lambda_{12}(t_2 | t_1) &= -\frac{\lambda_{03}(t_2)}{\frac{1}{\mathcal{L}_Z(s)} \frac{d}{ds} \mathcal{L}_Z(s)} \int_0^\infty Z^2 e^{-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1))} g(z) dz \\ &= -\frac{\lambda_{03}(t_2)}{\frac{1}{\mathcal{L}_Z(s)} \frac{d}{ds} \mathcal{L}_Z(s)} \frac{d^2}{ds^2} \mathcal{L}_Z(s) \\ &= -\lambda_{03}(t_2) \frac{d}{ds} \log \left(\frac{1}{ds} \mathcal{L}_Z(s) \right). \end{aligned} \quad (4.17)$$

onde, $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_2) - \Lambda_{03}(t_1)$.

□

Observação 4.2.3. Os resultados do Teorema 4.2.2, Itens (1.b) e (2.b) foram obtidos assumindo o modelo geral, isto é, quando $\lambda_{02}(t_2) \neq \lambda_{03}(t_2)$. Para obter estes resultados no caso do modelo restrito, basta fazer $\lambda_{02}(t_2) = \lambda_{03}(t_2)$ nos Itens (1.b) e (2.b) ou, equivalentemente, $\Lambda_{02}(t_2) = \Lambda_{03}(t_2)$.

4.2.2 O modelo de doença-morte com fragilidade compartilhada PVF

Nesta seção, consideramos a estrutura doença-morte para o modelo geral; os resultados correspondentes para o modelo restrito seguem diretamente assumindo $\lambda_{02}(\cdot) = \lambda_{03}(\cdot)$. No Teorema 4.2.2 consideramos a variável de fragilidade Z com distribuição PVF, na versão apresentada por Wienke (2010), com função densidade (3.49). Para evitar o problema de identificabilidade consideramos $\mathbb{E}[Z] = \mu = 1$ e $\text{Var}(Z) = \theta = 1/\sigma$. Assim, da transformada de Laplace dada em

(3.51) e do Teorema 4.2.2, as taxas de transição marginal são dadas por

$$\lambda_i(t_i) = \lambda_{0i}(t_i) \left(1 + \frac{\theta(\Lambda_{01}(t_i) + \Lambda_{02}(t_i))}{1 - \gamma} \right)^{\gamma-1}, \quad t_i > 0, \quad i = 1, 2. \quad (4.18)$$

$$\begin{aligned} \lambda_{12}(t_2|t_1) &= \lambda_{03}(t_2) \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1 - \gamma} \right)^{-1} \\ &\times \left[\theta + \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1 - \gamma} \right)^{\gamma} \right], \quad 0 < t_1 < t_2. \end{aligned} \quad (4.19)$$

Observe que, a taxa de transição marginal do estado “recorrência” para o estado “morte”, $\lambda_{12}(t_2|t_1)$, em (4.19) depende de t_1 e t_2 , portanto, não é markoviano em oposição à sua taxa de transição condicional correspondente, a menos que Z seja constante ($\theta = 0$). Neste modelo nem todos os sujeitos têm a mesma fragilidade, portanto, as taxas de transição condicionais são comparáveis apenas entre os indivíduos que compartilham fragilidade. Isso faz com que a interpretação das taxas de transição marginal e condicional seja diferente (LEE; NELDER *et al.*, 2004). Por outro lado, a partir do Teorema 4.2.2, podemos obter a função de sobrevivência bivariada marginal $S(t_1, t_2)$, bem como a função de sobrevivência marginal para o tempo até o evento terminal, dado que o evento não terminal ocorreu no instante t_1 , $S_{12}(t_2|t_1)$, isto é

$$S(t_1, t_2) = \exp \left\{ \frac{1 - \gamma}{\theta \gamma} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1 - \gamma} \right)^{\gamma} \right] \right\} \quad (4.20)$$

$$S_{12}(t_2|t_1) = \exp \left\{ \frac{1 - \gamma}{\theta \gamma} \left[1 - \left(1 + \frac{\theta \Lambda_{03}(t_1, t_2)}{1 - \gamma} \right)^{\gamma} \right] \right\}, \quad (4.21)$$

A distribuição PVF é uma família generalizada de famílias de distribuições que inclui as distribuições Gama, inversa Gaussiana (IG) e positiva estável (WIENKE, 2010). Na Tabela 2 mostramos as taxas de transição marginal e as funções de sobrevivência marginais $S(t_1, t_2)$ e $S_{12}(t_2|t_1)$, para os modelos IG e Gama. Primeiro, veja que, se fixamos $\gamma = 0.5$ em (4.18) e (4.19), podemos obter as três taxas de transição marginal para o modelo IG, e as funções de sobrevivência marginais são obtidas a partir de (4.20) e (4.21). Logo, se $\gamma \rightarrow 0$ em (4.18) e (4.19), as taxas de transição marginal para o modelo Gama, proposto por Xu, Kalbfleisch e Tai (2010), são obtidas. Similarmente, podemos obter as funções de sobrevivência marginal para o modelo Gama quando $\gamma \rightarrow 0$ em (4.20) e (4.21).

A distribuição Poisson composta (Pc), com parâmetros $\mu > 0$, $\sigma > 0$, e $\gamma < 0$ tem função densidade de probabilidade similar à densidade da distribuição PVF definida em (3.49), isto considerando uma parametrização conveniente (veja Seção 3.4.6). Assim, Wienke (2010) mostra que a transformada de Laplace da distribuição PVF, definida em (3.50), é igual ao da distribuição Pc (exceto para o intervalo do parâmetro γ , que pode ser negativo no modelo Pc). Consequentemente, as taxas de transição marginais considerando o modelo de fragilidade Pc são iguais às expressões em (4.18) e (4.19). Da mesma forma, conclui-se para as funções de sobrevivência marginal dadas em (4.20) e (4.21). Além disso, se $\gamma < 0$, pode-se mostrar que a integral das taxas (4.18) e (4.19) sobre $[0, \infty)$ é finita. Consequentemente, a função de sobrevivência é impropria devido a que uma fração de indivíduos têm fragilidade zero e nunca experimentarão o evento de interesse quando $t_2 \rightarrow \infty$, isto é, se $\gamma < 0$. Assim, essa fração de indivíduos é dada por

Tabela 2 – Taxa de transição marginal e Função de sobrevivência marginal para os modelos Inversa Gaussiana (IG) e Gama

| Modelo | Taxa de transição marginal | |
|-----------------------------------|--|---|
| | $\lambda_i(t_i) \ t_i > 0, \ i = 1, 2$ | $\lambda_{12}(t_2 t_1), \ 0 < t_1 \leq t_2$ |
| IG ($\gamma = 0.5$) | $\lambda_{0i}(t_i)(1 + 2\theta(\Lambda_{01}(t_i) + \Lambda_{02}(t_i)))^{-1/2}$ | $\lambda_{03}(t_2)(\theta + (1 + 2\theta s)^{1/2})(1 + 2\theta s)^{-1}$ |
| Gama ($\gamma \rightarrow 0$) | $\lambda_{0i}(t_i)(1 + \theta(\Lambda_{01}(t_i) + \Lambda_{02}(t_i)))^{-1}$ | $\lambda_{03}(t_2)(\theta + 1)(1 + \theta s)^{-1}$ |
| | | $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2)$ |
| Funções de sobrevivência marginal | | |
| | $S(t_1, t_2) \ 0 < t_1 \leq t_2,$ | $S_{12}(t_2 t_1), \ 0 < t_1 \leq t_2$ |
| IG($\gamma = 0.5$) | $\exp\left\{\frac{1}{\theta}\left(1 - (1 + 2\theta s)^{1/2}\right)\right\}$ | $\exp\left\{\frac{1}{\theta}\left(1 - (1 + 2\theta\Lambda_{03}(t_1, t_2))^{1/2}\right)\right\}$ |
| Gama ($\gamma \rightarrow 0$) | $(1 + \theta s)^{-1/\theta}$ | $(1 + \theta\Lambda_{03}(t_1, t_2))^{-1/\theta}$ |
| | $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2)$ | |

$$p_0 = \lim_{\substack{t_1 \rightarrow \infty \\ t_2 \rightarrow \infty}} S(t_1 t_2) = \exp\left(\frac{1 - \gamma}{\theta \gamma}\right). \quad (4.22)$$

Por outro lado, as probabilidades de transição marginal $\mathbb{P}_{01}(t_1)$, $\mathbb{P}_{02}(t_2)$ e $\mathbb{P}_{12}(t_2|t_1)$ são a probabilidade de estar no estado 1 no instante t_1 dado que o estado interior 0 foi inserido no tempo 0, a probabilidade de estar no estado 2 no instante t_2 dado que o estado anterior 0 foi inserido no tempo 0, e a probabilidade de estar no estado 2 no instante t_2 dado que o estado anterior foi 1 e foi inserido no tempo t_1 , respectivamente. Usando as taxas de transição condicionais e o laplaciano de Z , e desde que o tempo de transição para os três casos ($0 \rightarrow 1$, $0 \rightarrow 2$ e $1 \rightarrow 2$) são desconhecidos, integramos sobre os possíveis tempos de transição. Assim, podemos obter

$$\mathbb{P}_{0i}(t_i) = \int_0^{t_i} \lambda_{0i}(t) \left(1 + \frac{\theta s_i(t, t_i)}{1 - \gamma}\right)^{\gamma-1} \exp\left\{\frac{1 - \gamma}{\theta \gamma} \left[1 - \left(1 + \frac{\theta s_i(t, t_i)}{1 - \gamma}\right)^\gamma\right]\right\} dt, \quad i = 1, 2. \quad (4.23)$$

$$\mathbb{P}_{12}(t_2|t_1) = \int_{t_1}^{t_2} \lambda_{03}(t) \left(1 + \frac{\theta s_3(t, t_1)}{1 - \gamma}\right)^{\gamma-1} \exp\left\{\frac{1 - \gamma}{\theta \gamma} \left[1 - \left(1 + \frac{\theta s_3(t, t_1)}{1 - \gamma}\right)^\gamma\right]\right\} dt \quad (4.24)$$

onde $s_1(t, t_1) = \Lambda_{01}(t) + \Lambda_{02}(t) + \Lambda_{03}(t, t_1)$, $s_2(t) = \Lambda_{01}(t) + \Lambda_{02}(t)$ e $s_3(t, t_1) = \Lambda_{03}(t_1, t)$. As provas desses resultados encontram-se no Apêndice B. As integrais nas equações (4.23) e (4.24) serão resolvidos numericamente. Este procedimento é feito pelo método de quadratura de Gauss-Kronrod, usando a função `integrate()` do pacote `stats` no R (R Core Team, 2022).

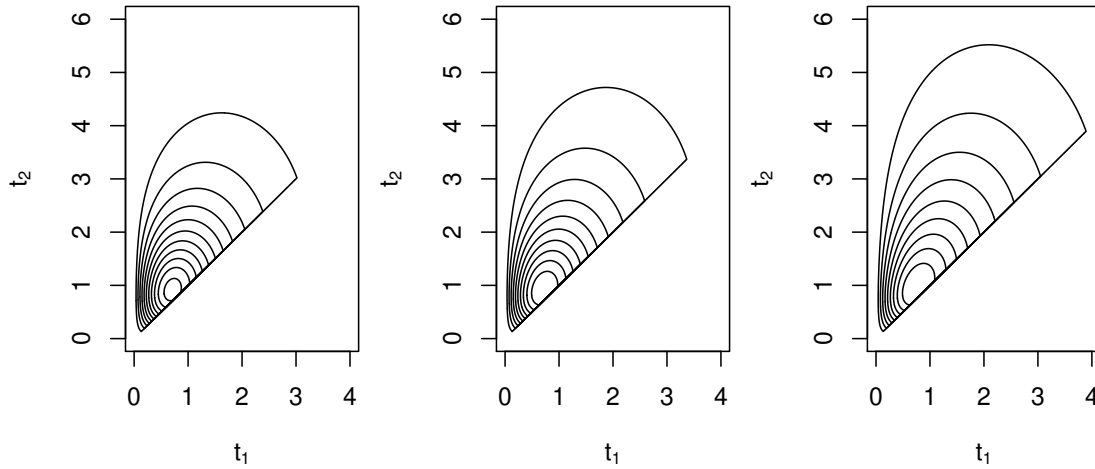


Figura 4 – Função densidade conjunta de (T_1, T_2) , em $T_1 \leq T_2$. Com $\gamma = -0.5$ (painel esquerdo), $\gamma = 0$ (painel central) e $\gamma = 0.5$ (painel direito)

4.2.3 Função de densidade de probabilidade e dependência local

No Apêndice A.1 é obtida a função densidade de probabilidade conjunta de (T_1, T_2) no quadrante superior $0 < t_1 \leq t_2$. A partir desse resultado e considerando o modelo condicional dado em (4.4)-(4.6), temos que

$$\begin{aligned} f(t_1, t_2|Z) &= \lambda_1(t_1|Z)\lambda_{12}(t_2|t_1, Z) \exp \left\{ - \int_0^{t_1} (\lambda_1(s|Z) + \lambda_2(s|Z)) ds + \int_{t_1}^{t_2} \lambda_{12}(s|t_1, Z) ds \right\} \\ &= \lambda_{01}(t_1)\lambda_{03}(t_2)Z^2 \exp \left\{ -Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2)) \right\} \end{aligned}$$

em que $\Lambda_{03}(t_1, t_2) = \Lambda_{03}(t_2) - \Lambda_{03}(t_1)$ e $\Lambda_{0i}(t) = \int_0^t \lambda_{0i}(s) ds$ para $i = 1, 2, 3$. Então a função densidade marginal de (T_1, T_2) no quadrante superior $0 < t_1 \leq t_2$, é obtida fazendo a integração com respeito à fragilidade Z com distribuição PVF e densidade $g(z)$ dada em (3.49), isto é,

$$f(t_1, t_2) = \lambda_{01}(t_1)\lambda_{03}(t_2) \int_0^\infty Z^2 \exp \left\{ -Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2)) \right\} g(z) dz$$

Logo, usando uma propriedade de transformada de Laplace, $\mathcal{L}_Z^{(r)}(s) = (-1)^r \mathbb{E}[Z^r e^{-Zs}]$, vista no Capítulo 3, dada em 3.5, concluímos que a função densidade de probabilidade conjunta de (T_1, T_2) no quadrante superior $0 < t_1 \leq t_2$, é dada por

$$\begin{aligned} f(t_1, t_2) &= \lambda_{01}(t_1)\lambda_{03}(t_2) \exp \left\{ \frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1-\gamma} \right)^\gamma \right] \right\} \\ &\quad \times \left[\theta + \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1-\gamma} \right)^\gamma \right] \\ &\quad \times \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1-\gamma} \right)^{\gamma-2} \end{aligned} \quad (4.25)$$

A Figura 4 ilustra o gráfico de contornos da função densidade conjunta $f(t_1, t_2)$, quando $T_1 \leq T_2$. De

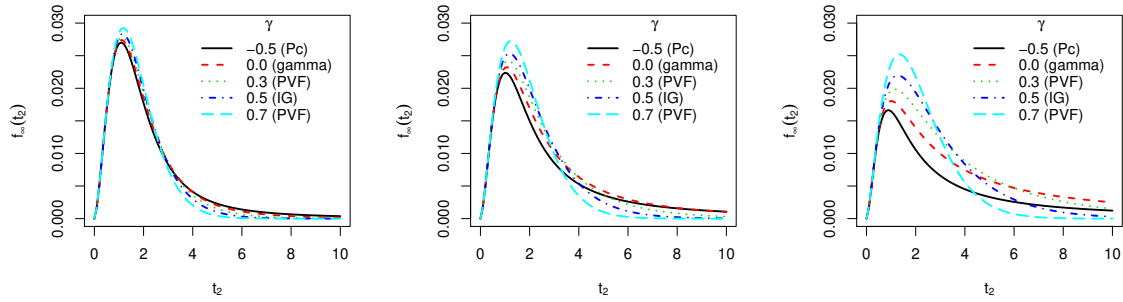


Figura 5 – Função densidade para T_2 , quando $T_1 > T_2$ ($T_1 = \infty$). Com $\theta = 0.5$ (painel esquerdo), $\theta = 1$ (painel central) e $\theta = 2$ (painel direito)

forma muito similar, a partir do Apêndice A.2, pode-se mostrar que a função densidade de probabilidade na linha $T_1 = \infty$, isto é, quando o evento terminal aconteceu primeiro, é dado por

$$f_{\infty}(t_2) = \lambda_{02}(t_2) \exp \left\{ \frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(t_2) + \Lambda_{02}(t_2))}{1-\gamma} \right)^{\gamma} \right] \right\} \\ \times \left(1 + \frac{\theta(\Lambda_{01}(t_2) + \Lambda_{02}(t_2))}{1-\gamma} \right)^{\gamma-1}.$$

A Figura 5 compara as funções de densidade para o evento terminal, dado que não aconteceu o evento não terminal, isto variando γ . Aqui, podemos observar que para $\theta = 0.5$ as curvas são semelhantes para os diferentes submodelos (veja painel esquerdo). No entanto, quanto maior o valor de θ , as curvas são diferentes umas das outras (veja por exemplo painel direito). Note que o parâmetro θ caracteriza a dependência dos tempos T_1 e T_2 , então a partir da Figura 5 podemos dizer que a estrutura de dependência é diferente para cada submodelo da família de distribuições PVF.

Por outro lado, a partir de (4.20), o modelo bivariado induzido pela fragilidade com distribuição PVF com média 1 e variância $\theta > 0$, inclui uma medida de associação local entre os tempos T_1 e T_2 , $\vartheta^*(t_1, t_2)$, definida em (3.26), como uma função de θ e é dada por

$$\vartheta^*(t_1, t_2) = 1 + \theta \left(1 + \frac{\theta(\Lambda_{01}(t_1) + \Lambda_{02}(t_1) + \Lambda_{03}(t_1, t_2))}{1-\gamma} \right)^{-\gamma}, \quad 0 < t_1 \leq t_2. \quad (4.26)$$

A Figura 6 ilustra o comportamento de $\vartheta^*(t_1, t_2)$ para diferentes valores de γ . Podemos observar que $\vartheta^*(t_1, t_2)$ em (4.26) $\vartheta^*(t_1, t_2) = 1 + \theta$ é constante em (t_1, t_2) , quando $\gamma \rightarrow 0$, (veja painel central), isto corresponde à fragilidade Gama, assim, neste caso, a associação entre T_1 e T_2 é menor quando θ é pequena e a associação cresce quando θ é maior. Porém, para a fragilidade PVF ($0 < \gamma \leq 1$), $\vartheta^*(t_1, t_2)$ depende de (t_1, t_2) , a dependência é menor quando os tempos crescem (veja painel direito), além disso $\vartheta^*(t_1, t_2) \rightarrow 1$ quando $t_1 \rightarrow \infty$ e $t_2 \rightarrow \infty$. Isso significa que para um período suficientemente grande, T_1 e T_2 são independentes. Por outro lado, se $\gamma < 0$, a medida de associação local entre T_1 e T_2 do modelo PVF corresponde ao modelo Pc, na qual $\vartheta^*(t_1, t_2) \rightarrow \infty$ quando $t_1 \rightarrow \infty$, isto é, quanto maior os tempos t_1 e t_2 maior a dependência entre eles (veja painel esquerdo).

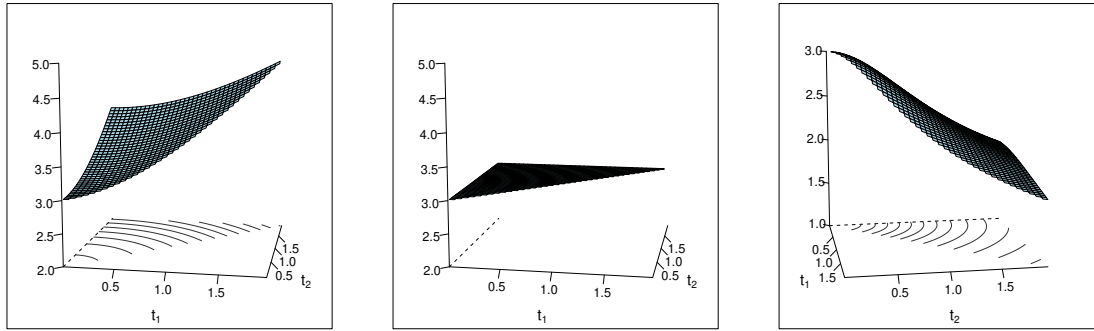


Figura 6 – Medida de associação local entre os tempos T_1 e T_2 , para $\gamma = -0.5$ (painel esquerdo), $\gamma = 0$ (painel central) e $\gamma = 0.5$ (painel direito)

Na seguinte seção, desenvolvemos o procedimento inferencial baseado na teoria da verossimilhança para o novo modelo de doença-morte dado na Equação (4.18)-(4.19). A seguir, assumimos uma distribuição Weibull para a função de risco de base $\lambda_{0i}(t) = \alpha_i \beta_i t^{\beta_i - 1}$, para $\alpha_i > 0$ e $\beta_i > 0$, $i = 1, 2, 3$.

4.3 Inferência para o novo modelo doença-morte

Supondo que n é o número de indivíduos em estudo. Sejam T_{j1} e T_{j2} os tempos até que o evento não terminal e terminal ocorram, respectivamente, C_j é o tempo de censura para o j -ésimo sujeito, com $j = 1, \dots, n$, a qual é considerada censura à direita. Logo, o conjunto de dados observados incluindo um vetor de p covariáveis, $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})'$ para o j -ésimo sujeito será denotado por $D = (Y_{j1}, Y_{j2}, \delta_{j1}, \delta_{j2}, \mathbf{x}_j)$, com $j = 1, \dots, n$, onde como foi visto na Seção 2.2, pode ser descrito como $Y_{j2} = \min\{T_{j2}, C_j\}$, $\delta_{j2} = I_{\{T_{j2} \leq C_j\}}$, $Y_{j1} = \min\{T_{j1}, Y_{j2}\}$ e $\delta_{j1} = I_{\{T_{j1} \leq Y_{j2}\}}$, $j = 1, \dots, n$.

A importância de incorporar covariáveis no modelo reside no fato de que as variáveis explicativas podem afetar o tempo de sobrevivência. Incluímos covariáveis no modelo de doença-morte, incorporando-as nas taxas de transição condicionais, análogas às equações (4.4), (4.5) e (4.6), em que cada taxa foi multiplicada por um termo que inclui as covariáveis, isto é,

$$\lambda_1(t_1|Z, \mathbf{x}) = Z\lambda_{01}(t_1) \exp(\boldsymbol{\varphi}'_1 \mathbf{x}), \quad t_1 > 0 \quad (4.27)$$

$$\lambda_2(t_2|Z, \mathbf{x}) = Z\lambda_{02}(t_2) \exp(\boldsymbol{\varphi}'_2 \mathbf{x}), \quad t_2 > 0 \quad (4.28)$$

$$\lambda_{12}(t_2|t_1, Z, \mathbf{x}) = Z\lambda_{03}(t_2) \exp(\boldsymbol{\varphi}'_3 \mathbf{x}), \quad 0 < t_1 \leq t_2 \quad (4.29)$$

onde $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ é um vetor de p covariáveis de um sujeito específico, $\boldsymbol{\varphi}_i = (\varphi_{i1}, \varphi_{i2}, \dots, \varphi_{ip})'$, $i = 1, 2, 3$ são os vetores de coeficientes e Z é a fragilidade compartilhada entre as taxas de transição de um sujeito específico, tomadas para serem independentemente distribuídas de \mathbf{x} . Logo, as correspondente taxas de transição marginais com covariáveis são as mesmas que nas equações (4.18), e (4.19) onde as funções de risco de base, $\lambda_{0i}(\cdot)$, e as funções de risco base acumuladas, $\Lambda_{0i}(\cdot)$, são multiplicadas com $\exp(\boldsymbol{\varphi}'_i \mathbf{x})$ com $i = 1, 2, 3$.

As funções de verossimilhança do modelo doença-morte são baseadas nas taxas de transição condicionais, que são obtidas ao encontrar primeiro as contribuições condicionais de cada um dos 4 possíveis

casos, apresentados na Tabela 1, ou seja, $(\delta_{j1}, \delta_{j2}) = (1, 1), (0, 1), (1, 0), (0, 0)$, e logo integrando sobre Z usando a função densidade de probabilidade $g(z)$, da fragilidade Z , dada em (3.49), mais especificamente, usando a transformada de Laplace da fragilidade Z , dada em (3.51). Para maiores detalhes, veja o Apêndice A, onde a obtenção da função de verossimilhança, dada em (4.30), é descrita em forma detalhada. Neste modelo assumimos que o efeito das covariáveis é o mesmo para todos os tempos e o tempo de censura C_j é independente de T_{j1}, T_{j2} dado \mathbf{x}_j . Seja $\boldsymbol{\vartheta} = (\theta, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3)$, o vetor de parâmetros a serem estimados, no qual $\boldsymbol{\eta}_i = (\alpha_i, \beta_i)$ é o vetor de parâmetros do modelo Weibull para cada taxa $\lambda_{0i}(t) = \alpha_i \beta_i t^{\beta_i - 1}$, em que α_i é o parâmetro de escala e β_i é o parâmetro de forma, e $\boldsymbol{\varphi}_i = (\varphi_{i1}, \varphi_{i2}, \dots, \varphi_{ip})'$ são os vetores de coeficientes, com $i = 1, 2, 3$. O logaritmo da função de verossimilhança para $\boldsymbol{\vartheta}$ dado os dados observados é dado por

$$\begin{aligned} \ell(\boldsymbol{\vartheta}) = & \sum_{j=1}^n \left\{ \delta_{j1} \log \lambda_{01}(Y_{j1} | \boldsymbol{\eta}_1) + \delta_{j2}(1 - \delta_{j1}) \log \lambda_{02}(Y_{j2} | \boldsymbol{\eta}_2) + \delta_{j1} \delta_{j2} \log \lambda_{03}(Y_{j2} | \boldsymbol{\eta}_3) \right\} \\ & + \sum_{j=1}^n \left\{ \delta_{j1} \boldsymbol{\varphi}'_1 \mathbf{x}_j + \delta_{j2}(1 - \delta_{j1}) \boldsymbol{\varphi}'_2 \mathbf{x}_j + \delta_{j1} \delta_{j2} \boldsymbol{\varphi}'_3 \mathbf{x}_j \right\} \\ & + \sum_{j=1}^n \left\{ \delta_{j1} \delta_{j2} \log \left[\theta + \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1 - \gamma} \right)^\gamma \right] + \delta_{j1}(\gamma - 1 - \delta_{j2}) \log \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1 - \gamma} \right) \right\} \\ & + \sum_{j=1}^n \left\{ \log \left[\frac{\delta_{j1}(1 - \gamma)}{\theta \gamma} \left(1 - \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1 - \gamma} \right)^\gamma \right) \right] \right\} \\ & + \sum_{j=1}^n \left\{ (\gamma - 1) \delta_{j2}(1 - \delta_{j1}) \log \left(1 + \frac{\theta K_2(\mathbf{x}_j)}{1 - \gamma} \right) \right\} \\ & + \sum_{j=1}^n \left\{ \log \left[\frac{(1 - \delta_{j1})(1 - \gamma)}{\theta \gamma} \left(1 - \left(1 + \frac{\theta K_2(\mathbf{x}_j)}{1 - \gamma} \right)^\gamma \right) \right] \right\} \end{aligned} \quad (4.30)$$

onde $K_1(\mathbf{x}_j) = (\Lambda_{01}(Y_{j1} | \boldsymbol{\eta}_1) e^{\boldsymbol{\varphi}'_1 \mathbf{x}_j} + \Lambda_{02}(Y_{j1} | \boldsymbol{\eta}_2) e^{\boldsymbol{\varphi}'_2 \mathbf{x}_j} + \Lambda_{03}(Y_{j1}, Y_{j2} | \boldsymbol{\eta}_3) e^{\boldsymbol{\varphi}'_3 \mathbf{x}_j}) e$

$$K_2(\mathbf{x}_j) = (\Lambda_{01}(Y_{j1} | \boldsymbol{\eta}_1) e^{\boldsymbol{\varphi}'_1 \mathbf{x}_j} + \Lambda_{02}(Y_{j1} | \boldsymbol{\eta}_2) e^{\boldsymbol{\varphi}'_2 \mathbf{x}_j}), \text{ para } j = 1, \dots, n.$$

O estimador de máxima verossimilhança de $\boldsymbol{\vartheta}$, denotado por $\hat{\boldsymbol{\vartheta}}$, pode ser obtido numericamente maximizando a equação (4.30). Embora todos os parâmetros do modelo proposto possam ser estimados por esse procedimento, também adotamos uma abordagem de verossimilhança perfilada para γ , devido a que este procedimento ofereceu as estimativas mais precisas e, ao mesmo tempo, provou ser computacionalmente o mais eficiente. A maximização de (4.30) segue as mesmas duas etapas para obter estimativas de $\boldsymbol{\vartheta}$. Na primeira etapa do processo iterativo fixamos diferentes valores para γ no intervalo $(-\tau, 1)$, com $\tau > 0$. Logo, encontramos as estimativas de máxima verossimilhança $\tilde{\theta}(\gamma)$, $\tilde{\boldsymbol{\eta}}_1(\gamma)$, $\tilde{\boldsymbol{\eta}}_2(\gamma)$, $\tilde{\boldsymbol{\eta}}_3(\gamma)$, $\tilde{\boldsymbol{\varphi}}_1(\gamma)$, $\tilde{\boldsymbol{\varphi}}_2(\gamma)$ e $\tilde{\boldsymbol{\varphi}}_3(\gamma)$ e o logaritmo da função de verossimilhança maximizada $\ell_{\max}(\gamma)$ é determinada. Nesta etapa, empregamos a função `optim()` do pacote `stats` no software R (R Core Team, 2022). Na segunda etapa a função log-verossimilhança $\ell_{\max}(\gamma)$ é maximizada em seguida $\hat{\gamma}$ é obtida. As estimativas de máxima verossimilhança de θ , $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, $\boldsymbol{\eta}_3$, $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$ são, respectivamente, dadas por $\hat{\theta} = \tilde{\theta}(\hat{\gamma})$, $\hat{\boldsymbol{\eta}}_1 = \tilde{\boldsymbol{\eta}}_1(\hat{\gamma})$, $\hat{\boldsymbol{\eta}}_2 = \tilde{\boldsymbol{\eta}}_2(\hat{\gamma})$, $\hat{\boldsymbol{\eta}}_3 = \tilde{\boldsymbol{\eta}}_3(\hat{\gamma})$, $\hat{\boldsymbol{\varphi}}_1 = \tilde{\boldsymbol{\varphi}}_1(\hat{\gamma})$, $\hat{\boldsymbol{\varphi}}_2 = \tilde{\boldsymbol{\varphi}}_2(\hat{\gamma})$ e $\hat{\boldsymbol{\varphi}}_3 = \tilde{\boldsymbol{\varphi}}_3(\hat{\gamma})$

Os procedimentos discutidos neste trabalho são desenvolvidos assumindo γ fixado. O intervalo de teste e estimativas para $\boldsymbol{\vartheta} = (\theta, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3)$, sob condições de regularidade adequadas (MALLER; ZHOU, 1996), pode-se demonstrar que a distribuição assintótica do EMV $\hat{\boldsymbol{\vartheta}} =$

$(\hat{\theta}, \hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3, \hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3)$ é normal multivariada com vetor de médias $\boldsymbol{\vartheta}$ e matriz de covariância

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\vartheta}}) = \left\{ -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\}^{-1} = \{-J(\boldsymbol{\vartheta})\}^{-1},$$

avaliada em $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$. Os elementos da matriz observada $J(\boldsymbol{\vartheta})$ são obtidas numericamente. O pacote `numDeriv` em (GILBERT; VARADHAN, 2019) fornece uma aproximação numérica precisa para essa matriz.

Intervalo de confiança assintótico com coeficiente α para cada parâmetro ϑ_r é dado por

$$\left(\hat{\vartheta}_r - z_{\alpha/2} \sqrt{\hat{\Sigma}^{r,r}}, \hat{\vartheta}_r + z_{\alpha/2} \sqrt{\hat{\Sigma}^{r,r}} \right),$$

onde $\hat{\Sigma}^{r,r}$ é o r -ésimo elemento diagonal de $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\vartheta}})$ estimado em $\hat{\boldsymbol{\vartheta}}$, para $r = 1, \dots, p + \dim(\boldsymbol{\varphi}) + 1$, $\dim(\cdot)$ denota a dimensão do espaço paramétrico, e $z_{\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição normal padrão.

4.4 Estudo de simulação

Nesta seção apresentamos um estudo de simulação para o modelo com estrutura de dados semicompetitivos apresentado no Capítulo 4. Para avaliar o desempenho das estimativas de máxima verossimilhança do parâmetro do modelo de doença-morte proposto em (4.18) - (4.19), realizamos um estudo de simulação. Os dados de riscos semicompetitivos foram gerados, usando o algoritmo proposto por Selle (2016), adaptando-se ao nosso modelo. O algoritmo é o seguinte:

1. Gerar $Z \sim PVF(\gamma, 1, \frac{1}{\theta})$, para $\gamma = 0.5$;
2. **Simule o primeiro tempo do evento t_1** : fazendo $p_1 = \mathbb{P}[\text{não ter nenhuma transição antes de } t_1]$, $u_1 \sim \text{Uniforme}(0,1)$ e resolvendo a equação $p_1 = u_1$ para t_1 .
3. **Simule o segundo tempo do evento t_2** : fazendo $p_2 = \text{probabilidade condicional de permanecer no estado 1 até } t_2$, $u_2 \sim \text{Uniforme}(0,1)$ e resolvendo a equação $p_2 = u_2$ para t_2 .
4. Fazer $p = \mathbb{P}[\text{ir ao estado 1}]$, e $u \sim \text{Uniforme}(0,1)$. **If** $u \leq p$ **then** $Y_1 = t_1$, $\delta_1 = 1$ e $Y_2 = t_2$ **else** $Y_1 = Y_2 = t_1$ e $\delta_1 = 0$.
5. **Simule o tempo de censura C**: Fazendo $u_3 \sim \text{Uniforme}(0,1)$. **If** $u_3 < 0.5$ **then** $C \sim \text{Uniforme}(v, w)$ **else** $C = w$.
6. **If** $C > Y_2$ **then** $\delta_2 = 1$ **else if** $C \geq Y_1$ e $C \leq Y_2$ **then** $Y_2 = C$ e $\delta_2 = 0$ **else** $Y_1 = Y_2 = C$ e $\delta_1 = \delta_2 = 0$.

No passo 2, a probabilidade de não ter nenhuma transição antes do tempo t é dado em (4.7) e considerando o modelo condicional com covariáveis (4.27)-(4.29) podemos obter $p_1 = \exp(-Z[\Lambda_{01}(t)e^{\boldsymbol{\varphi}'_1 \mathbf{x}} + \Lambda_{02}(t)e^{\boldsymbol{\varphi}'_2 \mathbf{x}}])$. Similarmente, se existe uma transição para o estado 1, o segundo tempo do evento, t_2 , é simulado no passo 3, onde a probabilidade condicional de permanecer no estado 1 até o tempo t_2 é dado em (4.8) e a partir de (4.27)-(4.29), podemos obter $p_2 = \exp(-Z[\Lambda_{03}(t_1, t_2)e^{\boldsymbol{\varphi}'_3 \mathbf{x}}])$. Logo, no passo 4, pode se provar que a probabilidade de ir para o estado 1 no instante t , é $p = \lambda_{01}(t)e^{\boldsymbol{\varphi}'_1 \mathbf{x}} / (\lambda_{01}(t)e^{\boldsymbol{\varphi}'_1 \mathbf{x}} + \lambda_{02}(t)e^{\boldsymbol{\varphi}'_2 \mathbf{x}})$.

No passo 5, o tempo de censura foi simulado a partir de uma distribuição de mistura (XU; KALBFLEISCH; TAI, 2010) com probabilidade 0.5 de uma distribuição uniforme(1.5,3) e com probabilidade 0.5 de um ponto de massa em 3, sob estas condições, a porcentagem média de observações censuradas varia de 10% a 20%.

Uma vez que os dados de riscos semicompetitivos são gerados, avaliamos o desempenho dos EMV dos parâmetros do modelo em (4.18)-(4.19) com covariáveis incorporadas a partir de (4.27)-(4.29). Assim na função log-verossimilhança (4.30) do modelo proposto a as covariáveis x_{1j} e x_{2j} são geradas a partir da distribuição Bernoulli(0.5) e da distribuição uniforme(0, 1), respectivamente, para $j = 1, \dots, n$. A função de risco de base segue uma distribuição Weibull(α_i, β_i), $i = 1, 2, 3$. Para o modelo restrito, consideramos os vetores de coeficientes $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_2 = \boldsymbol{\varphi}_3 = (1, 1)$, os parâmetros da distribuição Weibull $\log \alpha_1 = \log \alpha_2 = \log \alpha_3 = 1$, e $\beta_1 = \beta_2 = \beta_3 = 1$, de modo que $EHR = 1$. Para o modelo geral, consideramos $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_2 = \boldsymbol{\varphi}_3 = (1, 1)$, $\log \alpha_1 = \log \alpha_2 = 1$ e $\log \alpha_3 = 1.25$ e $\beta_1 = \beta_2 = \beta_3 = 1$ levando a $EHR > 1$. Finalmente consideramos o parâmetro de fragilidade ou de dependência $\theta = 1$.

Para as simulações, tomamos os tamanhos da amostra $n = 250, 500$ e 1000 . Para cada configuração, realizamos 1000 simulações para calcular as médias dos EMV, denotado por $\bar{\boldsymbol{\vartheta}}$. Calculamos o Desvio padrão (DP) $SD = \sqrt{\sum_{k=1}^N (\hat{\boldsymbol{\vartheta}}_k - \bar{\boldsymbol{\vartheta}})^2 / N}$, onde $\hat{\boldsymbol{\vartheta}}_k$ é o vetor dos EMV na k -ésima simulação, $k = 1, \dots, N$, as raízes dos erros quadráticos médios (REQM) dos estimadores $RMSE = \sqrt{\sum_{k=1}^N (\hat{\boldsymbol{\vartheta}}_k - \boldsymbol{\vartheta}^{(0)})^2 / N}$ onde $\boldsymbol{\vartheta}^{(0)}$ é o vetor de valor inicial dos parâmetros, a média do erro padrão (MEP), onde $SE_k = \sqrt{\text{diag}(\hat{\boldsymbol{\Sigma}}_k(\hat{\boldsymbol{\vartheta}}_k))}$, $k = 1, \dots, N$, $\text{diag}(\hat{\boldsymbol{\Sigma}}_k(\hat{\boldsymbol{\vartheta}}_k))$ é a diagonal da k -ésima matriz de variâncias e covariâncias estimadas, as probabilidades de cobertura (PC) empíricas de 95% para os parâmetros do modelo e o viés. Os resultados da simulação listados nas Tabelas 3 e 4 para os modelos geral e restrito, respectivamente, demonstram que os EMV estão próximos dos valores verdadeiros, o viés, REQM e erros padrão diminuem à medida que o tamanho da amostra aumenta e as PC empíricas estão mais próximos do nível de cobertura nominal à medida que o tamanho da amostra aumenta, o que é esperado se o esquema de estimativa subjacente estiver funcionando corretamente para produzir estimativas consistentes e assintoticamente normais. O estudo de simulação também foi repetido para $\theta = 0,5 < 1$ e $\theta = 1,15 > 1$, e os resultados obtidos foram muito semelhantes.

4.5 Aplicação: dados de câncer de cólon

Nesta seção será apresentada uma aplicação para o modelo com estrutura de dados de risco semicompetitivos. Aplicamos o modelo de doença-morte para um conjunto de dados num ensaio clínico de câncer de cólon com estágio III, relatado em 1989 por NCCTG (North Central Cancer Treatment Group) (MOERTEL *et al.*, 1990) consistindo em dados de riscos semicompetitivos. O conjunto de dados aparecem como dados de cólon no pacote `survival` de R (THERNEAU, 2021). Aqui, os pacientes que tiveram ressecções do tumor cancerígeno com intenção curativa foram atribuídos aleatoriamente a um dos tratamentos: (i) apenas observação, (ii) apenas Levamisole (Lev) e (iii) Lev + Fluorouracil (5-FU). Aqui, após a ressecção cirúrgica completa do tumor, 929 pacientes foram acompanhados por 5 anos ou

Tabela 3 – Resultados da simulação para o modelo geral ($EHR > 1$)

| Amostra | Taxas | Par. | Valor verd. | Média | Viés | MEP | DP | REQM | PC |
|----------------|----------------|-----------------|-------------|--------|--------|-------|-------|-------|------|
| $n = 250$ | λ_1 | $\log \theta$ | 0.0 | -0.182 | -0.182 | 0.932 | 0.823 | 0.843 | 99.4 |
| | | $\log \beta_1$ | 0.0 | -0.003 | -0.003 | 0.122 | 0.110 | 0.110 | 95.4 |
| | | $\log \alpha_1$ | 1.0 | 0.995 | -0.005 | 0.480 | 0.433 | 0.433 | 95.3 |
| | | φ_{11} | 1.0 | 0.997 | -0.003 | 0.251 | 0.243 | 0.243 | 94.8 |
| | | φ_{12} | 1.0 | 1.009 | 0.009 | 0.402 | 0.387 | 0.387 | 94.5 |
| | λ_2 | $\log \beta_2$ | 0.0 | -0.006 | -0.006 | 0.122 | 0.108 | 0.108 | 95.0 |
| | | $\log \alpha_2$ | 1.0 | 0.986 | -0.014 | 0.480 | 0.418 | 0.418 | 94.6 |
| | | φ_{21} | 1.0 | 1.002 | 0.002 | 0.251 | 0.251 | 0.251 | 94.9 |
| | | φ_{22} | 1.0 | 1.004 | 0.004 | 0.402 | 0.394 | 0.394 | 95.7 |
| | λ_{12} | $\log \beta_3$ | 0.0 | -0.013 | -0.013 | 0.222 | 0.211 | 0.211 | 94.4 |
| | | $\log \alpha_3$ | 1.25 | 1.266 | 0.016 | 0.381 | 0.359 | 0.359 | 95.5 |
| | | φ_{31} | 1.0 | 0.993 | -0.007 | 0.326 | 0.321 | 0.321 | 95.5 |
| φ_{32} | | 1.0 | 1.035 | 0.035 | 0.493 | 0.501 | 0.502 | 94.7 | |
| $n = 500$ | λ_1 | $\log \theta$ | 0 | -0.058 | -0.059 | 0.641 | 0.592 | 0.594 | 97.8 |
| | | $\log \beta_1$ | 0.0 | 0.004 | 0.004 | 0.091 | 0.084 | 0.084 | 95.2 |
| | | $\log \alpha_1$ | 1.0 | 1.030 | 0.031 | 0.364 | 0.341 | 0.342 | 94.7 |
| | | φ_{11} | 1.0 | 0.999 | -0.001 | 0.180 | 0.175 | 0.175 | 96.6 |
| | | φ_{12} | 1.0 | 0.992 | -0.008 | 0.286 | 0.287 | 0.286 | 94.1 |
| | λ_2 | $\log \beta_2$ | 0.0 | 0.002 | 0.002 | 0.091 | 0.085 | 0.085 | 94.6 |
| | | $\log \alpha_2$ | 1.0 | 1.024 | 0.024 | 0.364 | 0.340 | 0.340 | 94.2 |
| | | φ_{21} | 1.0 | 1.000 | 0.000 | 0.180 | 0.179 | 0.179 | 95.2 |
| | | φ_{22} | 1.0 | 0.997 | -0.003 | 0.285 | 0.288 | 0.288 | 94.3 |
| | λ_{12} | $\log \beta_3$ | 0.0 | 0.001 | 0.001 | 0.165 | 0.155 | 0.155 | 94.5 |
| | | $\log \alpha_3$ | 1.25 | 1.279 | 0.029 | 0.280 | 0.273 | 0.275 | 95.5 |
| | | φ_{31} | 1.0 | 1.002 | 0.002 | 0.237 | 0.216 | 0.216 | 96.1 |
| φ_{32} | | 1.0 | 1.017 | 0.017 | 0.352 | 0.350 | 0.350 | 94.3 | |
| $n = 1000$ | λ_1 | $\log \theta$ | 0 | 0.010 | 0.010 | 0.448 | 0.417 | 0.417 | 96.7 |
| | | $\log \beta_1$ | 0.0 | 0.010 | 0.010 | 0.066 | 0.062 | 0.063 | 95.5 |
| | | $\log \alpha_1$ | 1.0 | 1.034 | 0.034 | 0.264 | 0.251 | 0.253 | 96.1 |
| | | φ_{11} | 1.0 | 1.010 | 0.010 | 0.128 | 0.126 | 0.126 | 95.8 |
| | | φ_{12} | 1.0 | 1.013 | 0.013 | 0.202 | 0.201 | 0.201 | 95.5 |
| | λ_2 | $\log \beta_2$ | 0.0 | 0.006 | 0.007 | 0.066 | 0.062 | 0.063 | 94.6 |
| | | $\log \alpha_2$ | 1.0 | 1.030 | 0.030 | 0.264 | 0.254 | 0.256 | 96.0 |
| | | φ_{21} | 1.0 | 1.006 | 0.006 | 0.128 | 0.125 | 0.125 | 96.4 |
| | | φ_{22} | 1.0 | 1.009 | 0.009 | 0.202 | 0.206 | 0.206 | 94.8 |
| | λ_{12} | $\log \beta_3$ | 0.0 | 0.008 | 0.008 | 0.119 | 0.111 | 0.112 | 94.5 |
| | | $\log \alpha_3$ | 1.25 | 1.274 | 0.023 | 0.200 | 0.188 | 0.189 | 96.2 |
| | | φ_{31} | 1.0 | 1.014 | 0.014 | 0.170 | 0.168 | 0.168 | 95.0 |
| φ_{32} | | 1.0 | 1.015 | 0.015 | 0.249 | 0.248 | 0.248 | 95.1 | |

Tabela 4 – Resultados da simulação para o modelo restrito ($EHR = 1$)

| Amostra | Taxas | Par. | Valor verd. | Média | Viés | MEP | DP | REQM | PC | |
|-----------------|-----------------|-----------------|----------------|--------|--------|--------|-------|-------|-------|------|
| $n = 250$ | λ_1 | $\log \theta$ | 0.0 | -0.151 | -0.151 | 0.865 | 0.761 | 0.776 | 99.2 | |
| | | $\log \beta_1$ | 0.0 | -0.001 | -0.001 | 0.117 | 0.102 | 0.102 | 96.6 | |
| | | $\log \alpha_1$ | 1.0 | 1.002 | 0.002 | 0.464 | 0.413 | 0.412 | 96.1 | |
| | | φ_{11} | 1.0 | 0.998 | -0.002 | 0.249 | 0.241 | 0.241 | 95.1 | |
| | | φ_{12} | 1.0 | 1.011 | 0.011 | 0.401 | 0.389 | 0.389 | 94.6 | |
| | λ_2 | $\log \beta_2$ | 0.0 | -0.002 | -0.002 | 0.117 | 0.103 | 0.103 | 95.7 | |
| | | $\log \alpha_2$ | 1.0 | 1.008 | 0.008 | 0.463 | 0.416 | 0.416 | 95.0 | |
| | | φ_{21} | 1.0 | 0.994 | -0.006 | 0.249 | 0.242 | 0.242 | 95.2 | |
| | | φ_{22} | 1.0 | 0.999 | -0.001 | 0.401 | 0.391 | 0.391 | 96.4 | |
| | λ_{12} | $\log \beta_3$ | 0.0 | -0.004 | -0.004 | 0.209 | 0.194 | 0.194 | 95.3 | |
| | | $\log \alpha_3$ | 1.0 | 1.023 | 0.023 | 0.367 | 0.352 | 0.353 | 96.5 | |
| | | φ_{31} | 1.0 | 0.999 | -0.001 | 0.321 | 0.321 | 0.321 | 94.8 | |
| | | φ_{32} | 0.5 | 1.023 | 0.492 | 0.503 | 0.023 | 0.503 | 94.0 | |
| | $n = 500$ | λ_1 | $\log \theta$ | 0.0 | -0.025 | -0.025 | 0.597 | 0.559 | 0.559 | 97.7 |
| | | | $\log \beta_1$ | 0.0 | 0.009 | 0.008 | 0.087 | 0.082 | 0.083 | 94.7 |
| $\log \alpha_1$ | | | 1.0 | 1.049 | 0.049 | 0.350 | 0.336 | 0.340 | 95.4 | |
| φ_{11} | | | 1.0 | 1.003 | 0.003 | 0.178 | 0.176 | 0.176 | 96.3 | |
| φ_{12} | | | 1.0 | 0.989 | -0.011 | 0.285 | 0.287 | 0.287 | 94.0 | |
| λ_2 | | $\log \beta_2$ | 0.0 | 0.006 | 0.006 | 0.087 | 0.082 | 0.082 | 94.9 | |
| | | $\log \alpha_2$ | 1.0 | 1.039 | 0.039 | 0.349 | 0.329 | 0.331 | 95.4 | |
| | | φ_{21} | 1.0 | 1.009 | 0.009 | 0.178 | 0.176 | 0.177 | 95.7 | |
| | | φ_{22} | 1.0 | 1.000 | 0.000 | 0.285 | 0.286 | 0.286 | 94.8 | |
| λ_{12} | | $\log \beta_3$ | 0.0 | 0.009 | 0.009 | 0.154 | 0.145 | 0.145 | 94.2 | |
| | | $\log \alpha_3$ | 1.0 | 1.037 | 0.037 | 0.267 | 0.269 | 0.271 | 95.3 | |
| | | φ_{31} | 1.0 | 1.008 | 0.008 | 0.232 | 0.209 | 0.209 | 96.4 | |
| | | φ_{32} | 0.5 | 1.018 | 0.018 | 0.351 | 0.348 | 0.348 | 94.8 | |
| $n = 1000$ | | λ_1 | $\log \theta$ | 0.0 | 0.024 | 0.024 | 0.418 | 0.391 | 0.392 | 96.2 |
| | | | $\log \beta_1$ | 0.0 | 0.010 | 0.010 | 0.063 | 0.058 | 0.059 | 95.8 |
| | $\log \alpha_1$ | | 1.0 | 1.038 | 0.038 | 0.252 | 0.241 | 0.244 | 96.2 | |
| | φ_{11} | | 1.0 | 1.009 | 0.009 | 0.127 | 0.122 | 0.122 | 95.7 | |
| | φ_{12} | | 1.0 | 1.012 | 0.012 | 0.201 | 0.195 | 0.195 | 95.9 | |
| | λ_2 | $\log \beta_2$ | 0.0 | 0.008 | 0.008 | 0.063 | 0.060 | 0.060 | 95.2 | |
| | | $\log \alpha_2$ | 1.0 | 1.034 | 0.034 | 0.251 | 0.243 | 0.246 | 95.9 | |
| | | φ_{21} | 1.0 | 1.008 | 0.008 | 0.127 | 0.123 | 0.124 | 96.3 | |
| | | φ_{22} | 1.0 | 1.011 | 0.011 | 0.201 | 0.204 | 0.205 | 94.5 | |
| | λ_{12} | $\log \beta_3$ | 0.0 | 0.012 | 0.012 | 0.111 | 0.104 | 0.105 | 94.9 | |
| | | $\log \alpha_3$ | 1.0 | 1.029 | 0.029 | 0.189 | 0.183 | 0.185 | 96.2 | |
| | | φ_{31} | 1.0 | 1.014 | 0.014 | 0.165 | 0.161 | 0.161 | 95.2 | |
| | | φ_{32} | 0.5 | 1.013 | 0.013 | 0.248 | 0.243 | 0.244 | 95.4 | |

mais (tempo médio de acompanhamento = 6,5 anos). O objetivo científico aqui é avaliar os efeitos das covariáveis e do tratamento nas taxas de eventos terminais e não terminais, e a dependência entre esses eventos. Após a exclusão de indivíduos com dados incompletos e tempos de observação ausentes, temos um subconjunto de $n = 888$ pacientes com aproximadamente 50% de censura. Para cada paciente, temos t_{j1} : tempo até a recorrência do tumor(em anos), que pode ser atribuído a células tumorais com componente de metástase deixadas ativas após um tratamento inicial, t_{j2} : tempo até a morte (em anos) e x_j : tratamento (somente observação, Lev, Lev + 5-FU), $j = 1, \dots, 888$. O código R para implementação do nosso modelo está disponível no link do GitHub: <<https://github.com/bandyopd/PVF-Semicompeting>>.

No logaritmo da função de verossimilhança, dada em (4.30), consideramos a função de risco base do modelo Weibull, especificada por $\lambda_{0i}(t) = \alpha_i \beta_i t^{\beta_i - 1}$, $i = 1, 2, 3$ e o termo que incorpora as covariáveis é $\phi_{ij} = \exp(\varphi_{i1} x_{j1} + \varphi_{i2} x_{j2})$, $j = 1, \dots, n$, $i = 1, 2, 3$. Os efeitos da covariável (tratamento) são definidos por meio de variáveis dummy como:

$$x_{11} = \begin{cases} 1, & \text{se Levamisole;} \\ 0, & \text{caso contrário} \end{cases} \quad \text{e } x_{12} = \begin{cases} 1, & \text{se Levamisole+5-FU;} \\ 0, & \text{caso contrário.} \end{cases}$$

Ajustamos nosso modelo doença-morte com a fragilidade do PVF, com os casos particulares: Inversa

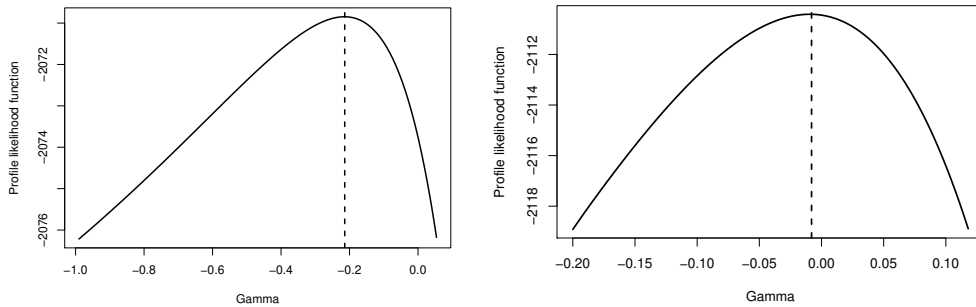


Figura 7 – Função de verossimilhança perfilada de γ . Modelo geral (painel esquerdo), modelo restrito (painel direito)

Gaussiana (IG, $\gamma = 0, 5$) e Gamma ($\gamma \rightarrow 0$). Conforme discutido na Seção 4.3, para estimar γ , O parâmetro γ foi estimado via verossimilhança perfilada; veja a Figura 7 onde são mostrados os respectivos gráficos da função de log-verossimilhança perfilada de γ correspondentes aos modelos geral e restrito. Os Estimadores de máxima verossimilhança (EMV), Erro padrão (EP) e o Intervalo de 95% de confiança (IC(95%)) do modelo proposto e submodelos são listados na Tabela 5, tanto para o modelo geral quanto para o modelo restrito. Em ambos os casos, observamos forte evidência de dependência entre o tempo de recorrência e o óbito (demostrado pela estimativa de θ). Observe que, para os três modelos de fragilidade, $\hat{\theta}$ é menor no modelo geral, indicando que parte da dependência entre T_1 e T_2 é capturada pelos diversos riscos de base. Também observamos que embora o tratamento com somente Levamisol não tenha efeito significativo sobre o risco de recorrência, a combinação (Lev + 5-FU) tem. Além disso, ambos os regimes não têm efeito significativo sobre o risco de morte, com ou sem recorrência tumoral, conforme revelado em ambos os modelos.

Tabela 5 – Análise dos dados de câncer de cólon, assumindo os modelos de fragilidade PVF, Inverse Gaussian (IG) e Gamma.

| Cov. | Par. | PVF | | | IG | | | Gamma | | | |
|----------------|------------------|---|------|----------------|--------|-------|----------------|--------|------|----------------|---|
| | | EMV | EP | IC (95%) | EMV | EP | IC (95%) | EMV | EP | IC (95%) | |
| | γ | -0.214 | - | - | - | - | - | - | - | - | - |
| | θ | 4.164 | 1.04 | (2.12; 6.20) | 9.151 | 2.96 | (3.36; 14.94) | 6.364 | 0.61 | (5.18; 7.55) | - |
| | β_1 | 1.764 | 0.12 | (1.54; 1.99) | 1.169 | 0.07 | (1.03; 1.31) | 1.875 | 0.11 | (1.66; 2.09) | - |
| | $\log(\alpha_1)$ | -0.443 | 0.24 | (-0.90; 0.02) | -0.555 | 0.20 | (-0.95; -0.16) | -0.150 | 0.21 | (-0.56; 0.26) | - |
| x_{Lev} | φ_{1_1} | 0.023 | 0.28 | (-0.53; 0.57) | -0.054 | 0.18 | (-0.40; 0.30) | 0.025 | 0.27 | (-0.50; 0.55) | - |
| $x_{Lev+5-FU}$ | φ_{1_2} | -0.585 | 0.29 | (-1.15; -0.02) | -0.827 | 0.19 | (-1.20; -0.45) | -0.747 | 0.28 | (-1.30; -0.19) | - |
| | | Taxa de transição para morte | | | | | | | | | |
| | β_2 | 2.543 | 0.22 | (2.11; 2.97) | 1.636 | 0.17 | (1.30; 1.97) | 2.597 | 0.23 | (2.15; 3.04) | - |
| | $\log(\alpha_2)$ | -3.676 | 0.41 | (-4.47; -2.88) | -3.654 | 0.39 | (-4.42; -2.89) | -3.313 | 0.39 | (-4.09; -2.54) | - |
| x_{Lev} | φ_{2_1} | -0.193 | 0.50 | (-1.18; 0.79) | -0.298 | 0.44 | (-1.17; 0.57) | -0.207 | 0.50 | (-1.18; 0.77) | - |
| $x_{Lev+5-FU}$ | φ_{2_2} | -0.217 | 0.47 | (-1.14; 0.71) | -0.389 | 0.41 | (-1.19; 0.41) | -0.385 | 0.47 | (-1.31; 0.54) | - |
| | | Taxa de transição da recorrência para a morte | | | | | | | | | |
| | β_3 | 1.933 | 0.15 | (1.64; 2.23) | 1.785 | 0.14 | (1.52; 2.05) | 2.222 | 0.14 | (1.94; 2.50) | - |
| | $\log(\alpha_3)$ | -1.590 | 0.22 | (-2.02; -1.16) | -1.136 | 0.20 | (-1.53; -0.75) | -1.578 | 0.20 | (-1.98; -1.18) | - |
| x_{Lev} | φ_{3_1} | 0.138 | 0.29 | (-0.43; 0.71) | 0.099 | 0.21 | (-0.32; 0.52) | 0.173 | 0.27 | (-0.36; 0.70) | - |
| $x_{Lev+5-FU}$ | φ_{3_2} | 0.188 | 0.30 | (-0.41; 0.78) | -0.022 | 0.24 | (-0.49; 0.44) | 0.076 | 0.29 | (-0.49; 0.64) | - |
| | | Taxa de transição para recorrência | | | | | | | | | |
| | γ | -0.01 | - | - | - | - | - | - | - | - | - |
| | θ | 8.92 | 1.32 | (6.35; 11.50) | 5.02* | 0.63* | (3.79; 6.24)* | 9.26 | 0.63 | (8.02; 10.50) | - |
| | β_1 | 2.44 | 0.10 | (2.24; 2.65) | 1.74 | 0.06 | (1.61; 1.86) | 2.45 | 0.1 | (2.26; 2.65) | - |
| | $\log(\alpha_1)$ | 0.43 | 0.26 | (-0.08; 0.93) | 0.99 | 0.60 | (0.18; 2.17) | 0.46 | 0.24 | (-0.01; 0.92) | - |
| x_{Lev} | φ_{1_1} | 0.11 | 0.32 | (-0.51; 0.74) | -0.12 | 0.21 | (-0.53; 0.30) | 0.11 | 0.32 | (-0.52; 0.74) | - |
| $x_{Lev+5-FU}$ | φ_{1_2} | -0.67 | 0.34 | (-1.34; 0.01) | -1.22 | 0.23 | (-1.68; -0.76) | -0.69 | 0.34 | (-1.36; -0.02) | - |
| | | Taxa de transição para morte | | | | | | | | | |
| | β_2 | 2.94 | 0.13 | (2.69; 3.19) | 2.31 | 0.09 | (2.14; 2.48) | 2.95 | 0.12 | (2.72; 3.19) | - |
| | $\log(\alpha_2)$ | -1.88 | 0.24 | (-2.35; -1.41) | -0.71 | 0.59 | (-1.87; 0.45) | -1.85 | 0.23 | (-2.31; -1.40) | - |
| x_{Lev} | φ_{2_1} | 0.22 | 0.32 | (-0.41; 0.85) | -0.06 | 0.22 | (-0.50; 0.38) | 0.22 | 0.33 | (-0.42; 0.86) | - |
| $x_{Lev+5-FU}$ | φ_{2_2} | 0.01 | 0.35 | (-0.67; 0.69) | -0.77 | 0.24 | (-1.24; -0.30) | -0.01 | 0.35 | (-0.69; 0.67) | - |

* As estimativas são para $\log(\theta)$

Na Tabela 6 aplicamos os conhecidos critérios de seleção de modelos candidatos, Akaike information criterion (AIC) (AKAIKE, 1974) e o Bayesian information criterion (BIC) (GELFAND; DEY, 1994). Para o modelo restrito, a fragilidade Gamma fornece um ajuste (marginalmente) melhor, também revelado na Tabela 5 (já que a estimativa $\hat{\gamma}$ é muito próxima de zero). Observe que, embora as Log-verossimilhança para PVF e Gamma fossem quase idênticas, o Δ BIC (diferença entre os BICs) foi mais proeminente (em comparação com Δ AIC), mas não excedendo a regra ≥ 10 (ANDERSON; BURNHAM, 2004) para gerar algum apoio substancial para o modelo Gamma. Para o modelo geral, tanto Δ AIC quanto Δ BIC foram < 10 , implicando consideravelmente menos suporte para o modelo Gamma. Embora ambos os modelos possam ser escolhidos, consideramos o modelo PVF/Pc em nossa análise subsequente. Como a estimativa do parâmetro γ é negativo, podemos sugerir que o modelo doença-morte com fragilidade compartilhada distribuição Poisson composta (Pc) se ajusta melhor aos dados. Portanto, devido a uma propriedade da fragilidade Pc a proporção estimada de pacientes que permanecem livres da doença, ou seja, \hat{p}_0 (dado em (4.22)) é de 0,2561, o que implica que após um longo período de tempo, aproximadamente 25% dos pacientes podem não apresentar nenhum tumor recorrência nem morte por câncer de cólon após ressecção completa do tumor cirurgicamente.

Tabela 6 – Log-Verossimilhança, AIC e BIC para os modelos ajustados.

| Model | Modelo geral | | | Modelo restrito | | |
|--------|---------------------|----------|----------|---------------------|----------|----------|
| | Log-Verossimilhança | AIC | BIC | Log-Verossimilhança | AIC | BIC |
| PVF/cP | -2070.847 | 4167.695 | 4230.538 | -2110.411 | 4240.823 | 4289.164 |
| IG | -2135.397 | 4296.793 | 4359.637 | -2304.823 | 4627.645 | 4671.152 |
| Gamma | -2073.759 | 4173.517 | 4236.361 | -2110.445 | 4238.890 | 4282.397 |

Na figura 8 mostramos a razão de risco explicativa condicional (**EHR**) incluindo covariáveis, dada por

$$\mathbf{EHR} = \frac{\lambda_{03}(t_2)}{\lambda_{02}(t_2)} \exp [(\boldsymbol{\varphi}_3^T - \boldsymbol{\varphi}_2^T)\mathbf{x}], \quad t_2 > t_1. \quad (4.31)$$

Aqui, o **EHR** descreve como o risco de morte muda ao longo do tempo, dado que a recorrência do tumor ocorreu no momento t_1 (LEE *et al.*, 2015). Em particular, $\mathbf{EHR} > 1$ implica que a recorrência do tumor tem um efeito sobre o risco de morte por câncer de cólon. Embora o **EHR** não dependa de t_1 devido à sua estrutura de Markov, sua interpretação depende da condição $t_2 > t_1$, para todo t_1 fixo. Para uma melhor visualização, ampliamos a curva **EHR** (veja a Figura 8, painel direito). Observamos que o **EHR** para pacientes sob os regimes concorrentes (somente observação, apenas Lev e Lev + 5-FU) é 1 em aproximadamente $t = 19,5, 33,5$ e 38 anos, respectivamente. Isso implica que, nesses instantes, o tempo de recorrência do tumor não tem efeito sobre a morte.

Para estudar os resultados visualmente, as taxas de transição marginal e probabilidades de transição marginal são obtidas usando as estimativas da Tabela 5. Desta forma, na Figura 9, traçamos as taxas de transição marginal, $\hat{\lambda}_1(t)$ e $\hat{\lambda}_2(t)$, estratificadas por tratamentos. Observamos que, no primeiro ano, os pacientes em regime de Lev+ 5-FU apresentam menor risco de recorrência tumoral, em comparação aos pacientes dos outros dois grupos. Esse resultado é consistente com a Tabela 5, onde observamos um efeito significativo apenas para o grupo Lev + 5-FU na recorrência do tumor. No entanto, a taxa de transição para a morte (de $\hat{\lambda}_2(t)$) é maior com Lev + 5-FU, embora a Tabela 5 revele que os tipos de

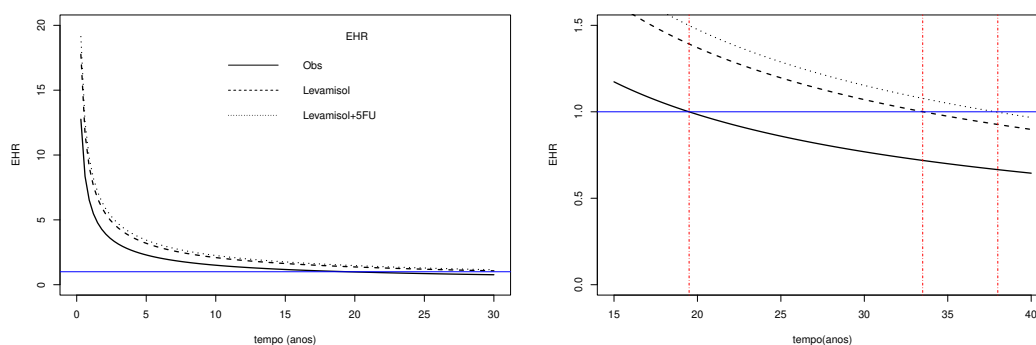


Figura 8 – razão de risco explicativa condicional estimado, \hat{EHR} , para cada tratamento.

regime não têm efeitos significativos sobre a morte. Mas, a taxa de transição de morte é muito menor comparada com a taxa de transição para a recorrência, e muito perto de zero.

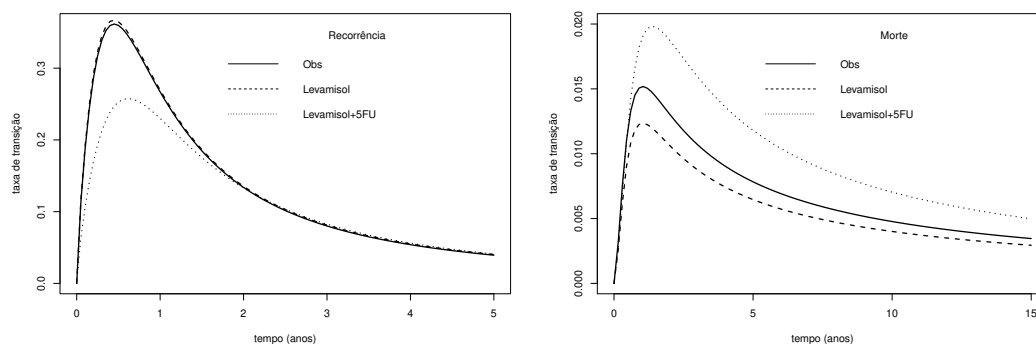


Figura 9 – Taxas de transição estimadas para a recorrência, $\hat{\lambda}_1(t)$, (painel esquerdo) e para a morte, $\hat{\lambda}_2(t)$, (painel direito), estratificado pela variável explicativa tratamento.

As taxas de transição estimadas para morte após a recorrência do tumor estão representadas na Figura 10. Observe que essa taxa de transição depende do tempo até a morte t_2 , além do tempo desde a recorrência do tumor, e os gráficos variam com os tempos de recorrência t_1 . Se compararmos essas taxas de transição com as taxas de transição para morte na Figura 9 (painel direito), as curvas sem recorrência do tumor (Figura 9 painel direito) ficam significativamente mais baixas do que as curvas com recorrência do tumor (Figura 10), implicando forte evidência do efeito do evento não terminal (recorrência tumoral) sobre o evento terminal (morte), e a dependência entre eles. Podemos inferir que um paciente com recidiva tumoral tem maior risco de morte do que outro sem recidiva tumoral. No entanto, também observamos que à medida que t_1 aumenta, as taxas de transição para óbito após a recorrência do tumor diminuem. Isso implica que o efeito da recorrência do tumor seja menos proeminente para os pacientes que apresentam recorrência do tumor no final do estudo, do que os pacientes que a experimentaram precocemente.

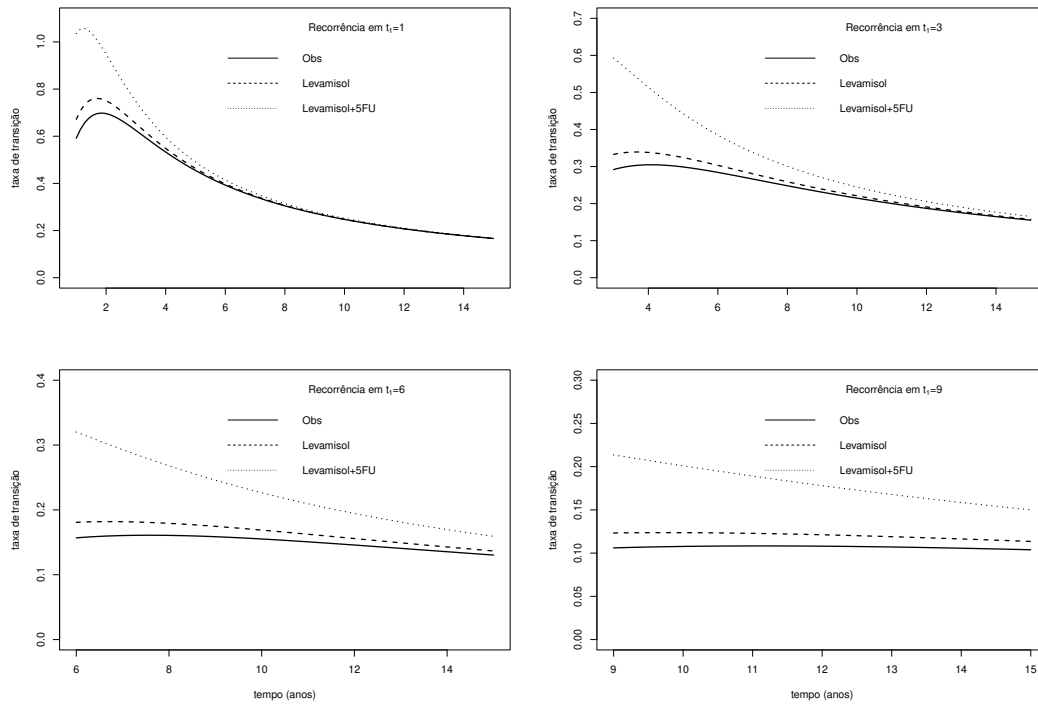


Figura 10 – Taxas de transição marginal estimadas para óbito, dado que um paciente teve recorrência do tumor em t_1 , $\hat{\lambda}_1(t_2|t_1)$, estratificado pela variável explicativa tratamento, para $t_1 = 1, 3, 6$ e 9 anos.

Finalmente, a Figura 11 apresenta as probabilidades de transição apresentadas na Subseção 4.2.2. A partir do gráfico de, $\hat{P}_{01}(t_1)$, (para recorrência do tumor, veja o painel superior esquerdo), observamos que a probabilidade de recorrência do tumor no instante t_1 é menor para o grupo Lev + 5-FU, o que corrobora com o obtido na Tabela 5. Para esses pacientes, os primeiros meses após a ressecção completa do tumor são críticos, conforme sugerido pela crescente probabilidade de recorrência até o instante t , além do qual diminui. No entanto, as probabilidades correspondentes de morte, ou seja, $\hat{P}_{02}(t_2)$ (veja o painel superior direito), estão aumentando para todos os pontos de tempo e tratamentos. Curiosamente, essa probabilidade é a maior para o grupo Lev + 5-FU. O gráfico no painel inferior esquerdo compara $\hat{P}_{01}(t_1)$ a $\hat{P}_{02}(t_1)$. Observamos que um paciente tem maior probabilidade de recidiva do que morte nos primeiros 2,5 anos após a ressecção, a partir dos quais a situação se inverte. Finalmente, a Figura 11 (canto inferior direito) mostra $\hat{P}_{12}(t|t_1)$, a probabilidade de transição para morte, dado que o paciente estava em estado de recorrência no momento t_1 , para várias opções de t_1 . Comparando com $\hat{P}_{02}(t_2)$ (canto superior direito), inferimos que um paciente com recorrência do tumor no momento t_1 tem maior probabilidade de evoluir para óbito, em comparação com outro paciente sem recorrência. A dependência entre a recorrência e a morte se manifesta agora através das probabilidades de transição.

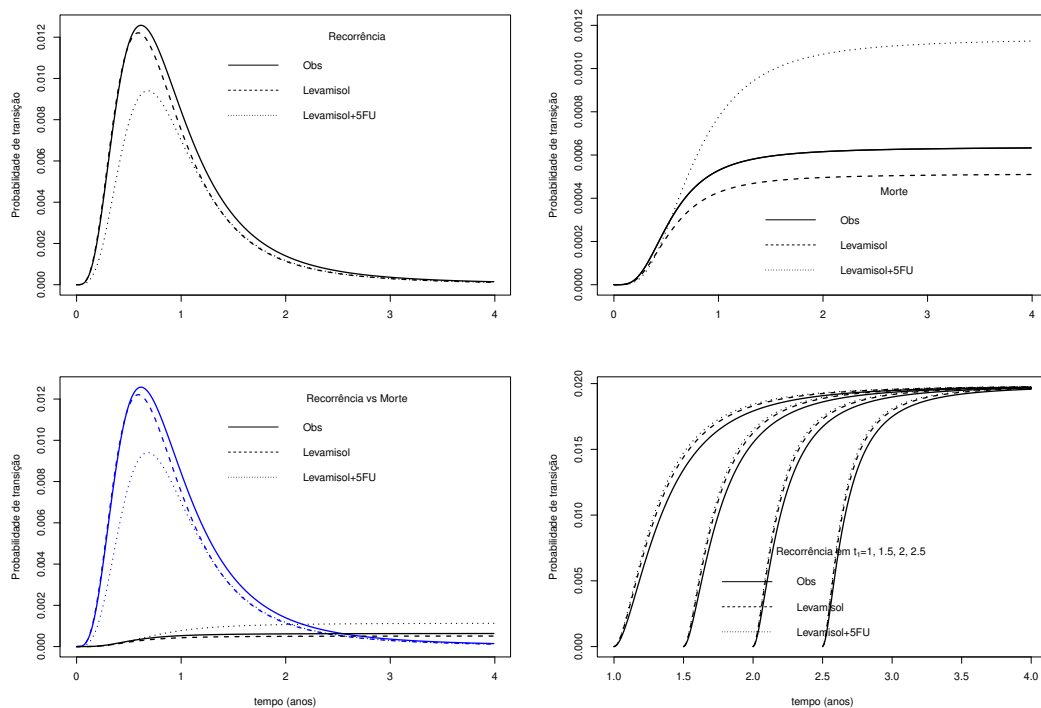


Figura 11 – Probabilidades de transição marginal estimadas para a recorrência, $\hat{P}_1(t)$, (painel superior esquerdo) e para a morte, $\hat{P}_2(t)$, (painel superior direito) no instante t , estratificado pela variável explicativa tratamento. Comparação entre probabilidade de transição marginal para recorrência (azul) e morte (preto) (painel inferior esquerdo). Probabilidade de transição marginal estimada para a morte, dado que um paciente estava no estado de recorrência no instante t_1 , $\hat{P}_{12}(t|t_1)$, (painel inferior direito).

MODELO DE SOBREVIVÊNCIA BIVARIADO INDUZIDO POR FRAGILIDADE DISCRETA

5.1 Introdução

Os modelos de sobrevivência de longa duração (também conhecidos como modelos de sobrevivência com fração de cura ou taxa de cura) mereceram muito interesse na literatura sob os títulos da análise de confiabilidade e sobrevivência. Os modelos de longa duração cobrem as situações em que existem unidades de amostragem não susceptíveis à ocorrência do evento de interesse. A proporção de tais unidades em muitos casos é denominada como a proporção ou fração de curados. Em estudos clínicos, o evento de interesse pode ser a morte de um paciente ou a recorrência de um tumor (que pode ser atribuído a células tumorais com componente de metástase deixadas ativas após um tratamento inicial). Por outro lado, em um sistema financeiro onde os eventos de interesse são o tempo até o cancelamento ou abandono de dois produtos, por exemplo cartões de crédito, podem existir diferentes causas de risco para que os clientes abandonem os produtos. No entanto, existe uma proporção de clientes fiéis que preferem continuar com os produtos oferecidos pela empresa financeira mesmo após um longo tempo. Nestas situações específicas, é apropriado considerar distribuições discretas de fragilidade, devido a que admitem uma fragilidade nula o que implica que existam unidades amostrais não susceptíveis ao evento de interesse, que pode ser por exemplo a recorrência de uma doença em um ensaio clínico ou o abandono de produtos de um sistema financeiro.

Nesse sentido, agora existe uma vasta literatura sobre “modelos com fração de cura ou de longa duração” para dados de sobrevivência, que foi desenvolvido, embora a maioria deles decorra de um dos modelos de taxa de cura, de 2 componentes de Mistura de [Berkson e Gage \(1952\)](#) ou o modelo de Risco acumulado limitado (ou modelo BCH) de [Yakovlev, Tsodikov e Asselain \(1996\)](#). Vários outros pesquisadores contribuíram para essa área, como os estudos apresentados em [Tsodikov, Ibrahim e Yakovlev \(2003\)](#), [Rodrigues et al. \(2009\)](#), [Ortega et al. \(2015\)](#), [Cordeiro et al. \(2016\)](#), [Koutras e Milienos \(2017\)](#), [Souza et al. \(2017\)](#), [Barriga et al. \(2019\)](#).

Várias extensões de modelos de longa duração foram desenvolvidas na literatura. No entanto, pouca atenção foi dada à pesquisa em modelos multivariados de taxa de cura. No âmbito frequentista, Chatterjee e Shih (2001) propuseram uma abordagem marginal usando modelos de cópulas bivariadas. Price e Manatunga (2001) impuseram a fragilidade para explicar a correlação e consideraram a estimativa da máxima verossimilhança sob uma premissa de modelo paramétrico. Ambos os métodos foram baseados no modelo de taxa de cura por mistura. Gallardo, Bolfarine e Pedroso-De-Lima (2016) considerando dados de sobrevivência agrupados, incluíram um efeito aleatório na função de sobrevivência dos indivíduos em risco e na proporção de curados, assumindo uma distribuição bivariada para o efeito em cada grupo. Kim (2017) propôs um modelo de mistura bivariado para analisar dados censurados em intervalos bivariados com taxa de cura.

No contexto bayesiano, o modelo Bounded cumulative hazard (BCH) introduzido por Tsodikov, Yakovlev e Asselain (1996) foi estendido para a modelagem de dados de sobrevivência multivariada. Por outro lado, para modelar a dependência entre variáveis latentes Chen, Ibrahim e Sinha (2002) introduziram um componente de fragilidade com distribuição positiva estável (PE) no parâmetro da distribuição de variáveis latentes. Enquanto que Cancho *et al.* (2016) propôs um modelo de sobrevivência multivariado com fração de cura, assumindo que há m tipos de causas (ou riscos) latentes e que o número de causas de cada tipo segue uma distribuição Poisson multivariada. Martins, Silva e Andreozzi (2017) desenvolveram um modelo conjunto flexível para os resultados longitudinais e de sobrevivência sob a presença de uma fração espacial de sobreviventes de longa duração.

Neste trabalho, propomos um novo modelo de sobrevivência multivariada com fração sobrevivente em termos de modelos de fragilidade, cuja distribuição de fragilidade considerado neste trabalho estende o modelo considerado por Chen, Ibrahim e Sinha (2002). O novo modelo também estende o modelo univariado proposto por Cancho, Rodrigues e Castro (2011). Além disso, construímos um modelo de regressão para avaliar os efeitos das covariáveis na proporção de unidades amostrais não susceptíveis ao evento de interesse. Por fim, apresentamos um estudo de simulação para avaliar o desempenho do método de estimação de máxima verossimilhança.

5.2 Descrição do modelo

Nesta seção, derivamos e discutimos a obtenção do modelo bivariado de longa duração, isto é, com proporção de indivíduos não susceptíveis ao evento de interesse. A extensão multivariada pode ser obtida de forma imediata. Sejam T_1 e T_2 , os tempos de sobrevivência que podem ser de dois indivíduos relacionados (por exemplo indivíduos que pertencem a um mesmo grupo), ou um indivíduo com tempos de dois eventos de interesse, por exemplo, T_1 : tempo até abandono do produto A e T_2 : tempo até abandono do produto B, oferecido por alguma empresa.

Para um indivíduo arbitrário na população em estudo, assumimos que a distribuição T_k ($k = 1, 2$) pode ser representado por um modelo básico de fragilidade, da seguinte forma,

$$\lambda_k(t_k|Z_k) = Z_k \lambda_{0k}(t_k), \quad k = 1, 2 \quad (5.1)$$

onde $\lambda_{0k}(t_k)$ é a função de risco de base para T_k . Considere Z_k uma variável aleatória discreta que denota o número não observável de causas de risco para o k -ésimo evento de interesse. Assumimos que Z_k ($k = 1, 2$) tem distribuição Poisson com média $\omega \rho_k$ onde $\rho_k > 0$ é constante, enquanto ω é uma componente de fragilidade no modelo para a qual consideramos a distribuição PVF, descrita na Seção 3.4.5, com média 1 e variância $1 - \gamma := \theta > 0$ (HOUGAARD; HOUGAARD, 2000). Z_1 e Z_2 são condicionalmente independentes (dada ω). Assim, marginalmente cada variável, Z_k , tem $\mathbb{E}[Z_k] = \rho_k$, $\text{Var}(Z_k) = \rho_k(1 + \theta \rho_k)$ e $\text{Cov}(Z_1, Z_2) = \theta \rho_1 \rho_2$, para $k = 1, 2$. A quantidade ω é uma componente de fragilidade no modelo a qual induz uma correlação entre as variáveis latentes Z_1 e Z_2 , assim como a heterogeneidade não observada de cada variável Z_k , $k = 1, 2$. Sob a suposições dadas acima, podemos mostrar que a função de sobrevivência bivariada condicional dado $W = \omega$, é dado por

$$S(t_1, t_2|\omega) = \exp\{-\omega(\rho_1(F_{01}(t_1)) + \rho_2(F_{02}(t_2)))\}, \quad (5.2)$$

onde $F_{0k}(t_k) = 1 - \exp\{-\int_0^{t_k} \lambda_{0k}(u) du\}$ é função de distribuição acumulada de base para o tempo T_k , $k = 1, 2$. A prova de (5.2) encontra-se no Apêndice C.1. Usando a transformada de Laplace da distribuição PVF, dada em (3.60), a função de sobrevivência não condicional conjunta para T_1 e T_2 é dada por

$$S(t_1, t_2) = \exp\left\{-\frac{\mu}{\gamma} [(\sigma + \rho_1 F_{01}(t_1) + \rho_2 F_{02}(t_2))^\gamma - \sigma^\gamma]\right\}, \quad t_1 > 0, \quad t_2 > 0. \quad (5.3)$$

De (5.3), as funções de sobrevivência marginal são dadas por

$$S(t_k) = \exp\left\{-\frac{\mu}{\gamma} [(\sigma + \rho_k F_{0k}(t_k))^\gamma - \sigma^\gamma]\right\}, \quad k = 1, 2. \quad (5.4)$$

È importante notar que em (5.4) quando $\gamma \rightarrow 0$ e $\mu = \sigma = 1/\theta$ a distribuição marginal tem a estrutura do modelo univariado proposto por Cancho, Rodrigues e Castro (2011), dada por $S(t) = (1 + \theta \rho F_0(t))^{-1/\theta}$, a qual corresponde a uma distribuição gama. Assim, o modelo proposto estende o modelo de sobrevivência univariado com fração de cura proposto por Cancho, Rodrigues e Castro (2011). Além disso, nosso modelo estende ao modelo proposto por Chen, Ibrahim e Sinha (2002), pois para $\sigma = 0$ e $\mu = \gamma$ em (5.3) o modelo corresponde à função de sobrevivência conjunta considerando a fragilidade positiva estável, proposta por Chen, Ibrahim e Sinha (2002).

Logo, considerando a suposição $\mu = \sigma = 1$ em (5.3), de modo que $\mathbb{E}[\omega] = 1$ e $\text{Var}(\omega) = 1 - \gamma = \theta > 0$, temos que a função de sobrevivência marginal conjunta é dada por

$$S(t_1, t_2) = \exp\left\{-\frac{1}{\gamma} [(1 + \rho_1 F_{01}(t_1) + \rho_2 F_{02}(t_2))^\gamma - 1]\right\}, \quad t_1 > 0, \quad t_2 > 0. \quad (5.5)$$

A função de sobrevivência $S(t_1, t_2)$, em (5.5), não é própria, isto é,

$$\lim_{t_1, t_2 \rightarrow \infty} S(t_1, t_2) = p_{00} = \exp\left\{-\frac{1}{\gamma} [(1 + \rho_1 + \rho_2)^\gamma - 1]\right\} > 0 \quad (5.6)$$

onde p_{00} é a função de cura conjunta. De (5.5), as funções de sobrevivência marginal são dadas por

$$S(t_k) = \exp\left\{-\frac{1}{\gamma} [(1 + \rho_k F_{0k}(t_k))^\gamma - 1]\right\}, \quad k = 1, 2. \quad (5.7)$$

A equação (5.7) indica que a função de sobrevivência tem uma estrutura de um modelo de sobrevivência com probabilidade de cura,

$$p_{0k} = \exp \left\{ -\frac{1}{\gamma} [(1 + \rho_k)^\gamma - 1] \right\}, \quad k = 1, 2. \quad (5.8)$$

Por outro lado, a probabilidade de que um indivíduo em risco do tipo k ($k = 1, 2$), é imune ou curado da doença dado que sobreviveu até o tempo $t > 0$, depois do tratamento, denotado por $\pi_k(t)$, é dado por

$$\pi_k(t) = \mathbb{P}[Z_k = 0 | T_k > t] = \frac{\mathbb{P}[T_k > t | Z_k = 0] \mathbb{P}[Z_k = 0]}{S_k(t)},$$

Note que $\mathbb{P}[T_k > t | Z_k = 0] = 1$ e $\mathbb{P}[Z_k = 0] = \lim_{t_k \rightarrow \infty} \mathbb{P}[T_k > t] = p_{0k}$ que está dado em (5.8), com isto e usando (5.7), temos que

$$\pi_k(t) = \left[\frac{\exp \{ (1 + \rho_k F_{0k}(t))^\gamma \}}{\exp \{ (1 + \rho_k)^\gamma \}} \right]^{\frac{1}{\gamma}}, \quad k = 1, 2. \quad (5.9)$$

A probabilidade em (5.9) é uma função crescente em t . Para $t = 0$, a probabilidade é igual a $\pi_k(0) = p_{0k}$, corresponde a nenhuma informação sobre a imunidade de um indivíduo, exceto a probabilidade geral de estar imune. A probabilidade tende a um quando $t \rightarrow \infty$, correspondendo certamente à imunidade se o tempo de vida do indivíduo for muito longa.

A partir de (5.5) a distribuição bivariada de T_1 e T_2 tem função de densidade conjunta dada por

$$f(t_1, t_2) = \rho_1 \rho_2 f_{01}(t_1) f_{02}(t_2) S(t_1, t_2) (1 + \rho_1 F_{01}(t_1) + \rho_2 F_{02}(t_2))^{\gamma-2} \\ \times [1 - \gamma + (1 + \rho_1 F_{01}(t_1) + \rho_2 F_{02}(t_2))^\gamma], \quad (5.10)$$

onde $\theta > 0$, $\rho_k > 0$ e $f_{0k}(t_k) = \frac{\partial F_{0k}(t_k)}{\partial t_k}$ para $k = 1, 2$. A Figura 12 mostra as funções de densidade e seus respectivos gráficos de contorno (5.10) para diferentes valores do parâmetro θ do novo modelo, mostrando a flexibilidade do modelo em termos do parâmetro de dependência $\theta > 0$ ($\theta = 1 - \gamma$). Por outro lado, seguindo Clayton (1978), podemos calcular a medida de associação local entre T_1 e T_2 , $\vartheta^*(t_1, t_2)$, definida em (3.26), isto é

$$\vartheta^*(t_1, t_2) = 1 + (1 - \gamma) (1 + \rho_1 F_{01}(t_1) + \rho_2 F_{02}(t_2))^{-\gamma}. \quad (5.11)$$

Observe que, de forma similar ao modelo de riscos semicompetitivos, $\vartheta^*(t_1, t_2) = 1 + \theta = 2 - \gamma$ é constante em (t_1, t_2) , quando $\gamma \rightarrow 0$, isto corresponde à fragilidade Gama, veja a Figura 13 (painel central). Porém, para a fragilidade PVF, $\vartheta^*(t_1, t_2)$ depende de (t_1, t_2) , a dependência é menor quando os tempos crescem, além disso $\vartheta^*(t_1, t_2) \rightarrow 1 + (1 - \gamma)(1 + \rho_1 + \rho_2)^{-\gamma}$ quando $t_1 \rightarrow \infty$ e $t_2 \rightarrow \infty$. Isso significa que para um período suficientemente grande a dependência de T_1 e T_2 se torna constante, veja a Figura 13 (painel direito). Por outro lado, se $\gamma < 0$, a medida de associação local entre T_1 e T_2 do modelo PVF corresponde ao modelo Pc, na qual $\vartheta^*(t_1, t_2) \rightarrow 1 + (1 - \gamma)(1 + \rho_1 + \rho_2)^{-\gamma}$ quando $t_1 \rightarrow \infty$ e $t_2 \rightarrow \infty$, isto significa que a dependência cresce quando os tempos t_1 e t_2 crescem, mas a partir certo tempo a dependência de T_1 e T_2 também se torna constante, ver Figura 13 (painel esquerdo).

O Teorema a seguir evidencia uma forma de obter uma nova classe de modelos de sobrevivência multivariados com fração sobrevivente em termos da transformada de Laplace da fragilidade da distribuição.

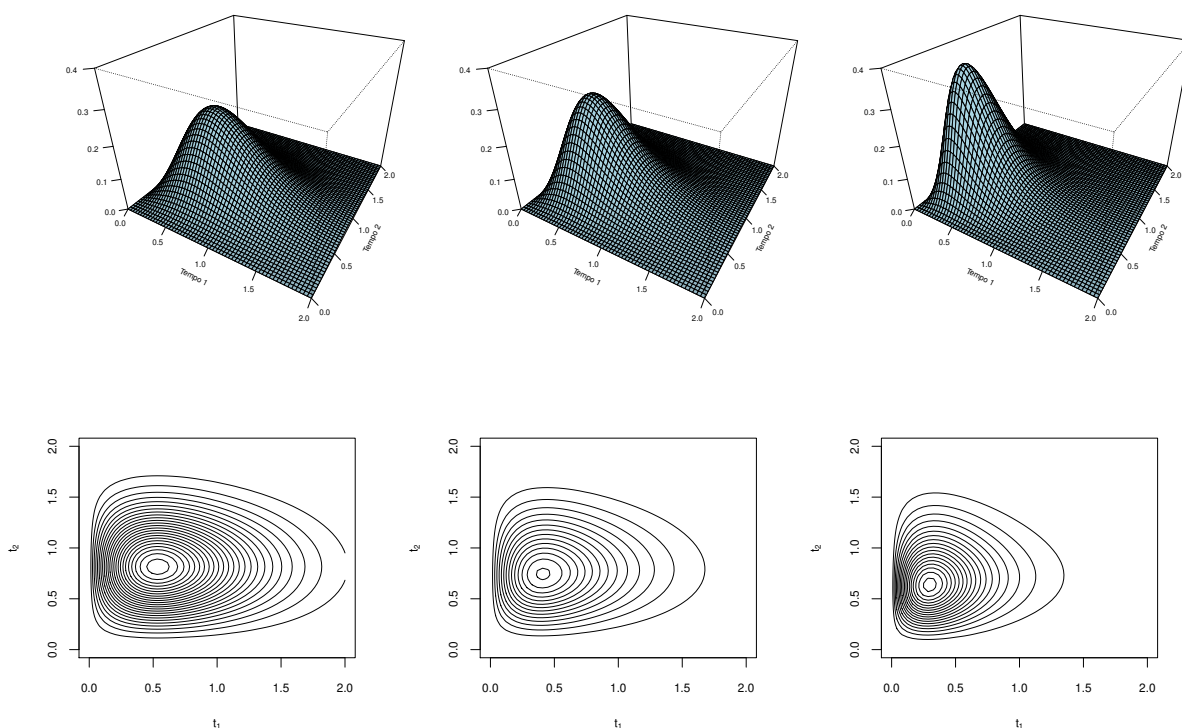


Figura 12 – Função de densidade dada em (5.10) e seu gráfico de contorno do novo modelo para os parâmetros $\theta = 0$ (painel esquerdo), $\theta = 1$ (painel central), $\theta = 4$ (painel direito), $\rho_1 = 1$, $\rho_2 = 0.4$ e função de sobrevivência $S_{01}(t_1) = e^{-t_1^2}$ e $S_{01}(t_2) = e^{-t_2^3}$.

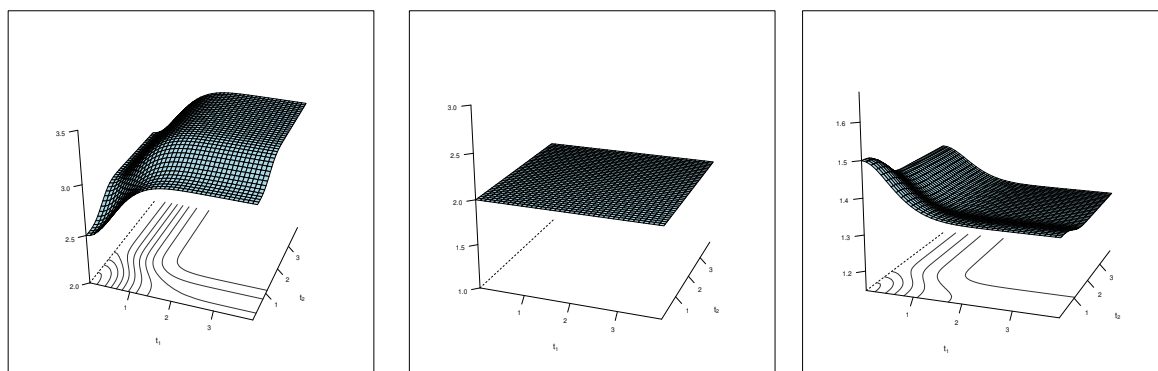


Figura 13 – Medida de associação local entre T_1 e T_2 , $\vartheta^*(t_1, t_2)$, do novo modelo para os parâmetros $\gamma = -0.5$ e $\theta = 1.5$ (painel esquerdo), $\gamma = 0$ e $\theta = 1$ (painel central), $\gamma = 0.5$ e $\theta = 0.5$ (painel direito), $\rho_1 = 1$, $\rho_2 = 0.4$ e função de sobrevivência $F_{01}(t_1) = 1 - e^{-t_1^2}$ e $F_{02}(t_2) = 1 - e^{-t_2^3}$.

Teorema 5.2.1. Sejam T_1, \dots, T_m variáveis aleatórias que representam os m tipos de tempos observáveis para o evento de interesse, de modo que a distribuição de T_k , $k = 1, \dots, m$ podem ser representadas pelo modelo básico de fragilidade $\lambda_k(t_k|Z_k) = Z_k \lambda_{0k}(t_k)$, $k = 1, \dots, m$, onde $\lambda_{0k}(t_k)$ é a função de risco de base para T_k e Z_k é uma variável aleatória discreta com distribuição $\text{Poisson}(\omega \rho_k)$ sendo $\rho_k > 0$ uma constante real e $\omega > 0$ uma componente aleatória (fragilidade). Se a distribuição de ω tem transformada de Laplace, $\mathcal{L}_\omega(s)$, com média 1 e variância finita, $\theta > 0$ e supondo que dado $Z_k = z_k$ os T_1 e T_2 são condicionalmente

independentes, então a função de sobrevivência não condicional de (T_1, \dots, T_m) , é dada por

$$S(t_1, \dots, t_m) = \mathcal{L}_\omega \left(\sum_{k=1}^m \theta_k F_{0k}(t_k) \right), \quad (5.12)$$

onde $F_{0k}(t_k) = 1 - \exp \left\{ - \int_0^{t_k} \lambda_{0k}(u) du \right\}$ é a função de distribuição acumulada de base para o tempo T_k , para $k = 1, \dots, m$.

Prova: Para as variáveis aleatórias não negativas, T_k e as condições do Teorema 5.2.1, a função de sobrevivência condicional dado Z_k é dada por

$$S(t_k | Z_k) = \exp \left\{ - \int_0^{t_k} Z_k \lambda_{0k}(u) du \right\} = (S_{0k}(t_k))^{Z_k}, \quad (5.13)$$

onde $S_{0k}(t_k) = 1 - F_{0k}(t_k)$ é a função de sobrevivência de base da variável aleatória T_k , para $k = 1, \dots, m$. Logo, desde que dado $Z_k = z_k$, as variáveis aleatórias T_k são condicionalmente independentes e dado $W = \omega$, as variáveis aleatórias Z_k , são condicionalmente independentes, podemos escrever

$$\begin{aligned} \mathbb{P}[T_1 > t_1, T_2 > t_2, Z_1, \dots, Z_m, \omega] &= \prod_{k=1}^m (S(t_k | Z_k) \mathbb{P}[Z_k | \omega]) g(\omega) \\ &= \prod_{k=1}^m \left([S_{0k}(t_k)]^{Z_k} \mathbb{P}[Z_k | \omega] \right) g(\omega) \end{aligned}$$

onde $g(\omega)$ é a função densidade de probabilidade da componente de fragilidade ω . Integrando sobre ω e somando sobre todo Z_1, \dots, Z_m , obtemos a função de sobrevivência não condicional que dado por

$$S(t_1, \dots, t_m) = \mathbb{E}_\omega \left[\prod_{k=1}^m \mathbb{E}_{Z_k | \omega} \left[(S_{0k}(t_k))^{Z_k} \right] \right]. \quad (5.14)$$

Note que, a variável aleatória condicional $Z_k | \omega \sim \text{Poisson}(\omega \rho_k)$ e $\mathbb{E}_{Z_k | \omega} \left[(S_{0k}(t_k))^{Z_k} \right] = \psi_{Z_k | \omega}(S_{0k}(t_k)) = e^{-\omega \rho_k (1 - S_{0k}(t_k))}$ é a função geradora de probabilidade da distribuição Poisson avaliada em $S_{0k}(t_k)$, $k = 1, \dots, m$. Desta forma

$$S(t_1, \dots, t_m) = \mathbb{E}_\omega \left[e^{-\omega (\sum_{k=1}^m \rho_k (1 - S_{0k}(t_k)))} \right] = \mathcal{L}_\omega \left(\sum_{k=1}^m \theta_k F_{0k}(t_k) \right). \quad (5.15)$$

□

A versão multivariada do modelo proposto em (5.5) pode ser obtida a partir do Teorema 5.2.1. Para ser mais específico, a variável aleatória, W com distribuição PVF com média 1 e variância $\theta > 0$, tem transformada de Laplace dada em (3.51), isto implica que a função de sobrevivência multivariada é dado por

$$S(t_1, \dots, t_m) = \exp \left\{ - \frac{1}{\gamma} \left[\left(1 + \sum_{k=1}^m \rho_k (F_{0k}(t_k)) \right)^\gamma - 1 \right] \right\}, \quad t_1 > 0, \quad t_2 > 0. \quad (5.16)$$

5.3 Inferência

Nesta seção, construímos a função de verossimilhança para parâmetros do modelo de modelo de cura bivariado em (5.5) da seguinte maneira. Suponha que temos n indivíduos na amostra e seja Z_{kj} o número de riscos latentes do tipo k do evento de interesse para o j -ésimo indivíduo, com $j = 1, \dots, n$ e $k = 1, 2$. Além disso, assumimos que os Z_{kj} são variáveis aleatórias Poisson independentes com média $\omega_j \rho_k$, $j = 1, \dots, n$ e $k = 1, 2$. Também assumimos que as ω_j são variáveis aleatórias i.i.d com distribuição PVF de média 1 e variância $\theta > 0$. Seja T_{kj} o tempo de falha para o j -ésimo indivíduo e o k -ésimo tipo de evento de interesse, onde T_{kj} pode ser censurado pela direita. Suponha que a distribuição de T_{kj} , com $j = 1, \dots, n$ e $k = 1, 2$, pode ser representado por um modelo básico de fragilidade $\lambda_k(t_{kj}|Z_{kj}) = Z_{kj} \lambda_{0k}(t_{kj}|\boldsymbol{\eta}_k)$, onde $\lambda_{0k}(t_{kj}|\boldsymbol{\eta}_k)$ é a função de risco de base para T_{kj} com vetor de parâmetros $\boldsymbol{\eta}_k$. Suponha também que Z_{kj} e T_{kj} são condicionalmente independentes. Denote por C_{kj} os tempos de censura de modo que $y_{jk} = \min\{T_{kj}, C_{kj}\}$ e $\delta_{kj} = I(T_{kj} \leq C_{kj})$ é igual a 1 se T_{kj} é um tempo de falha e 0 se o tempo de falha for censurado pela direita. Assim, seja $\mathbf{y}_k = (y_{k1}, \dots, y_{kn})$, $\boldsymbol{\delta}_k = (\delta_{k1}, \dots, \delta_{kn})$, $\mathbf{Z}_k = (Z_{k1}, \dots, Z_{kn})$ com $k = 1, 2$ e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$. Logo, os dados completos são dados por $D = (n, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \mathbf{Z}_1, \mathbf{Z}_2, \boldsymbol{\omega})$, onde $\mathbf{Z}_1, \mathbf{Z}_2$ e $\boldsymbol{\omega}$ são vetores aleatórios não observáveis, então os dados observados são dados por $D_{obs} = (n, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. Adicionalmente seja $\boldsymbol{\vartheta} = (\gamma, \theta, \rho_1, \rho_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, o vetor de parâmetros a serem estimados. Pode-se provar que a função de verossimilhança para $\boldsymbol{\vartheta}$ dado os dados completos é dada por

$$L(\boldsymbol{\vartheta}|D) = \prod_{k=1}^2 \prod_{j=1}^n [S_{0k}(y_{kj}|\boldsymbol{\eta}_k)]^{Z_{kj}-\delta_{kj}} [Z_{kj} f_{0k}(y_{kj}|\boldsymbol{\eta}_k)]^{\delta_{kj}} \\ \times \prod_{k=1}^2 \exp \left\{ \sum_{j=1}^n [Z_{kj} \log(\omega_j \rho_k) - \log(Z_{kj}!) - \omega_j \rho_k] \right\} \prod_{j=1}^n g(\omega_j) \quad (5.17)$$

onde $f_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ é a função densidade de base e $S_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ é a função de sobrevivência de base. A prova deste resultado encontra-se no Apêndice D.1. Neste trabalho supomos a função densidade Weibull para $f_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ com vetor de parâmetro $\boldsymbol{\eta}_k = (\alpha_k, \beta_k)$ em que α_k é o parâmetro de escala e β_k é o parâmetro de forma, com $k = 1, 2$. A função de verossimilhança dos dados observados é obtida integrando (5.17) em relação a $(\mathbf{Z}_1, \mathbf{Z}_2, \boldsymbol{\omega})$ assumindo uma distribuição PVF para cada ω_j com média 1 e variância $\theta > 0$ e função de densidade denotadas por $g(\cdot)$. No seguinte teorema podemos ver este resultado.

Teorema 5.3.1. A função de verossimilhança, baseada nos dados observados e denotada por $L(\boldsymbol{\vartheta}|D_{obs})$, é dada por

$$L(\boldsymbol{\vartheta}|D_{obs}) = \prod_{j=1}^n (\rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1))^{\delta_{1j}} (\rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_2))^{\delta_{2j}} \\ \times \prod_{j=1}^n \exp \left\{ -\frac{1}{\gamma} [(1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^\gamma - 1] \right\} \quad (5.18) \\ \times \prod_{j=1}^n (1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^{(1-\gamma)(\delta_{1j} + \delta_{2j}) - \gamma \delta_{1j} \delta_{2j}} \\ \times \prod_{j=1}^n [1 - \gamma + (1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^\gamma]^{\delta_{1j} \delta_{2j}},$$

onde $f_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ e $F_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ são as funções de densidade e distribuição acumulada de base, respectivamente, para $j = 1, \dots, n$ e $k = 1, 2$.

A prova deste Teorema é apresentada no Apêndice D.2. Incorporamos covariáveis para o modelo de taxa de cura (5.5) através dos parâmetros de taxa de cura ρ_1 e ρ_2 . Quando as covariáveis são incorporadas, temos um parâmetro de taxa de cura diferente, ρ_{kj} , para cada indivíduo, $j = 1, \dots, n$. Seja $\mathbf{x}'_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ o vetor de p covariáveis do j -ésimo indivíduo, e seja $\boldsymbol{\varphi}_k = (\varphi_{k1}, \varphi_{k2}, \dots, \varphi_{kp})'$ o correspondente vetor de coeficientes de regressão para a variável aleatória de tempo de falha T_k , $k = 1, 2$. Relacionamos o parâmetro ρ_{kj} com as covariáveis a traves da ligação logarítmica, isto é, $\rho_{kj} = \exp\{\mathbf{x}'_j \boldsymbol{\varphi}_k\}$, de modo que a taxa de cura para o j -ésimo e o evento k é,

$$p_{0kj} = \exp \left\{ -\frac{1}{\gamma} \left[(1 + \exp\{\mathbf{x}'_j \boldsymbol{\varphi}_k\})^\gamma - 1 \right] \right\}, \quad (5.19)$$

para $j = 1, \dots, n$ e $k = 1, 2$. Seja agora $\boldsymbol{\vartheta} = (\gamma, \boldsymbol{\theta}, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, então podemos escrever o logaritmo da função verossimilhança de $\boldsymbol{\vartheta}$ dadas as observados, $\ell(\boldsymbol{\vartheta}|D_{obs})$, sob a censura não informativa, isto é

$$\begin{aligned} \ell(\boldsymbol{\vartheta}|D_{obs}) &= \sum_{j=1}^n \delta_{1j} \mathbf{x}'_j \boldsymbol{\varphi}_1 + \delta_{2j} \mathbf{x}'_j \boldsymbol{\varphi}_2 + \delta_{1j} \log(f_{01}(y_{1j}|\boldsymbol{\eta}_1)) + \delta_{1j} \log(f_{02}(y_{2j}|\boldsymbol{\eta}_2)) \\ &+ \sum_{j=1}^n \frac{1}{\gamma} \left[1 - \left(1 + e^{\mathbf{x}'_j \boldsymbol{\varphi}_1} F_{01}(y_{1j}|\boldsymbol{\eta}_1) + e^{\mathbf{x}'_j \boldsymbol{\varphi}_2} F_{02}(y_{2j}|\boldsymbol{\eta}_2) \right)^\gamma \right] \\ &+ \sum_{j=1}^n ((1 - \gamma)(\delta_{1j} + \delta_{2j}) - \gamma \delta_{1j} \delta_{2j}) \log \left(1 + e^{\mathbf{x}'_j \boldsymbol{\varphi}_1} F_{01}(y_{1j}|\boldsymbol{\eta}_1) + e^{\mathbf{x}'_j \boldsymbol{\varphi}_2} F_{02}(y_{2j}|\boldsymbol{\eta}_2) \right) \\ &+ \sum_{j=1}^n \delta_{1j} \delta_{2j} \log \left[1 - \gamma + \left(1 + e^{\mathbf{x}'_j \boldsymbol{\varphi}_1} F_{01}(y_{1j}|\boldsymbol{\eta}_1) + e^{\mathbf{x}'_j \boldsymbol{\varphi}_2} F_{02}(y_{2j}|\boldsymbol{\eta}_2) \right)^\gamma \right], \end{aligned} \quad (5.20)$$

Desde que a distribuição de base é especificada (modelo Weibull), os estimadores de máxima verossimilhança (EMVs) de $\boldsymbol{\vartheta}$ podem ser obtidos por maximização direta da função de log-verossimilhança (5.20), usando por exemplo o procedimento de otimização tal como o metodo BFGS (Broyden-Fletcher-Goldfarb-Shanno) (ZIEGEL, 1987). Sob condições de regularidade (MALLER; ZHOU, 1996, Cap. 7) pode ser demonstrado que a distribuição assintótica do estimador de máxima verossimilhança $\hat{\boldsymbol{\vartheta}}$ é normal multivariada com média $\boldsymbol{\vartheta}$ e matriz de variância $\boldsymbol{\Sigma}(\hat{\boldsymbol{\vartheta}})$, que pode ser estimado por

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\vartheta}}) = \left\{ -\frac{\partial^2 \ell(\boldsymbol{\vartheta}|D_{obs})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\}^{-1},$$

aplicado em $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$. As segundas derivadas requeridas são calculadas numericamente.

5.4 Estudo de simulação

Para avaliar o desempenho das estimativas de máxima verossimilhança dos parâmetros do modelo proposto, realizamos um estudo de simulação considerando o modelo bivariado com distribuição de base Weibull para os tempos de falha T_k , com $k = 1, 2$, cujos parâmetros são $\boldsymbol{\eta}_1 = (1, 1)$ e $\boldsymbol{\eta}_2 = (1, 1)$. Para cada indivíduo j , o número de causas de risco, Z_{kj} do evento de interesse do tipo k para o j -ésimo indivíduo, com $j = 1, \dots, n$ e $k = 1, 2$, são gerados a partir de uma distribuição Poisson com média $\omega_j \rho_{kj}$, onde $\log(\rho_{kj}) = \varphi_{k0} + \varphi_{k1} x_j$ com $\varphi_{k0} = -0.5$ e $\varphi_{k1} = -0.5$, e a componente de fragilidade ω_j é gerada a partir de uma distribuição PVF com média 1 e variância $1 - \gamma$ descrita na Seção 3.4.5, para valores de $\gamma = -0.5(Pc), 0.3(PVF), 0.5(IG)$ e $0.7(PVF)$. Os tempos de censura (C_{1j}, C_{2j}) são gerados a partir de uma distribuição uniforme no intervalo $(0, \tau_k)$, onde τ_k foi definido para controlar a

proporção de observações censuradas. Neste estudo, a proporção de observações censuradas foi em média aproximadamente igual a 50% e 55%, respectivamente.

Para as simulações, tomamos os tamanhos de amostra $n = 250, 500$ e 1000 . Para cada configuração, realizamos 1000 simulações para calcular as médias dos EMV, média do erro padrão (MEP), desvios padrão (DP), viés, raiz dos erros quadráticos médios (REQM) das estimativas e as probabilidades de cobertura empírica (PC) correspondentes aos intervalos de confiança de 95% para os parâmetros do modelo (5.5). Os resultados da simulação listados nas Tabelas 7 e 8 revelam que os EMV estão próximos dos valores reais. Além disso, as REQMs e os DPs diminuem e os CPs empíricos estão mais próximos dos níveis nominais de cobertura, o que é esperado se o esquema de estimativa subjacente estiver funcionando corretamente para produzir estimativas consistentes e assintoticamente normais.

Tabela 7 – média dos EMVs, média do Erro padrão (MEP), Desvios padrão (DP), Viés, raízes dos erros quadráticos médios (REQM) e probabilidades de cobertura (PC) dos parâmetros do novo modelo quando $\gamma = 0.3$ e 0.7

| γ | Parâmetro | $n = 250$ | | | | | | $n = 500$ | | | | | | $n = 1000$ | | | | | |
|-------------|-------------|-----------|-------|-------|--------|--------|-------|-----------|-------|-------|--------|--------|-------|------------|-------|-------|--------|-------|-------|
| | | Média | MEP | DP | VIES | REQM | PC | Média | MEP | DP | VIES | REQM | PC | Média | MEP | DP | VIES | REQM | PC |
| 0.3 (PVF) | γ | 0.405 | 0.192 | 0.132 | 0.105 | 0.169 | 0.99 | 0.300 | 0.199 | 0.195 | -0.000 | 0.195 | 0.96 | 0.305 | 0.140 | 0.139 | 0.005 | 0.139 | 0.94 |
| | α_1 | 0.976 | 0.066 | 0.022 | -0.024 | 0.033 | 0.99 | 0.966 | 0.068 | 0.035 | -0.034 | 0.049 | 0.97 | 0.973 | 0.048 | 0.024 | -0.027 | 0.035 | 0.97 |
| | α_2 | 0.984 | 0.082 | 0.017 | -0.016 | 0.023 | 1.00 | 0.959 | 0.084 | 0.043 | -0.041 | 0.059 | 0.98 | 0.969 | 0.059 | 0.030 | -0.031 | 0.043 | 0.97 |
| | β_1 | 0.952 | 0.138 | 0.051 | -0.048 | 0.070 | 0.99 | 0.930 | 0.145 | 0.081 | -0.070 | 0.107 | 0.98 | 0.948 | 0.100 | 0.054 | -0.052 | 0.075 | 0.98 |
| | β_2 | 0.887 | 0.229 | 0.092 | -0.113 | 0.146 | 1.00 | 0.880 | 0.250 | 0.149 | -0.120 | 0.191 | 1.00 | 0.918 | 0.168 | 0.093 | -0.082 | 0.123 | 1.00 |
| | ϕ_{11} | -0.435 | 0.131 | 0.090 | 0.065 | 0.111 | 0.98 | -0.495 | 0.134 | 0.134 | 0.005 | 0.134 | 0.96 | -0.497 | 0.094 | 0.089 | 0.003 | 0.089 | 0.97 |
| | ϕ_{12} | -0.534 | 0.178 | 0.129 | -0.034 | 0.134 | 0.99 | -0.495 | 0.185 | 0.184 | 0.005 | 0.184 | 0.95 | -0.498 | 0.131 | 0.127 | 0.002 | 0.127 | 0.96 |
| | ϕ_{21} | -0.512 | 0.156 | 0.105 | -0.012 | 0.105 | 0.99 | -0.468 | 0.161 | 0.149 | 0.032 | 0.152 | 0.97 | -0.483 | 0.110 | 0.103 | 0.017 | 0.104 | 0.97 |
| | ϕ_{22} | -0.274 | 0.185 | 0.125 | 0.226 | 0.258 | 0.86 | -0.506 | 0.195 | 0.195 | -0.006 | 0.195 | 0.96 | -0.503 | 0.137 | 0.139 | -0.003 | 0.139 | 0.95 |
| | 0.7 (PVF) | γ | 0.690 | 0.276 | 0.241 | -0.010 | 0.241 | 0.97 | 0.699 | 0.193 | 0.181 | -0.001 | 0.180 | 0.98 | 0.708 | 0.137 | 0.133 | 0.008 | 0.133 |
| α_1 | | 0.956 | 0.094 | 0.046 | -0.044 | 0.063 | 0.98 | 0.967 | 0.066 | 0.032 | -0.033 | 0.046 | 0.98 | 0.975 | 0.047 | 0.022 | -0.025 | 0.033 | 0.98 |
| α_2 | | 0.949 | 0.119 | 0.055 | -0.051 | 0.075 | 0.98 | 0.961 | 0.082 | 0.042 | -0.039 | 0.057 | 0.97 | 0.970 | 0.057 | 0.028 | -0.030 | 0.041 | 0.98 |
| β_1 | | 0.912 | 0.199 | 0.107 | -0.088 | 0.138 | 0.99 | 0.935 | 0.136 | 0.073 | -0.065 | 0.098 | 0.99 | 0.953 | 0.094 | 0.048 | -0.047 | 0.067 | 0.99 |
| β_2 | | 0.839 | 0.361 | 0.200 | -0.161 | 0.257 | 0.99 | 0.888 | 0.234 | 0.133 | -0.112 | 0.173 | 0.99 | 0.919 | 0.159 | 0.089 | -0.081 | 0.120 | 0.99 |
| ϕ_{11} | | -0.491 | 0.176 | 0.159 | 0.009 | 0.159 | 0.97 | -0.489 | 0.121 | 0.113 | 0.011 | 0.113 | 0.97 | -0.499 | 0.085 | 0.081 | 0.001 | 0.081 | 0.96 |
| ϕ_{12} | | -0.501 | 0.240 | 0.231 | -0.001 | 0.231 | 0.96 | -0.500 | 0.169 | 0.161 | -0.000 | 0.161 | 0.97 | -0.499 | 0.119 | 0.110 | 0.001 | 0.110 | 0.97 |
| ϕ_{21} | | -0.449 | 0.222 | 0.193 | 0.051 | 0.199 | 0.98 | -0.471 | 0.146 | 0.128 | 0.029 | 0.131 | 0.98 | -0.481 | 0.101 | 0.092 | 0.019 | 0.094 | 0.97 |
| ϕ_{22} | | -0.516 | 0.255 | 0.259 | -0.016 | 0.259 | 0.95 | -0.505 | 0.180 | 0.177 | -0.005 | 0.177 | 0.96 | -0.497 | 0.126 | 0.126 | 0.003 | 0.126 | 0.95 |

Tabela 8 – Média dos EMVs, Média do Erro padrão (MEP), Desvios padrão (DP), vies, raízes dos erros quadráticos médios (REQM) e probabilidades de cobertura (PC) dos parâmetros do novo modelo quando $\gamma = -0.5$ e 0.5

| γ | Parâmetro | $n = 250$ | | | | | | $n = 500$ | | | | | | $n = 1000$ | | | | | |
|-----------|-------------|-----------|-------|-------|--------|-------|------|-----------|-------|-------|--------|-------|------|------------|-------|-------|--------|-------|------|
| | | Média | MEP | DP | VIES | REQM | PC | Média | MEP | DP | VIES | REQM | PC | Média | MEP | DP | VIES | REQM | PC |
| | γ | -0.521 | 0.305 | 0.300 | -0.021 | 0.301 | 0.95 | -0.508 | 0.213 | 0.215 | -0.008 | 0.215 | 0.94 | -0.507 | 0.150 | 0.150 | -0.007 | 0.150 | 0.96 |
| | α_1 | 0.955 | 0.103 | 0.049 | -0.045 | 0.067 | 0.98 | 0.965 | 0.071 | 0.036 | -0.035 | 0.050 | 0.97 | 0.973 | 0.050 | 0.024 | -0.027 | 0.036 | 0.98 |
| | α_2 | 0.944 | 0.128 | 0.061 | -0.056 | 0.083 | 0.98 | 0.959 | 0.089 | 0.043 | -0.041 | 0.059 | 0.98 | 0.967 | 0.062 | 0.032 | -0.033 | 0.046 | 0.97 |
| | β_1 | 0.889 | 0.242 | 0.135 | -0.111 | 0.175 | 0.99 | 0.920 | 0.163 | 0.089 | -0.080 | 0.120 | 0.98 | 0.942 | 0.113 | 0.061 | -0.058 | 0.084 | 0.98 |
| -0.5 (cP) | β_2 | 0.809 | 0.436 | 0.233 | -0.191 | 0.301 | 1.00 | 0.869 | 0.280 | 0.150 | -0.131 | 0.199 | 1.00 | 0.902 | 0.188 | 0.109 | -0.098 | 0.147 | 0.99 |
| | ϕ_{11} | -0.477 | 0.232 | 0.223 | 0.023 | 0.224 | 0.96 | -0.482 | 0.159 | 0.152 | 0.018 | 0.153 | 0.96 | -0.494 | 0.112 | 0.108 | 0.006 | 0.108 | 0.96 |
| | ϕ_{12} | -0.523 | 0.311 | 0.317 | -0.023 | 0.317 | 0.95 | -0.508 | 0.218 | 0.220 | -0.008 | 0.220 | 0.95 | -0.498 | 0.154 | 0.154 | 0.002 | 0.154 | 0.96 |
| | ϕ_{21} | -0.437 | 0.291 | 0.250 | 0.063 | 0.257 | 0.99 | -0.465 | 0.190 | 0.172 | 0.035 | 0.175 | 0.98 | -0.473 | 0.130 | 0.117 | 0.027 | 0.120 | 0.97 |
| | ϕ_{22} | -0.508 | 0.322 | 0.326 | -0.008 | 0.326 | 0.96 | -0.507 | 0.227 | 0.233 | -0.007 | 0.233 | 0.95 | -0.507 | 0.160 | 0.165 | -0.007 | 0.165 | 0.94 |
| | γ | 0.492 | 0.280 | 0.270 | -0.008 | 0.270 | 0.98 | 0.506 | 0.196 | 0.189 | 0.006 | 0.189 | 0.96 | 0.508 | 0.138 | 0.134 | 0.008 | 0.135 | 0.96 |
| | α_1 | 0.957 | 0.096 | 0.048 | -0.043 | 0.064 | 0.97 | 0.964 | 0.067 | 0.035 | -0.036 | 0.050 | 0.97 | 0.973 | 0.047 | 0.023 | -0.027 | 0.035 | 0.97 |
| | α_2 | 0.947 | 0.120 | 0.057 | -0.053 | 0.078 | 0.98 | 0.958 | 0.083 | 0.042 | -0.042 | 0.060 | 0.97 | 0.970 | 0.058 | 0.028 | -0.030 | 0.041 | 0.98 |
| | β_1 | 0.910 | 0.205 | 0.105 | -0.090 | 0.138 | 0.99 | 0.936 | 0.141 | 0.073 | -0.064 | 0.097 | 0.99 | 0.950 | 0.097 | 0.056 | -0.050 | 0.075 | 0.97 |
| 0.5 (IG) | β_2 | 0.850 | 0.369 | 0.191 | -0.150 | 0.243 | 1.00 | 0.883 | 0.242 | 0.139 | -0.117 | 0.182 | 0.99 | 0.919 | 0.162 | 0.086 | -0.081 | 0.118 | 0.99 |
| | ϕ_{11} | -0.480 | 0.185 | 0.176 | 0.020 | 0.177 | 0.96 | -0.496 | 0.127 | 0.118 | 0.004 | 0.118 | 0.97 | -0.497 | 0.090 | 0.092 | 0.003 | 0.092 | 0.94 |
| | ϕ_{12} | -0.513 | 0.251 | 0.253 | -0.013 | 0.254 | 0.95 | -0.514 | 0.178 | 0.180 | -0.014 | 0.180 | 0.95 | -0.502 | 0.125 | 0.128 | -0.002 | 0.128 | 0.94 |
| | ϕ_{21} | -0.452 | 0.231 | 0.208 | 0.048 | 0.213 | 0.99 | -0.466 | 0.153 | 0.133 | 0.034 | 0.137 | 0.98 | -0.483 | 0.105 | 0.097 | 0.017 | 0.098 | 0.97 |
| | ϕ_{22} | -0.522 | 0.266 | 0.263 | -0.022 | 0.263 | 0.96 | -0.508 | 0.187 | 0.182 | -0.008 | 0.182 | 0.96 | -0.502 | 0.132 | 0.131 | -0.002 | 0.131 | 0.95 |

5.5 Aplicação: dados de churn de clientes brasileiros

Para ilustrar nossa proposta de modelagem discutida até agora, nesta seção consideramos uma amostra de uma carteira de clientes brasileiros, que compreende o tempo até o churn (em anos) em duas carteiras de produtos diferentes de cartão de crédito. O churn de clientes, normalmente conhecida como rotatividade de clientes (customer chur), mas também, desistência de clientes. A perda de clientes é uma grande preocupação das empresas ou prestadores de serviços atualmente. O foco central é o churn voluntário, que ocorre por decisão pessoal do cliente, que decide migrar para outra empresa ou prestador de serviço. Para os dados de churn de clientes brasileiros observamos dois tempos em anos até o churn, T_1 e T_2 , em dois produtos de cartão de crédito, de agora em diante denominado de Produto 1 e Produto 2. O período de acompanhamento dos clientes para ambos produtos foi de 7 anos. O estudo foi realizado com 900 portadores de cartão de crédito de um banco brasileiro para mensurar o churn de clientes. A censura é de aproximadamente 52% e 73% para cada um dos produtos respectivamente, e podem ser consideradas as frações de clientes fiéis, que permanecem no produto mesmo após vários anos. Na Figura 14 são plotadas o gráfico de dispersão (painel esquerdo), e as estimativas de Kaplan-Meier das funções de sobrevivência (painel direito) para esse conjunto de dados sobre os tempos até o abandono de clientes dos produtos 1 e 2 onde o primeiro gráfico mostra uma estrutura bivariada dos dados, conforme mostra a a Figura 1 (painel esquerdo), e as funções de sobrevivência estimadas tendem a níveis positivos abaixo para um limite horizontal superior. Esse fato, segundo [Yakovlev, Tsodikov e Asselain \(1996\)](#), pode ser pensado como um indicativo da presença de uma proporção de clientes que não abandonam os produtos após um longo tempo. Esse comportamento indica que modelos que não consideram a possibilidade de cura podem não ser adequados para modelar esses dados.

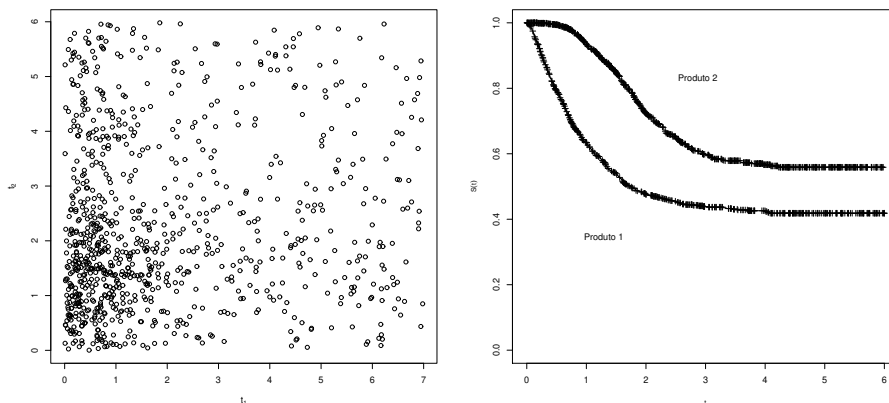


Figura 14 – Gráfico de dispersão (painel esquerdo) e Estimativas de Kaplan-Meier das funções de sobrevivência do produto 1 e 2 (painel direito)

O objetivo aqui é verificar se existe alguma associação entre os tempos até o churn de clientes do produto 1 e 2, investigar se há diferença na proporção de clientes fiéis (taxa de cura) entre os produtos 1 e 2 e avaliar os efeitos das covariáveis na fração curada. Para nossos propósitos, consideramos como variáveis de resposta os tempos até o abandono dos produtos 1 e 2, as quais serão modeladas consideramos a distribuição Weibull. As seguintes variáveis são coletadas de cada cliente:

- t_{kj} : Tempo observado (em anos) do k -ésimo produto, com $k = 1, 2$.

- x_{1j} : Sexo (0= feminino (51%), 1=masculino(49%).
- x_{2j} : Renda (1 = renda < 3.000, 2 = 3.000 ≤ renda < 7.000, 3 = renda ≥ 7000), $j = 1, \dots, 900$.

Para este conjunto de dados, ajustamos o modelo descrito na Seção 5.2 com todas as covariáveis no parâmetro ρ_{kj} , para $k = 1, 2$ e $j = 1, \dots, 900$, isto é,

$$\log(\rho_{kj}) = \Phi_{k,Intercepto} + \Phi_{k,sexo_1}x_{j,sexo_1} + \Phi_{k,renda_2}x_{j,renda_2} + \Phi_{k,renda_3}x_{j,renda_3}. \quad (5.21)$$

Onde os efeitos das covariáveis (sexo e renda) são definidas por meio de variáveis dummy, da seguinte forma:

$$x_{j,sexo_1} = \begin{cases} 1, & \text{se for Masculino;} \\ 0, & \text{caso contrário} \end{cases},$$

$$x_{j,renda_2} = \begin{cases} 1, & \text{se } 3.000 \leq \text{renda} < 7.000; \\ 0, & \text{caso contrário.} \end{cases} \quad \text{e} \quad x_{j,renda_3} = \begin{cases} 1, & \text{se } \text{renda} \geq 7000; \\ 0, & \text{caso contrário.} \end{cases}$$

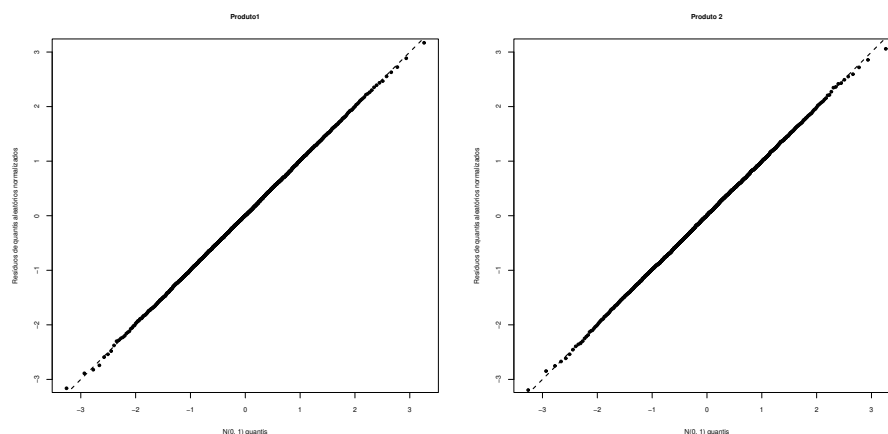


Figura 15 – QQ-plot dos resíduos quantílicos randomizados normalizados com linha de identidade para a função de sobrevivência marginal

Para avaliar a adequação do ajuste, obtemos o QQ-plot dos resíduos de quantis randomizados normalizados na Figura 15 que sugere que o modelo produz um ajuste aceitável. Cada ponto da Figura 15 corresponde à mediana de cinco conjuntos de resíduos ordenados. As estimativas de máxima verossimilhança dos parâmetros do modelo de longa duração bivariada, com os respectivos erro padrões (EP) e os intervalos de confiança (IC (95%)), estão listados na Tabela 9, considerando a fragilidade PVF e os casos particulares: Inversa Gaussiana (IG, $\gamma = 0,5$) e Gamma ($\gamma \rightarrow 0$). Estas estimativas indicam que, para um nível de significância de 5%, observamos que todas as covariáveis têm efeito significativo na fração de cura, tanto para o produto 1 quanto para o produto 2, exceto pela variável renda 2 ($3.000 \leq \text{renda} < 7.000$), a qual não tem efeito significativo na fração de cura para o o produto 1. Por outro lado, lembrando que o modelo PVF proposto tem a vantagem de poder sugerir qual de seus modelos específicos melhor se ajustaria aos dados, com base nas estimativas de seu parâmetro γ . Desta forma, vendo a estimação do parâmetro γ , ele assume um valor muito próximo de zero, o que sugere que o modelo gama poderia ser o escolhido para a modelagem. No entanto, olhando para a Tabela 10, onde é apresentada os critérios

de seleção de modelos, podemos perceber que embora os resultados apresentados nesta tabela sejam favoráveis ao modelo PVF, não há diferença suficiente com os valores obtidos a partir modelos Gama e IG, para poder optar por um modelo definitivo (ANDERSON; BURNHAM, 2004).

Por outro lado, na Tabela 9, note que a estimativa do parâmetro de dependência θ , do modelo PVF é dada por $\hat{\theta} = 1 - \hat{\gamma} = 0.949$, muito próximo ao valor 0.956, estimado no modelo Gama. A partir de aqui consideraremos as estimativas do modelo PVF. Assim, calculamos a estatística de razão de verossimilhança (Δ_n) para testar se a dependência entre os tempos para o abandono dos produtos é significativa, ou seja, $H_0 : \theta = 0$ (modelo independente) versus $H_1 : \theta > 0$ (modelo dependente). É importante notar que sob a hipótese nula H_0 , a estatística, Δ_n é assintoticamente distribuída como uma mistura simétrica de uma distribuição qui-quadrado com um grau de liberdade e uma massa pontual em zero. Assim, Δ_n é igual a 76.494, com p -valor = 0,000, o que fornece forte evidência a favor de H_1 , indicando que há dependência significativa entre os tempos.

Tabela 9 – EMV dos parâmetros para o modelo de taxa de cura bivariada para conjunto de dados do banco, considerando os modelos de fragilidade Gamma, IG e PVF.

| Par | Gamma | | | IG | | | PVF | | |
|-------------------------------|--------|-------|------------------|--------|-------|------------------|--------|-------|------------------|
| | EMV | EP | IC(95%) | EMV | EP | IC(95%) | EMV | EP | IC(95%) |
| β_1 | 1.586 | 0.065 | (1.460, 1.713) | 1.599 | 0.065 | (1.472, 1.727) | 1.581 | 0.062 | (1.459, 1.703) |
| α_1 | -0.871 | 0.104 | (-1.074, -0.668) | -0.837 | 0.096 | (-1.024, -0.649) | -0.858 | 0.101 | (-1.055, -0.660) |
| Prod. 1 $\phi_{1,Intercepto}$ | 0.924 | 0.159 | (0.612, 1.236) | 0.991 | 0.172 | (0.653, 1.329) | 0.909 | 0.150 | (0.616, 1.202) |
| $\phi_{1,sexo_1}$ | -1.951 | 0.153 | (-2.250, -1.652) | -1.930 | 0.151 | (-2.225, -1.635) | -1.938 | 0.149 | (-2.230, -1.647) |
| $\phi_{1,renda_2}$ | 0.137 | 0.162 | (-0.181, 0.455) | 0.156 | 0.163 | (-0.164, 0.477) | 0.137 | 0.161 | (-0.179, 0.453) |
| $\phi_{1,renda_3}$ | 1.183 | 0.164 | (0.863, 1.504) | 1.109 | 0.156 | (0.804, 1.414) | 1.169 | 0.162 | (0.850, 1.487) |
| β_2 | 3.079 | 0.146 | (2.793, 3.366) | 3.137 | 0.150 | (2.843, 3.432) | 3.075 | 0.144 | (2.793, 3.357) |
| α_2 | -2.860 | 0.160 | (-3.173, -2.547) | -2.919 | 0.165 | (-3.242, -2.597) | -2.851 | 0.155 | (-3.155, -2.547) |
| Prod. 1 $\phi_{2,Intercepto}$ | 1.247 | 0.170 | (0.913, 1.580) | 1.349 | 0.186 | (0.985, 1.713) | 1.236 | 0.161 | (0.920, 1.553) |
| $\phi_{2,sexo_2}$ | -1.018 | 0.161 | (-1.333, -0.702) | -1.048 | 0.165 | (-1.371, -0.724) | -1.017 | 0.160 | (-1.329, -0.704) |
| $\phi_{2,renda_2}$ | -0.979 | 0.178 | (-1.328, -0.631) | -0.950 | 0.180 | (-1.302, -0.597) | -0.973 | 0.177 | (-1.320, -0.626) |
| $\phi_{2,renda_3}$ | -2.373 | 0.232 | (-2.827, -1.918) | -2.432 | 0.239 | (-2.901, -1.964) | -2.365 | 0.229 | (-2.813, -1.917) |
| θ | 0.956 | 0.146 | (0.670, 1.242) | 1.725 | 0.411 | (0.920, 2.530) | - | - | - |
| γ | - | - | - | - | - | - | 0.051 | 0.085 | (-0.116, 0.217) |

Tabela 10 – Critérios de comparação de modelos, Log-Verossimilhança, AIC e BIC, para os dados do banco

| | Log-Likelihood | AIC | BIC |
|-------|----------------|----------|----------|
| Gamma | -1379.631 | 2785.263 | 2847.694 |
| IG | -1380.034 | 2786.068 | 2848.499 |
| PVF | -1379.494 | 2784.988 | 2847.419 |

A partir dos EMV, também estimamos a probabilidade de os clientes não cancelarem um determinado produto (proporção de cura) após acompanhamento $t > 0$ ($k = 1, 2$) dado por

$$\pi_k(t) = \left[\frac{\exp \left\{ \left(1 + \rho_k (1 - \exp \{-e^{\alpha_k t \beta_k}\}) \right)^\gamma \right\}}{\exp \left\{ (1 + \rho_k)^\gamma \right\}} \right]^{\frac{1}{\gamma}}, \quad k = 1, 2. \quad (5.22)$$

onde ρ_k é dado em (5.21). A partir deste resultado, estimamos a proporção de clientes fiéis, para 6 clientes hipotéticos A, B, C, D, E e F após um período de observação com valores para as covariáveis dadas na Tabela 11.

A Tabela 11 apresenta as EMV das proporção de clientes que não abandonarão ambos produtos, p_{00} , o produto 1, p_{01} , e o produto 2, p_{02} (taxa de cura). Por exemplo, para clientes com fatores de risco iguais ao cliente A, a probabilidade de não abandonar ambos produtos é $p_{00} = 0,131$, e para os produtos 1 e 2 são $p_{01} = 0,276$ e $p_{02} = 0,212$ respectivamente. Por outro lado, as probabilidades desses clientes continuarem com os produtos 1 e 2 após 3 anos são de $\pi_1(3) = 0,932$ e $\pi_2(3) = 0,848$, respectivamente. A partir da Tabela 11 observamos que a proporção de clientes que continuarem com ambos produtos, p_{00} , é menor para clientes femininas do que para clientes masculinos, para qualquer nível de renda. Esta relação de proporções de clientes fiéis (entre mulheres e homens) se mantém para os produtos 1 e 2 de forma individual. Observe que a fração clientes que não cancelaram o produto 2 aumenta a medida que a renda aumenta, tanto para homens e mulheres. No entanto, para clientes do produto 1 essa proporção diminui a medida que a renda aumenta.

Tabela 11 – Estimativas de máxima verossimilhança da proporção de não abandono dos produtos 1 e 2 do sistema financeiro .

| Cliente | Renda | Sexo | Fração de não abandono dos produtos | | | | |
|---------|-----------------------|-----------|-------------------------------------|-----------|------------|-----------|------------|
| | | | Produtos 1 e 2 p_{00} | Produto 1 | | Produto 2 | |
| | | | | p_{01} | $\pi_1(3)$ | p_{02} | $\pi_2(3)$ |
| A | < 3000 | Feminino | 0.131 | 0.276 | 0.932 | 0.212 | 0.848 |
| B | | Masculino | 0.375 | 0.735 | 0.976 | 0.438 | 0.894 |
| C | ≥ 3000 e < 7000) | Feminino | 0.181 | 0.248 | 0.929 | 0.427 | 0.892 |
| D | | Masculino | 0.526 | 0.707 | 0.973 | 0.677 | 0.940 |
| E | > 7000 | Feminino | 0.094 | 0.098 | 0.911 | 0.754 | 0.955 |
| F | | Masculino | 0.434 | 0.458 | 0.950 | 0.895 | 0.981 |

A Figura 16 mostra o comportamento das probabilidades de que os clientes hipotéticos, mostrados na Tabela 11, não abandone os produtos após um período de observação $t \in [0, 6]$. Aqui, podemos observar que essa proporção aumenta com o tempo, chegando perto de 1 a partir do quarto ano aproximadamente. Finalmente, na Figura 17, as estimativas da função de risco cumulativo são apresentadas, onde podemos observar a aderência do modelo proposto com a estimativa de Kaplan-Meier.

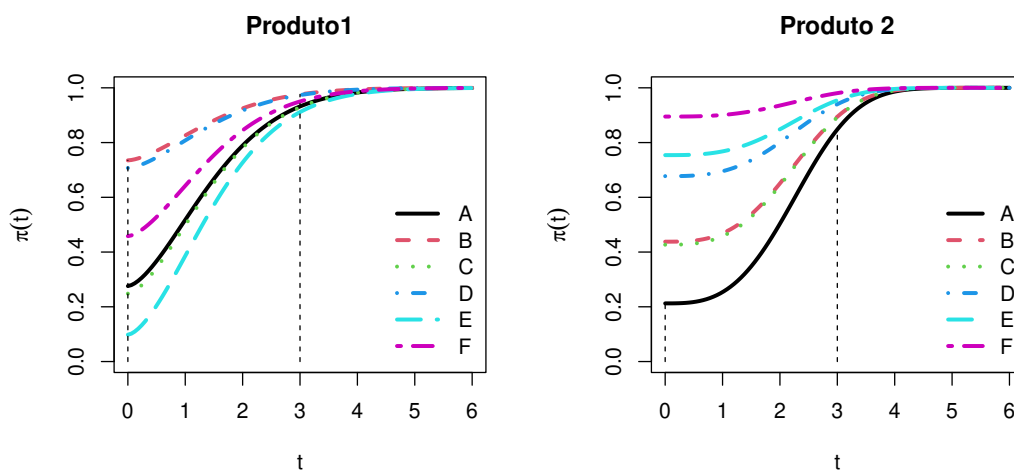


Figura 16 – A probabilidade de não cancelar os produtos 1 (painel esquerdo) e 2 (painel direito) de seis clientes de uma banco brasileiro, dado que não cancelaram após o terceiro ano

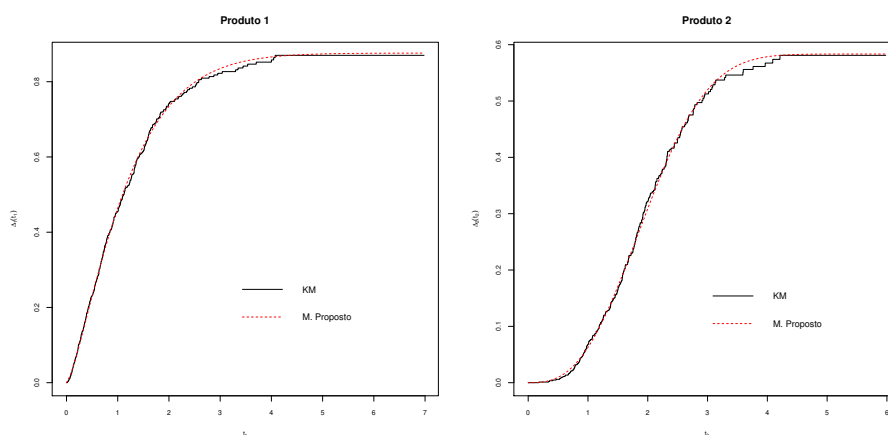


Figura 17 – Estimativas de Kaplan-Meier das funções de risco acumulado e do modelo proposto para o produto 1 (painel esquerdo) e 2 (painel direito)

CONSIDERAÇÕES FINAIS

6.1 Conclusões

Nesta tese, consideramos a modelagem de dados semicompetitivos que surgem em ensaios clínicos de câncer por meio da estrutura de doença-morte com fragilidade compartilhada. Nossa contribuição central é evitar a abordagem do tempo de falha latente (que vem com limitações bem exploradas) e explorar a especificação doença-morte por meio da fragilidade PVF, apresentada por [Wienke \(2010\)](#), que é uma classe flexível e geral contendo as densidades gama, inversa gaussiana e positiva estável. A conveniente transformada de Laplace do PVF permite representações matematicamente tratáveis das funções de riscos, probabilidades de transição e funções de sobrevivência. O pacote R `SmoothHazard` ([TOURAINÉ; GERDS; JOLY, 2017](#)) também se ajusta a um modelo doença-morte com especificações Weibull (paramétricas) ou M-splines penalizadas (semi-paramétricas) de risco de base para dados de sobrevivência arbitrariamente censurados. Para comparar e contrastar, nossa abordagem incorpora uma estrutura de dependência entre o tempo de recorrência e o tempo de morte usando uma fragilidade compartilhada entre as taxas de transição condicional, que atua como um efeito multiplicador nas taxas de risco de base. Essa dependência é caracterizada pelo parâmetro de fragilidade θ . O estudo de simulação indica um desempenho de amostra finita adequado com tamanhos de amostra crescentes. Finalmente, o modelo foi aplicado a um conjunto de dados reais de câncer de cólon, onde, neste caso específico, o modelo de fragilidade do PVF se destaca como o melhor, de acordo com os critérios de informação de Akaike (AIC) e o critério de informação bayesiano (BIC). Além disso, os resultados obtidos foram bastante consistentes.

Por outro lado, nesta tese, também propomos um novo modelo de sobrevivência bivariada de longa duração cujo número de causas de risco, para k diferentes tipos de eventos de interesse, é modelado por meio de um modelo de fragilidade discreta Poisson com média que a sua vez depende de uma variável de fragilidade PVF, sendo esta última a responsável de incorporar um estrutura de dependência entre os k tipos de eventos. Examinamos algumas de suas propriedades onde podemos observar a flexibilidade do modelo graças ao fato de que a fragilidade do PVF proposta por [Hougaard e Hougaard \(2000\)](#) que usamos aqui inclui as distribuições Gama, Inversa Gaussiana e positiva estável. Dessa forma, na inferência,

dependendo da estimação do parâmetro γ , pode-se concluir qual modelo melhor se ajusta aos dados, além do fato de que se essa estimação fosse negativa, a distribuição Poisson composta, que inclui um fator de risco zero, seria a escolhida. Um estudo de simulação foi apresentado considerando diferentes valores do parâmetro γ onde em cada caso podemos observar um desempenho de amostra finita adequado com tamanhos de amostra crescentes. Discutimos sua aplicabilidade em dados de churn de clientes de um sistema financeiro brasileiro, em que nosso modelo oferece o melhor ajuste. A estimação dos parâmetros do modelo foi realizada usando a abordagem de máxima verossimilhança, que pode ser implementada de forma direta usando software R disponível (R Core Team, 2018). Este fato torna a abordagem bastante poderosa e acessível aos profissionais da área.

6.2 Sugestões para pesquisa futura

- No modelo de riscos semicompetitivos, a taxa de transição condicional para o evento terminal dado que um evento não terminal foi especificado como um modelo de Markov, no entanto, uma especificação semi-Markov (LEE *et al.*, 2015) também é possível.
- Uma expansão do modelo doença-morte com fragilidade compartilhada, pode ser estudada, permitindo que o evento terminal consista em dois eventos competitivos e o modelo agora consiste em cinco taxas de transição em vez de três, conforme representado na Figura 18. As taxas marginais de transição são definidas como:

$$\begin{aligned}\lambda_1(t_1) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_1 \in [t_1, t_1 + \Delta) | T_1 \geq t_1, T_2 \geq t_1]}{\Delta}, \quad t_1 > 0 \\ \lambda_{2a}(t_2) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta), K = a | T_1 \geq t_2, T_2 \geq t_2]}{\Delta}, \quad t_2 > 0 \\ \lambda_{2b}(t_2) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta), K = b | T_1 \geq t_2, T_2 \geq t_2]}{\Delta}, \quad t_2 > 0 \\ \lambda_{12a}(t_2 | t_1) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta), K = a | T_1 = t_1, T_2 \geq t_2]}{\Delta}, \quad 0 < t_1 \leq t_2. \\ \lambda_{12b}(t_2 | t_1) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2 + \Delta), K = b | T_1 = t_1, T_2 \geq t_2]}{\Delta}, \quad 0 < t_1 \leq t_2.\end{aligned}$$

- Nossa proposta atual nesta tese de um risco de base Weibull (paramétrico) é certamente pela flexibilidade do modelo. Extensões futuras podem incluir várias escolhas populares e flexíveis do risco de base não paramétrica e semi-paramétrica (SPRINGER,), como constante por partes, e estudar seu impacto tanto na estimativa de parâmetros quanto no ganho computacional.
- O procedimento inferencial clássico foi proposto nesta tese. Também pode ser explorada sob um paradigma bayesiano para ambos modelos, de riscos semicompetitivos do Capítulo 4 e de longa duração no Capítulo 5.

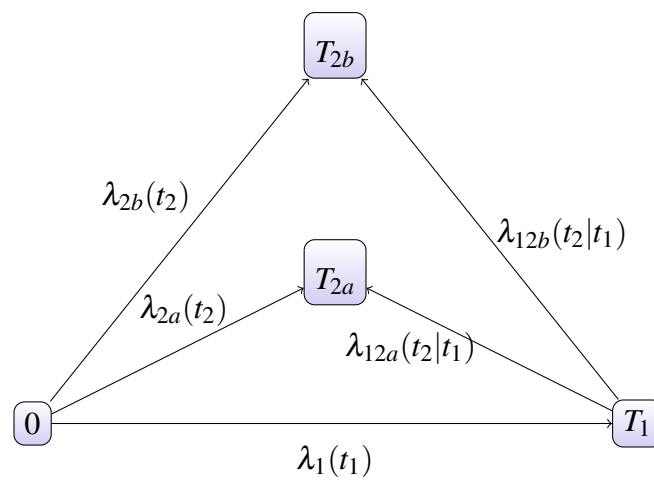


Figura 18 – Uma situação de riscos semicompetitivos em que a falha do evento terminal consiste em dois eventos separados, $2a$ e $2b$. As taxas de transição para cada transição estão incluídas.

REFERÊNCIAS

- AALEN, O. O. Heterogeneity in survival analysis. **Statistics in medicine**, Wiley Online Library, v. 7, n. 11, p. 1121–1137, 1988. Citado nas páginas 24 e 48.
- AALEN, O. O. *et al.* Modelling heterogeneity in survival analysis by the compound poisson distribution. **The Annals of Applied Probability**, Institute of Mathematical Statistics, v. 2, n. 4, p. 951–972, 1992. Citado na página 48.
- AKAIKE, H. A new look at the statistical model identification. In: **Selected Papers of Hirotugu Akaike**. [S.l.]: Springer, 1974. p. 215–222. Citado na página 72.
- ANDERSEN, P. K.; KEIDING, N. Multi-state models for event history analysis. **Statistical methods in medical research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 11, n. 2, p. 91–115, 2002. Citado na página 52.
- ANDERSON, D.; BURNHAM, K. Model selection and multi-model inference. **Second**. NY: Springer-Verlag, v. 63, n. 2020, p. 10, 2004. Citado nas páginas 72 e 90.
- BANDEEN-ROCHE, K. *et al.* A diagnostic for association in bivariate survival models. **Lifetime Data Analysis**, Springer, v. 11, n. 2, p. 245–264, 2005. Citado na página 52.
- BARRIGA, G. D.; CANCHO, V. G.; GARIBAY, D. V.; CORDEIRO, G. M.; ORTEGA, E. M. A new survival model with surviving fraction: An application to colorectal cancer data. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 28, n. 9, p. 2665–2680, 2019. Citado na página 77.
- BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, Taylor & Francis, v. 47, n. 259, p. 501–515, 1952. Citado nas páginas 25 e 77.
- CANCHO, V. G.; LOUZADA, F.; DEY, D. K.; BARRIGA, G. D. A new lifetime model for multivariate survival data with a surviving fraction. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 86, n. 2, p. 279–292, 2016. Citado na página 78.
- CANCHO, V. G.; RODRIGUES, J.; CASTRO, M. de. A flexible model for survival data with a cure rate: a bayesian approach. **Journal of Applied Statistics**, Taylor & Francis, v. 38, n. 1, p. 57–70, 2011. Citado nas páginas 78 e 79.
- CHATTERJEE, N.; SHIH, J. A bivariate cure-mixture approach for modeling familial association in diseases. **Biometrics**, Wiley Online Library, v. 57, n. 3, p. 779–786, 2001. Citado na página 78.
- CHEN, M.-H.; IBRAHIM, J. G.; SINHA, D. Bayesian inference for multivariate survival data with a cure fraction. **Journal of Multivariate Analysis**, Elsevier, v. 80, n. 1, p. 101–126, 2002. Citado nas páginas 78 e 79.
- CHEN, Y.-H. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. **Lifetime data analysis**, Springer, v. 18, n. 1, p. 36–57, 2012. Citado na página 52.
- CLAYTON, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. **Biometrika**, Oxford University Press, v. 65, n. 1, p. 141–151, 1978. Citado nas páginas 32, 41, 43, 52, 55 e 80.

- CORDEIRO, G. M.; CANCHO, V. G.; ORTEGA, E. M.; BARRIGA, G. D. A model with long-term survivors: negative binomial birnbaum-saunders. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 45, n. 5, p. 1370–1387, 2016. Citado na página 77.
- COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. Citado nas páginas 23 e 36.
- CROWDER, M. J. **Multivariate survival analysis and competing risks**. [S.l.]: Chapman and Hall/CRC, 2012. Citado na página 29.
- DUCHATEAU, L.; JANSSEN, P. **The frailty model**. [S.l.]: Springer, 2008. Citado nas páginas 23, 24, 39, 40, 41, 42, 43, 44 e 45.
- DURANTE, F.; SEMPI, C. Copula theory: an introduction. In: **Copula theory and its applications**. [S.l.]: Springer, 2010. p. 3–31. Citado na página 31.
- FELLER, W. Law of large numbers for identically distributed variables. **An introduction to probability theory and its applications**, Wiley, v. 2, p. 231–234, 1971. Citado na página 45.
- FINE, J. P.; JIANG, H.; CHAPPELL, R. On semi-competing risks data. **Biometrika**, Oxford University Press, v. 88, n. 4, p. 907–919, 2001. Citado nas páginas 9, 11, 25, 30, 51 e 52.
- GALLARDO, D. I.; BOLFARINE, H.; PEDROSO-DE-LIMA, A. C. Promotion time cure rate model with bivariate random effects. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 45, n. 2, p. 603–624, 2016. Citado na página 78.
- GELFAND, A. E.; DEY, D. K. Bayesian model choice: asymptotics and exact calculations. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 56, n. 3, p. 501–514, 1994. Citado na página 72.
- GILBERT, P.; VARADHAN, R. **numDeriv: Accurate Numerical Derivatives**. [S.l.], 2019. R package version 2016.8-1.1. Disponível em: <<https://CRAN.R-project.org/package=numDeriv>>. Citado na página 66.
- HA, I. D.; XIANG, L.; PENG, M.; JEONG, J.-H.; LEE, Y. Frailty modelling approaches for semi-competing risks data. **Lifetime data analysis**, Springer, v. 26, n. 1, p. 109–133, 2020. Citado na página 53.
- HAN, B.; YU, M.; DIGNAM, J. J.; RATHOUZ, P. J. Bayesian approach for flexible modeling of semicompeting risks data. **Statistics in medicine**, Wiley Online Library, v. 33, n. 29, p. 5111–5125, 2014. Citado na página 53.
- HANEUSE, S.; LEE, K. H. Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal. **Circulation: Cardiovascular Quality and Outcomes**, Am Heart Assoc, v. 9, n. 3, p. 322–331, 2016. Citado nas páginas 51 e 52.
- HOUGAARD, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. **Biometrika**, Oxford University Press, v. 71, n. 1, p. 75–83, 1984. Citado nas páginas 24, 36 e 43.
- _____. A class of multivariate failure time distributions. **Biometrika**, Oxford University Press, v. 73, n. 3, p. 671–678, 1986. Citado nas páginas 24, 33, 44 e 45.
- _____. Survival models for heterogeneous populations derived from stable distributions. **Biometrika**, Oxford University Press, v. 73, n. 2, p. 387–396, 1986. Citado nas páginas 24, 33, 36 e 46.
- _____. **Analysis of multivariate survival data**. [S.l.]: Springer Science & Business Media, 2012. Citado nas páginas 25, 41 e 42.

- HOUGAARD, P.; HOUGAARD, P. **Analysis of multivariate survival data**. [S.l.]: Springer, 2000. v. 564. Citado nas páginas 21, 23, 24, 26, 41, 42, 47, 49, 79 e 95.
- HSIEH, J.-J.; HUANG, Y.-T. Regression analysis based on conditional likelihood approach under semi-competing risks data. **Lifetime data analysis**, Springer, v. 18, n. 3, p. 302–320, 2012. Citado na página 52.
- JIANG, F.; HANEUSE, S. A semi-parametric transformation frailty model for semi-competing risks survival data. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 44, n. 1, p. 112–129, 2017. Citado na página 53.
- KIM, J.; KIM, J.; KIM, S. W. Additive-multiplicative hazards regression models for interval-censored semi-competing risks data with missing intermediate events. **BMC medical research methodology**, BioMed Central, v. 19, n. 1, p. 49, 2019. Citado na página 53.
- KIM, Y.-J. Cure rate model with bivariate interval censored data. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 46, n. 9, p. 7116–7124, 2017. Citado na página 78.
- KOUTRAS, M.; MILIENOS, F. A flexible family of transformation cure rate models. **Statistics in medicine**, Wiley Online Library, v. 36, n. 16, p. 2559–2575, 2017. Citado na página 77.
- LAKHAL, L.; RIVEST, L.-P.; ABDOUS, B. Estimating survival and association in a semicompeting risks model. **Biometrics**, Wiley Online Library, v. 64, n. 1, p. 180–188, 2008. Citado na página 52.
- LAU, B.; COLE, S. R.; GANGE, S. J. Competing risk regression models for epidemiologic data. **American journal of epidemiology**, Oxford University Press, v. 170, n. 2, p. 244–256, 2009. Citado na página 51.
- LEE, C.; GILSANZ, P.; HANEUSE, S. Fitting a shared frailty illness-death model to left-truncated semi-competing risks data to examine the impact of education level on incident dementia. **BMC Medical Research Methodology**, BioMed Central, v. 21, n. 1, p. 1–13, 2021. Citado na página 53.
- LEE, K. H.; DOMINICI, F.; SCHRAG, D.; HANEUSE, S. Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. **Journal of the American Statistical Association**, Taylor & Francis, v. 111, n. 515, p. 1075–1095, 2016. Citado na página 53.
- LEE, K. H.; HANEUSE, S.; SCHRAG, D.; DOMINICI, F. Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 64, n. 2, p. 253–273, 2015. Citado nas páginas 53, 72 e 96.
- LEE, Y.; NELDER, J. A. *et al.* Conditional and marginal models: another view. **Statistical Science**, Institute of Mathematical Statistics, v. 19, n. 2, p. 219–238, 2004. Citado na página 60.
- MALLER, R. A.; ZHOU, X. **Survival analysis with long-term survivors**. [S.l.]: John Wiley & Sons, 1996. Citado nas páginas 65 e 84.
- MARTINS, R.; SILVA, G. L.; ANDREOZZI, V. Joint analysis of longitudinal and survival aids data with a spatial fraction of long-term survivors: A bayesian approach. **Biometrical Journal**, Wiley Online Library, v. 59, n. 6, p. 1166–1183, 2017. Citado na página 78.
- MEIRA-MACHADO, L.; UÑA-ÁLVAREZ, J. de; CADARSO-SUAREZ, C.; ANDERSEN, P. K. Multi-state models for the analysis of time-to-event data. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 18, n. 2, p. 195–222, 2009. Citado nas páginas 53 e 54.

- MOERTEL, C. G.; FLEMING, T. R.; MACDONALD, J. S.; HALLER, D. G.; LAURIE, J. A.; GOODMAN, P. J.; UNGERLEIDER, J. S.; EMERSON, W. A.; TORMEY, D. C.; GLICK, J. H. *et al.* Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. **New England Journal of Medicine**, Mass Medical Soc, v. 322, n. 6, p. 352–358, 1990. Citado nas páginas 51 e 67.
- MOERTEL, C. G.; FLEMING, T. R.; MACDONALD, J. S.; HALLER, D. G.; LAURIE, J. A.; TANGEN, C. M.; UNGERLEIDER, J. S.; EMERSON, W. A.; TORMEY, D. C.; GLICK, J. H. *et al.* Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage iii colon carcinoma: a final report. **Annals of internal medicine**, American College of Physicians, v. 122, n. 5, p. 321–326, 1995. Citado na página 51.
- NELSEN, R. B. **An introduction to copulas**. [S.l.]: Springer Science & Business Media, 2007. Citado nas páginas 31 e 32.
- OAKES, D. A model for association in bivariate survival data. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 414–422, 1982. Citado na página 41.
- _____. A note on the kaplan—meier estimator. **The American Statistician**, Taylor & Francis, v. 47, n. 1, p. 39–40, 1993. Citado na página 51.
- ORTEGA, E. M.; CORDEIRO, G. M.; CAMPELO, A. K.; KATTAN, M. W.; CANCHO, V. G. A power series beta weibull regression model for predicting breast carcinoma. **Statistics in medicine**, Wiley Online Library, v. 34, n. 8, p. 1366–1388, 2015. Citado na página 77.
- PENG, L.; FINE, J. P. Regression modeling of semicompeting risks data. **Biometrics**, Wiley Online Library, v. 63, n. 1, p. 96–108, 2007. Citado na página 52.
- PENG, M.; XIANG, L.; WANG, S. Semiparametric regression analysis of clustered survival data with semi-competing risks. **Computational Statistics & Data Analysis**, Elsevier, v. 124, p. 53–70, 2018. Citado na página 52.
- PRICE, D. L.; MANATUNGA, A. K. Modelling survival data with a cured fraction using frailty models. **Statistics in medicine**, Wiley Online Library, v. 20, n. 9-10, p. 1515–1527, 2001. Citado na página 78.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>. Citado na página 96.
- _____. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 61 e 65.
- RODRIGUES, J.; CANCHO, V. G.; CASTRO, M. de; LOUZADA-NETO, F. On the unification of long-term survival models. **Statistics & Probability Letters**, Elsevier, v. 79, n. 6, p. 753–759, 2009. Citado nas páginas 25, 38 e 77.
- SANTOS, D. M. dos; DAVIES, R. B.; FRANCIS, B. Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer. **Journal of Statistical Planning and Inference**, Elsevier, v. 47, n. 1-2, p. 111–127, 1995. Citado nas páginas 24, 41 e 42.
- SELLE, M. L. **Modelling of Semi-Competing Risks Using the Illness-Death Model with Shared Frailty**. Dissertação (Mestrado) — NTNU, 2016. Citado na página 66.
- SOUZA, D. de; CANCHO, V. G.; RODRIGUES, J.; BALAKRISHNAN, N. Bayesian cure rate models induced by frailty in survival analysis. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 26, n. 5, p. 2011–2028, 2017. Citado na página 77.
- SPRINGER (Ed.). **Análisis de supervivencia bayesiano**. Citado na página 96.

- THERNEAU, T. M. **A Package for Survival Analysis in R**. [S.l.], 2021. R package version 3.2-13. Disponível em: <<https://CRAN.R-project.org/package=survival>>. Citado na página 67.
- TOURAINÉ, C.; GERDS, T. A.; JOLY, P. Smoothhazard: An r package for fitting regression models to interval-censored observations of illness-death models. **Journal of Statistical Software**, v. 79, p. 1–22, 2017. Citado na página 95.
- TRIVEDI, P. K.; ZIMMER, D. M. *et al.* Copula modeling: an introduction for practitioners. **Foundations and Trends® in Econometrics**, Now Publishers, Inc., v. 1, n. 1, p. 1–111, 2007. Citado na página 31.
- TSODIKOV, A.; IBRAHIM, J.; YAKOVLEV, A. Estimating cure rates from survival data: an alternative to two-component mixture models. **Journal of the American Statistical Association**, Taylor & Francis, v. 98, n. 464, p. 1063–1078, 2003. Citado nas páginas 25, 38 e 77.
- TSODIKOV, A. D.; YAKOVLEV, A. Y.; ASSELAIN, B. **Stochastic models of tumor latency and their biostatistical applications**. [S.l.]: World Scientific, 1996. v. 1. Citado na página 78.
- TWEEDIE, M. An index which distinguishes between some important exponential families. In: **Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference**. [S.l.: s.n.], 1984. v. 579, p. 579–604. Citado nas páginas 24 e 46.
- VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. **Demography**, Springer, v. 16, n. 3, p. 439–454, 1979. Citado nas páginas 23, 24, 37, 41 e 42.
- WANG, W. Estimating the association parameter for copula models under dependent censoring. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 65, n. 1, p. 257–273, 2003. Citado na página 52.
- WIENKE, A. **Frailty models in survival analysis**. Tese (Doutorado), 2007. Citado na página 35.
- WIENKE, A. **Frailty models in survival analysis**. [S.l.]: CRC Press, 2010. Citado nas páginas 21, 23, 24, 25, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 48, 53, 54, 57, 59, 60 e 95.
- XU, J.; KALBFLEISCH, J. D.; TAI, B. Statistical analysis of illness–death processes and semicompeting risks data. **Biometrics**, Wiley Online Library, v. 66, n. 3, p. 716–725, 2010. Citado nas páginas 9, 11, 25, 52, 53, 54, 55, 60 e 67.
- YAKOVLEV, A. Y.; TSODIKOV, A.; ASSELAIN, B. **Stochastic Models of Tumor Latency and Their Biostatistical Applications**. [S.l.]: World Scientific, 1996. v. 1. Citado nas páginas 25, 77 e 88.
- ZHOU, R.; ZHU, H.; BONDY, M.; NING, J. Semiparametric model for semi-competing risks data with application to breast cancer study. **Lifetime data analysis**, Springer, v. 22, n. 3, p. 456–471, 2016. Citado na página 52.
- ZIEGEL, E. **Numerical recipes: The art of scientific computing**. [S.l.]: Taylor & Francis Group, 1987. Citado na página 84.

FUNÇÃO DE VEROSSIMILHANÇA

Primeiro veremos alguns resultados necessários para a obtenção da função densidade de probabilidade. Esses resultados são os seguintes

A.1 Função densidade de probabilidade conjunta

Derivamos a função densidade conjunta denotado por $f(t_1, t_2)$, no quadrante superior $0 < t_1 < t_2$, isto é

$$\begin{aligned} f(t_1, t_2) &= \mathbb{P}[T_1 = t_1, T_2 = t_2, T_2 > t_1] = \mathbb{P}[T_1 = t_1, T_2 = t_2] \\ &= \mathbb{P}[T_1 > t_1, T_2 > t_1] \left(\frac{\mathbb{P}[T_1 = t_1, T_2 > t_1]}{\mathbb{P}[T_1 > t_1, T_2 > t_1]} \right) \left(\frac{\mathbb{P}[T_1 = t_1, T_2 = t_2]}{\mathbb{P}[T_1 = t_1, T_2 \geq t_2]} \right) \left(\frac{\mathbb{P}[T_1 = t_1, T_2 \geq t_2]}{\mathbb{P}[T_1 = t_1, T_2 > t_1]} \right) \end{aligned}$$

logo, a partir da definição 4.2.1 e desde que $S(t_1) = \mathbb{P}[T_1 > t_1, T_2 > t_1]$, $S_{12}(t_2|t_1) = \mathbb{P}[T_2 \geq t_2|T_1 = t_1]$ e $\mathbb{P}[T_2 > t_1|T_1 = t_1] = 1$ obtemos que a função densidade de probabilidade é dado por

$$\begin{aligned} f(t_1, t_2) &= \lambda_1(t_1)\lambda_{12}(t_2|t_1)S(t_1)S_{12}(t_2|t_1) \\ &= \lambda_1(t_1)\lambda_{12}(t_2|t_1) \exp \left\{ - \int_0^{t_1} (\lambda_1(s) + \lambda_2(s)) ds \right\} \exp \left\{ - \int_{t_1}^{t_2} \lambda_{12}(s|t_1) ds \right\} \quad (\text{A.1}) \end{aligned}$$

Lembrando que $S(t)$ é a função de sobrevivência geral ou total.

A.2 Função densidade de probabilidade na linha $T_1 = \infty$

Veja que

$$\begin{aligned} f_\infty(t_2) &= \mathbb{P}[T_1 = \infty, T_2 = t_2] \\ &= \frac{\mathbb{P}[T_1 > t_2, T_2 = t_2]}{\mathbb{P}[T_1 > t_2, T_2 > t_2]} \mathbb{P}[T_1 > t_2, T_2 > t_2], \end{aligned}$$

logo, a partir da definição 4.2.1 e desde que $S(t_2) = \mathbb{P}[T_1 > t_2, T_2 > t_2]$, obtemos que

$$f_\infty(t_2) = \lambda_2(t_2)S(t_2) = \lambda_2(t_2) \exp \left\{ - \int_0^{t_2} (\lambda_1(s) + \lambda_2(s)) ds \right\} \quad (\text{A.2})$$

A.3 Construção da função de verossimilhança

Seja $\boldsymbol{\vartheta} = (\theta, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3)$, o vetor de parâmetros a serem estimados e $(Y_1, Y_2, \delta_1, \delta_2)$ a observação para cada indivíduo onde $Y_2 = \min\{T_2, C\}$, $\delta_2 = I_{\{T_2 \leq C\}}$, $Y_1 = \min\{T_1, Y_2\}$ e $\delta_1 = I_{\{T_1 \leq Y_2\}}$. Construímos a função de Verossimilhança para cada um dos quatro possíveis casos observáveis apresentados na seguinte Tabela.

Tabela 12 – Possíveis observações para dados de riscos semicompetitivos.

| Casos | Cenário | (Y_1, Y_2) | (δ_1, δ_2) |
|-------|--|--------------|------------------------|
| 1 | Recorrência e censura antes da morte | (T_1, C) | $(1, 0)$ |
| 2 | Morte depois da recorrência | (T_1, T_2) | $(1, 1)$ |
| 3 | Morte sem recorrência | (T_2, T_2) | $(0, 1)$ |
| 4 | Censura antes da recorrência e a Morte | (C, C) | $(0, 0)$ |

Caso 1: Quando a primeira observação é o tempo de recorrência e depois a censura, ou seja, $(\delta_1, \delta_2) = (1, 0)$ e $(Y_1, Y_2) = (T_1, C)$, então

$$\begin{aligned} L_1(\boldsymbol{\vartheta}, Z) &= \mathbb{P}[T_1 = Y_1, T_2 > Y_2 | \boldsymbol{\vartheta}, Z] \\ &= \int_{Y_2}^{\infty} f(Y_1, v | \boldsymbol{\vartheta}, Z) dv \end{aligned}$$

usando (A.1) e considerando o modelo condicional dado em (4.4)-(4.6), obtemos que

$$L_1(\boldsymbol{\vartheta}, Z) = Z \lambda_{01}(Y_1 | \boldsymbol{\vartheta}) \exp \left\{ -Z (\Lambda_{01}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{02}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{03}(Y_2 | \boldsymbol{\vartheta}) - \Lambda_{03}(Y_1 | \boldsymbol{\vartheta})) \right\}$$

em que $\Lambda_{0i}(t | \boldsymbol{\vartheta}) = \int_0^t \lambda_{0i}(s | \boldsymbol{\vartheta}) ds$, é a função de risco de base acumulada, para $i = 1, 2, 3$. Logo, integrando sobre a fragilidade Z com distribuição PVF, temos que a função de verossimilhança marginal é dada por

$$\begin{aligned} L_1(\boldsymbol{\vartheta}) &= \lambda_{01}(Y_1 | \boldsymbol{\vartheta}) \left(1 + \frac{\theta (\Lambda_{01}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{02}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{03}(Y_2 | \boldsymbol{\vartheta}) - \Lambda_{03}(Y_1 | \boldsymbol{\vartheta}))}{1 - \gamma} \right)^{\gamma-1} \\ &\times \exp \left\{ \frac{1 - \gamma}{\gamma \theta} \left[1 - \left(1 + \frac{\theta (\Lambda_{01}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{02}(Y_1 | \boldsymbol{\vartheta}) + \Lambda_{03}(Y_2 | \boldsymbol{\vartheta}) - \Lambda_{03}(Y_1 | \boldsymbol{\vartheta}))}{1 - \gamma} \right)^{\gamma} \right] \right\} \end{aligned} \quad (\text{A.3})$$

Caso 2: Quando a primeira observação é o tempo de recorrência e depois a morte, ou seja, $(\delta_1, \delta_2) = (1, 1)$ e $(Y_1, Y_2) = (T_1, T_2)$, então

$$L_2(\boldsymbol{\vartheta}, Z) = \mathbb{P}[T_1 = Y_1, T_2 = Y_2 | \boldsymbol{\vartheta}, Z] = f(Y_1, Y_2 | \boldsymbol{\vartheta}, Z)$$

De forma similar a (A.3) pode-se mostrar que

$$\begin{aligned}
L_2(\boldsymbol{\vartheta}) &= \lambda_{01}(Y_1|\boldsymbol{\vartheta})\lambda_{03}(Y_2|\boldsymbol{\vartheta}) \left(1 + \frac{\theta(\Lambda_{01}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{03}(Y_2|\boldsymbol{\vartheta}) - \Lambda_{03}(Y_1|\boldsymbol{\vartheta}))}{1-\gamma} \right)^{\gamma-2} \\
&= \times \left[\theta + \left(1 + \frac{\theta(\Lambda_{01}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{03}(Y_2|\boldsymbol{\vartheta}) - \Lambda_{03}(Y_1|\boldsymbol{\vartheta}))}{1-\gamma} \right)^\gamma \right] \\
&\times \exp \left\{ \frac{1-\gamma}{\gamma\theta} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_1|\boldsymbol{\vartheta}) + \Lambda_{03}(Y_2|\boldsymbol{\vartheta}) - \Lambda_{03}(Y_1|\boldsymbol{\vartheta}))}{1-\gamma} \right)^\gamma \right] \right\}
\end{aligned} \tag{A.4}$$

Caso 3: Quando o tempo de recorrência foi censurado pelo tempo de morte, ou seja, $(\delta_1, \delta_2) = (0, 1)$ e $(Y_1, Y_2) = (T_1, T_2)$, então

$$L_3(\boldsymbol{\vartheta}, Z) = \mathbb{P}[T_1 = \infty, T_2 = Y_2 | \boldsymbol{\vartheta}, Z] = f_\infty(Y_2 | \boldsymbol{\vartheta}, Z)$$

usando A.2, temos

$$L_3(\boldsymbol{\vartheta}, Z) = Z\lambda_{02}(Y_2|\boldsymbol{\vartheta}) \exp \{-Z(\Lambda_{01}(Y_2|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_2|\boldsymbol{\vartheta}))\}$$

logo, de forma similar a A.3, temos que

$$\begin{aligned}
L_3(\boldsymbol{\vartheta}) &= \lambda_{02}(Y_2|\boldsymbol{\vartheta}) \left(1 + \frac{\theta(\Lambda_{01}(Y_2|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_2|\boldsymbol{\vartheta}))}{1-\gamma} \right)^{\gamma-1} \\
&\times \exp \left\{ \frac{1-\gamma}{\gamma\theta} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(Y_2|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_2|\boldsymbol{\vartheta}))}{1-\gamma} \right)^\gamma \right] \right\}
\end{aligned} \tag{A.5}$$

Caso 4: Quando os dois tempos de falha são cesurados, ou seja, $(\delta_1, \delta_2) = (0, 0)$ e $(Y_1, Y_2) = (C, C) = (Y_2, Y_2) = (Y_1, Y_1)$, então

$$\begin{aligned}
L_4(\boldsymbol{\vartheta}, Z) &= \mathbb{P}[T_1 > Y_2, T_2 > Y_2 | \boldsymbol{\vartheta}, Z] \\
&= \exp \{-Z(\Lambda_{01}(Y_2|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_2|\boldsymbol{\vartheta}))\}
\end{aligned}$$

Logo, usando o Laplaciano da distribuição PVF, a função verossimilhança marginal é dado por

$$L_4(\boldsymbol{\vartheta}) = \exp \left\{ \frac{1-\gamma}{\gamma\theta} \left[1 - \left(1 + \frac{\theta(\Lambda_{01}(Y_2|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_2|\boldsymbol{\vartheta}))}{1-\gamma} \right)^\gamma \right] \right\} \tag{A.6}$$

Assim juntando (A.3), (A.4), (A.5) e (A.6) e para uma amostra de n , a função de verossimilhança é dado por

$$\begin{aligned}
L(\boldsymbol{\vartheta}) &= \prod_{j=1}^n \lambda_{01}(Y_{j1}|\boldsymbol{\vartheta})^{\delta_{j1}} \lambda_{02}(Y_{j2}|\boldsymbol{\vartheta})^{\delta_{j2}(1-\delta_{j1})} \lambda_{03}(Y_{j2}|\boldsymbol{\vartheta})^{\delta_{j1}\delta_{j2}} \left\{ \theta + \left(1 + \frac{\theta K_{j1}}{1-\gamma} \right)^\gamma \right\}^{\delta_{j1}\delta_{j2}} \\
&\times \left(1 + \frac{\theta K_{j1}}{1-\gamma} \right)^{\delta_{j1}(\gamma-1-\delta_{j2})} \left\{ \exp \left(\frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta K_{j1}}{1-\gamma} \right)^\gamma \right] \right) \right\}^{\delta_{j1}} \\
&\times \left(1 + \frac{\theta K_{j2}}{1-\gamma} \right)^{(\gamma-1)\delta_{j2}(1-\delta_{j1})} \left\{ \exp \left(\frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta K_{j2}}{1-\gamma} \right)^\gamma \right] \right) \right\}^{1-\delta_{j1}}
\end{aligned} \tag{A.7}$$

onde $K_{j1} = \Lambda_{01}(Y_{j1}|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_{j1}|\boldsymbol{\vartheta}) + \Lambda_{03}(Y_{j2}|\boldsymbol{\vartheta}) - \Lambda_{03}(Y_{j1}|\boldsymbol{\vartheta})$, $K_{j2} = \Lambda_{01}(Y_{j2}|\boldsymbol{\vartheta}) + \Lambda_{02}(Y_{j2}|\boldsymbol{\vartheta})$ para $j = 1, \dots, n$.

PROBABILIDADE DE TRANSIÇÃO MARGINAL

(1) Probabilidade de transição para o estado de recorrência:

Veja que $\mathbb{P}_{01}(t)$ é probabilidade de estar no estado 1 no tempo t dado que o estado 0 começou no tempo $t = 0$ que podemos expressar como a probabilidade de não ter transição fora do estado 0 antes da transição de interesse multiplicado pela transição para o estado 1 antes do tempo t multiplicado por a probabilidade de permanecer no estado 1 até o tempo t , isto é

$$\begin{aligned}\mathbb{P}_{01}(t) &= \mathbb{P}[\text{Estar no estado 1 no tempo } t | \text{O estado 0 começou no tempo } t = 0] \\ &= \mathbb{P}[\text{não ter transição fora do estado 0 antes da transição de interesse}] \\ &\quad \times (\text{transição para o estado 1 antes do tempo } t) \\ &\quad \times \mathbb{P}[\text{permanecer no estado 1 até o tempo } t]\end{aligned}$$

Usando as taxas condicionais e por tanto tomando a media sobre Z e desde que o tempo de transição para o estado 1 é desconhecido também devemos integrar sobre todo t , isto é

$$\begin{aligned}\mathbb{P}_{01}(t) &= \int_0^\infty \int_0^t \mathbb{P}[\text{não ter transição fora do estado 0 antes do tempo } s | z] \\ &\quad \times (\text{transição para o estado 1 no tempo } s \text{ dado } z) \\ &\quad \times \mathbb{P}[\text{permanecer no estado 1 até o tempo } t | z] g(z) ds dz,\end{aligned}$$

onde $g(z)$ representa a função densidade de probabilidade da variável de fragilidade Z . Isto é equivalente a

$$\mathbb{P}_{01}(t) = \int_0^\infty \int_0^t S(s|z) \lambda_1(s|z) S_{12}(t|s, z) g(z) ds dz$$

onde $S(s|Z)$, $\lambda_1(s|Z)$ e $S_{12}(t|s, Z)$ são dados em (4.7), (4.4) e (4.8), respectivamente. Desta forma

$$\mathbb{P}_{01}(t) = \int_0^t \lambda_{01}(s) \int_0^\infty e^{-z(\Lambda_{01}(s) + \Lambda_{02}(s) + \Lambda_{03}(s,t))} z g(z) ds dz,$$

onde $\Lambda_{0i}(t) = \int_0^t \lambda_{0i}(s) ds$ para $i = 1, 2, 3$, são as funções de risco de base acumulado e $\Lambda_{03}(u, t) = \Lambda_{03}(t) - \Lambda_{03}(u)$. Concluímos usando a transformada de Laplace da fragilidade Z com distribuição

PVF com media 1 e variância $\theta > 0$, dada em (3.51), desta forma obtemos,

$$\mathbb{P}_{01}(t) = \int_0^t \lambda_{01}(s) \left(1 + \frac{\theta s_1(s,t)}{1-\gamma}\right)^{\gamma-1} \exp\left\{\frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta s_1(s,t)}{1-\gamma}\right)^\gamma\right]\right\} ds.$$

onde $s_1(t, t_1) = \Lambda_{01}(s) + \Lambda_{02}(s) + \Lambda_{03}(s, t)$.

(2) Probabilidade de transição para o estado de morte:

A derivação de $\mathbb{P}_{02}(t)$ é feita de forma análoga e é dada por

$$\mathbb{P}_{02}(t) = \int_0^t \lambda_{02}(s) \left(1 + \frac{\theta s_2(s)}{1-\gamma}\right)^{\gamma-1} \exp\left\{\frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta s_2(s)}{1-\gamma}\right)^\gamma\right]\right\} ds.$$

onde $s_2(s) = \Lambda_{01}(ts) + \Lambda_{02}(s)$. Nesta função, não temos uma contribuição para permanecer no estado 2 até o tempo t , pois o estado 2 é absorvente.

(3) Probabilidade de transição do estado de recorrência para o estado de morte:

Passamos para a probabilidade de transição $\mathbb{P}_{12}(t|t_1)$, a qual é condicional no estado 1 que foi atingido no tempo t_1 . Então a probabilidade de estar no estado 2 no tempo t é dada pela probabilidade de permanecer no estado 1 até algum tempo s e ter uma transição para o estado 2 no tempo $s \leq t$. De forma similar as probabilidades de transição anteriores e novamente, como o tempo de transição do estado 1 para o estado 2 é desconhecido, integramos os possíveis tempos de transição. Usando taxas de transição condicionais e média sobre Z , obtemos o seguinte

$$\mathbb{P}_{12}(t|t_1) = \int_0^\infty \int_0^t S_{12}(s|t_1, z) \lambda_{12}(s|t_1, z) g(z) ds dz$$

onde $S_{12}(t|s, Z)$ e $\lambda_{12}(s|t_1, z)$ são dados em (4.8) e (4.6), respectivamente. Desta forma

$$\mathbb{P}_{01}(t) = \int_0^t \lambda_{03}(s) \int_0^\infty e^{-z\Lambda_{03}(s,t)} z g(z) ds dz$$

Finalmente usando o Laplaciano de Z , e para $s_3(s, t_1) = \Lambda_{03}(t_1, s)$. obtemos

$$\mathbb{P}_{12}(t|t_1) = \int_0^t \lambda_{02}(s) \left(1 + \frac{\theta s_3(s, t)}{1-\gamma}\right)^{\gamma-1} \exp\left\{\frac{1-\gamma}{\theta\gamma} \left[1 - \left(1 + \frac{\theta s_3(s, t)}{1-\gamma}\right)^\gamma\right]\right\} ds.$$

FUNÇÃO DE SOBREVIVÊNCIA CONJUNTA CONDICIONAL

C.1 Função de sobrevivência conjunta condicional em ω

Sejam T_1, \dots, T_m variáveis aleatórias que representam os m tipos de tempos observáveis ao evento de interesse, tal que a distribuição de T_k ($k = 1, \dots, m$) pode ser representado por um modelo básico de fragilidade dado em (5.1). Queremos provar que

$$S(t_1, \dots, t_m | \omega) = \exp \left\{ -\omega \left(\sum_{k=1}^m \rho_k F_{0k}(t_k) \right) \right\}, \quad (\text{C.1})$$

Considerando o fato que T_k e Z_k , ($k = 1, \dots, m$) são condicionalmente independentes, além disso, Z_i e Z_j ($i \neq j = 1, \dots, m$) são condicionalmente independentes (dado ω), pode-se mostrar que

$$\begin{aligned} \mathbb{P}[T_1 > t_1, \dots, T_m > t_m | \omega] &= \prod_{k=1}^m \sum_{z_k \geq 0} (S(t_k | Z_k) \mathbb{P}[Z_k | \omega]) \\ &= \prod_{k=1}^m \psi_{Z_k}(S_{0k}(t_k)) \end{aligned}$$

A última igualdade é obtida a partir do modelo (5.1), onde podemos obter que $S(t_k | Z_k) = [S_{0k}(t_k)]^{Z_k}$, para mais detalhe veja a Seção 3.1, em que $\psi_{Z_k}(\cdot)$ é a função geradora de probabilidade da variável de fragilidade Z_k , para $k = 1, \dots, m$. Logo assumindo que Z_k ($k = 1, \dots, m$) tem distribuição Poisson com media $\omega \rho_k$ onde $\rho_k > 0$ é constante, enquanto ω é uma componente de fragilidade. Temos que, $\psi_{Z_k}(s) = e^{-\omega \rho_k (1-s)}$, para $k = 1, \dots, m$, desta forma

$$S(t_1, \dots, t_m | \omega) = \prod_{k=1}^m e^{-\omega \rho_k (1-S_{0k}(t_k))} = \exp \left\{ -\omega \left(\sum_{k=1}^m \rho_k (F_{0k}(t_k)) \right) \right\}, \quad (\text{C.2})$$

onde $F_{0k}(t_k)$ é a função de distribuição acumulada de base para o tempo T_k .

FUNÇÃO DE VEROSSIMILHANÇA

D.1 Função verossimilhança para dados completos

Suponha que temos n indivíduos na amostra, e seja Z_{kj} o número de riscos latentes do tipo k do evento de interesse para o j -ésimo indivíduo, com $j = 1, \dots, n$ e $k = 1, 2$. Além disso, assumimos que os Z_{kj} são variáveis aleatórias Poisson independentes com média $\omega_j \rho_k$, $j = 1, \dots, n$ e $k = 1, 2$. Também assumimos que as ω_j são variáveis aleatórias i.i.d com distribuição PVF de média 1 e variância $1 - \gamma = \theta > 0$. Seja T_{kj} o tempo de falha para o j -ésimo indivíduo e o k -ésimo tipo de evento de interesse, onde T_{kj} pode ser censurado pela direita. Suponha que a distribuição de T_{kj} , com $j = 1, \dots, n$ e $k = 1, 2$, pode ser representado por um modelo básico de fragilidade

$$\lambda_k(t_{kj}|Z_{kj}) = Z_{kj} \lambda_{0k}(t_{kj}|\boldsymbol{\eta}_k)$$

onde $\lambda_{0k}(t_{kj}|\boldsymbol{\eta}_k)$ é a função de risco de base para T_{kj} com vetor de parâmetros $\boldsymbol{\eta}_k$. Suponha também que Z_{kj} e T_{kj} são condicionalmente independentes. Denote por C_{kj} os tempos de censura de modo que

$$y_{jk} = \min\{T_{kj}, C_{kj}\} \text{ e } \delta_{kj} = \begin{cases} 1, & T_{kj} \leq C_{kj} \\ 0, & T_{kj} > C_{kj} \end{cases}$$

Assim, sejam $\mathbf{y}_k = (y_{k1}, \dots, y_{kn})$, $\boldsymbol{\delta}_k = (\delta_{k1}, \dots, \delta_{kn})$, $\mathbf{Z}_k = (Z_{k1}, \dots, Z_{kn})$ com $k = 1, 2$ e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$. Logo, os dados completos são dados por $D = (n, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \mathbf{Z}_1, \mathbf{Z}_2, \boldsymbol{\omega})$, onde $\mathbf{Z}_1, \mathbf{Z}_2$ e $\boldsymbol{\omega}$ são vetores aleatórios não observáveis, então os dados observados são dados por $D_{obs} = (n, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. Adicionalmente seja $\boldsymbol{\vartheta} = (\gamma, \rho_1, \rho_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, o vetor de parâmetros a serem estimados. A função verossimilhança para $\boldsymbol{\vartheta}$ dado os dados completos pode ser construída da seguinte forma. Considerando o fato de que as observações dos tempos de eventos são condicionalmente independentes dada as fragilidades, além disso, Z_{1j} e Z_{2j} são condicionalmente independentes dado $\boldsymbol{\omega}$, podemos obter

$$L(\boldsymbol{\vartheta}|D) = \prod_{j=0}^n f_1(y_{1j}, \delta_{1j}|Z_{1j}, \omega_j, \boldsymbol{\eta}_1) f_2(y_{2j}, \delta_{2j}|Z_{2j}, \omega_j, \boldsymbol{\eta}_2) f_1(Z_{1j}|\omega_j, \rho_1) f_2(Z_{2j}|\omega_j, \rho_1) g(\omega_j|\gamma),$$

onde $g(\cdot)$ é a função densidade da componente de fragilidade $\boldsymbol{\omega}$. Logo usando o fato de que $f(t, \delta) = f(t)^\delta S(t)^{1-\delta} = (\lambda(t)S(t))^\delta S(t)^{1-\delta}$ e $\lambda(t) = \frac{f(t)}{S(t)}$ em análise de sobrevivência, e usando o modelo de

fragilidade dada em (5.1), considerando $Z_{kj}|\omega_j \sim \text{Poisson}(\rho_k \omega_j)$, com $k = 1, 2$ e $j = 1, \dots, n$, podemos obter

$$\begin{aligned} L(\boldsymbol{\vartheta}|D) &= \prod_{j=0}^n Z_{1j}^{\delta_{1j}} f_{01}(y_{1j}|\boldsymbol{\eta}_1)^{\delta_{1j}} (S_{01}y_{1j}|\boldsymbol{\eta}_1)^{Z_{1j}-\delta_{1j}} Z_{2j}^{\delta_{2j}} f_{02}(y_{2j}|\boldsymbol{\eta}_2)^{\delta_{2j}} S_{02}(y_{2j}|\boldsymbol{\eta}_2)^{Z_{2j}-\delta_{2j}} \\ &\times \prod_{j=0}^n \frac{e^{-\rho_1 \omega_j} (\rho_1 \omega_j)^{Z_{1j}}}{Z_{1j}!} \frac{e^{-\rho_2 \omega_j} (\rho_2 \omega_j)^{Z_{2j}}}{Z_{2j}!} g(\omega_j|\gamma). \end{aligned}$$

Logo, podemos reescrever $L(\boldsymbol{\vartheta}|D)$ como

$$\begin{aligned} L(\boldsymbol{\vartheta}|D) &= \prod_{k=1}^2 \prod_{j=1}^n [S_{0k}(y_{kj}|\boldsymbol{\eta}_k)]^{Z_{kj}-\delta_{kj}} [Z_{kj} f_{0k}(y_{kj}|\boldsymbol{\eta}_k)]^{\delta_{kj}} \\ &\times \prod_{k=1}^2 \exp \left\{ \sum_{j=1}^n [Z_{kj} \log(\omega_j \rho_k) - \log(Z_{kj}!) - \omega_j \rho_k] \right\} \prod_{j=1}^n g(\omega_j). \end{aligned} \quad (\text{D.1})$$

D.2 Função verossimilhança para dados observados

D.2.1 Prova do Teorema 5.3.1

Demonstração. Para obter a função verossimilhança para dados observados $L(\boldsymbol{\vartheta}|D_{obs})$, dada em (5.18), devemos fazer a soma e a integração sobre todos valores de Z_{kj} e ω_j , respectivamente, para $k = 1, 2$ e $j = 1, \dots, n$. Para isto, a partir de (D.1) temos que

$$\begin{aligned}
L(\boldsymbol{\vartheta}|D_{obs}) &= \prod_{k=1}^2 \prod_{j=0}^n \int_0^\infty \left(\frac{f_{0k}(y_{kj}|\boldsymbol{\eta}_k)}{S_{0k}(y_{kj}|\boldsymbol{\eta}_k)} \right)^{\delta_{kj}} \sum_{Z_{kj}} \left(\frac{Z_{kj}^{\delta_{kj}} (\rho_k \omega_j S_{0k}(y_{kj}|\boldsymbol{\eta}_k))^{Z_{kj}}}{Z_{kj}!} \right) e^{-\rho_k \omega_j} g(\omega_j|\gamma) d\omega_j \\
&= \prod_{k=1}^2 \prod_{j=0}^n \left(\frac{f_{0k}(y_{kj}|\boldsymbol{\eta}_k)}{S_{0k}(y_{kj}|\boldsymbol{\eta}_k)} \right)^{\delta_{kj}} \\
&\quad \times \prod_{k=1}^2 \prod_{j=0}^n \int_0^\infty e^{-\rho_k \omega_j (1 - S_{0k}(y_{kj}|\boldsymbol{\eta}_k))} \sum_{Z_{kj}} \left(\frac{Z_{kj}^{\delta_{kj}} (\rho_k \omega_j S_{0k}(y_{kj}|\boldsymbol{\eta}_k))^{Z_{kj}} e^{-\rho_k \omega_j S_{0k}(y_{kj}|\boldsymbol{\eta}_k)}}{Z_{kj}!} \right) g(\omega_j|\gamma) d\omega_j \\
&= \prod_{k=1}^2 \prod_{j=0}^n \left(\frac{f_{0k}(y_{kj}|\boldsymbol{\eta}_k)}{S_{0k}(y_{kj}|\boldsymbol{\eta}_k)} \right)^{\delta_{kj}} \int_0^\infty e^{-\rho_k \omega_j (1 - S_{0k}(y_{kj}|\boldsymbol{\eta}_k))} (\rho_k \omega_j S_{0k}(y_{kj}|\boldsymbol{\eta}_k))^{\delta_{kj}} g(\omega_j|\gamma) d\omega_j \\
&= \prod_{k=1}^2 \prod_{j=0}^n \rho_k^{\delta_{kj}} f_{0k}(y_{kj}|\boldsymbol{\eta}_k)^{\delta_{kj}} \int_0^\infty \omega_j^{\delta_{kj}} e^{-\omega_j \rho_k (1 - S_{0k}(y_{kj}|\boldsymbol{\eta}_k))} g(\omega_j|\gamma) d\omega_j \\
&= \prod_{j=1}^n \rho_1^{\delta_{1j}} \rho_2^{\delta_{2j}} \left(f_{01}(y_{1j}|\boldsymbol{\eta}_1)^{\delta_{1j}} f_{02}(y_{2j}|\boldsymbol{\eta}_2)^{\delta_{2j}} \right) \\
&\quad \times \prod_{j=1}^n \int_0^\infty \omega_j^{\delta_{2j} + \delta_{2j}} e^{-\omega_j (\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))} g(\omega_j|\gamma) d\omega_j,
\end{aligned}$$

onde $F_{02}(y_{kj}|\boldsymbol{\eta}_k)$ é a função de distribuição acumulada de base. A partir de aqui, construímos a função de Verossimilhança para cada um dos quatro possíveis casos de $(\delta_{1j}, \delta_{2j})$, para o j -ésimo indivíduo, isto é

Caso 1: $(\delta_{1j}, \delta_{2j}) = (0, 0)$, ou seja $(y_{1j}, y_{2j}) = (C_{1j}, C_{2j})$, ambos tempos de eventos foram censurados. Então a função verossimilhança é dado por:

$$L_{00} = \int_0^\infty e^{-\omega_j (\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))} g(\omega_j|\gamma) d\omega_j.$$

Esta integral é a transformada de Laplace da fragilidade ω_j , aplicado em $\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)$, isto é,

$$L_{00} = \mathcal{L}_{\omega_j}(\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)).$$

Caso 2: $(\delta_{1j}, \delta_{2j}) = (1, 0)$, ou seja $(y_{1j}, y_{2j}) = (T_{1j}, C_{2j})$, o primeiro evento foi observado e o segundo foi censurado. Então a função verossimilhança é dado por:

$$L_{10} = \rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1) \int_0^\infty \omega_j e^{-\omega_j (\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))} g(\omega_j|\gamma) d\omega_j.$$

Usando a propriedade da transformada de Laplace $\int_0^\infty \omega e^{-\omega s} g(\omega) d\omega = -\frac{\partial}{\partial s} \mathcal{L}_\omega(s)$, temos

$$\mathbf{L}_{10} = -\rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1) \frac{\partial}{\partial s} \mathcal{L}_{\omega_j}(s).$$

em que $s = \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)$.

Caso 3: $(\delta_{1j}, \delta_{2j}) = (0, 1)$, ou seja $(y_{1j}, y_{2j}) = (C_{1j}, T_{2j})$, o primeiro evento foi censurado e o segundo foi observado. De forma análogo ao Caso 2, pode-se mostrar que

$$\mathbf{L}_{01} = -\rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_1) \frac{\partial}{\partial s} \mathcal{L}_{\omega_j}(s).$$

em que $s = \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)$.

Caso 4: $(\delta_{1j}, \delta_{2j}) = (1, 1)$, ou seja $(y_{1j}, y_{2j}) = (T_{1j}, T_{2j})$, ambos tempos de eventos foram observados. Então a função verossimilhança é dado por:

$$\mathbf{L}_{11} = \rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1) \rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_2) \int_0^\infty \omega_j^2 e^{-\omega_j(\rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))} g(\omega_j|\gamma) d\omega_j.$$

Usando a propriedade da transformada de Laplace $\int_0^\infty \omega^2 e^{-\omega s} g(\omega) d\omega = \frac{\partial^2}{\partial s^2} \mathcal{L}_\omega(s)$, temos

$$\mathbf{L}_{11} = \rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1) \rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_2) \frac{\partial^2}{\partial s^2} \mathcal{L}_{\omega_j}(s).$$

em que $s = \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)$.

Juntando todos os casos e considerando todos os indivíduos a função verossimilhança para dados observados e considerando a transformada de Laplace do modelo de fragilidade PVF dada em (3.61), a função de verossimilhança é dada por:

$$\begin{aligned} \mathbf{L}(\boldsymbol{\vartheta}|D_{obs}) &= \prod_{j=1}^n (\rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1))^{\delta_{1j}} (\rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_2))^{\delta_{2j}} \left(\exp \left\{ -\frac{1}{\gamma} ((1+s)^\gamma - 1) \right\} \right)^{(1-\delta_{1j})(1-\delta_{2j})} \\ &\times \prod_{j=1}^n \left(\exp \left\{ -\frac{1}{\gamma} ((1+s)^\gamma - 1) \right\} (1+s)^{\gamma-1} \right)^{\delta_{1j}(1-\delta_{2j}) + \delta_{2j}(1-\delta_{1j})} \\ &\times \prod_{j=1}^n \left(\exp \left\{ -\frac{1}{\gamma} ((1+s)^\gamma - 1) \right\} (1+s)^{\gamma-2} (1-\gamma + (1+s)^\gamma) \right)^{\delta_{1j}\delta_{2j}} \end{aligned}$$

onde $s = \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2)$. Ou equivalentemente

$$\begin{aligned} \mathbf{L}(\boldsymbol{\vartheta}|D_{obs}) &= \prod_{j=1}^n (\rho_1 f_{01}(y_{1j}|\boldsymbol{\eta}_1))^{\delta_{1j}} (\rho_2 f_{02}(y_{2j}|\boldsymbol{\eta}_2))^{\delta_{2j}} \\ &\times \prod_{j=1}^n \exp \left\{ -\frac{1}{\gamma} [(1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^\gamma - 1] \right\} \\ &\times \prod_{j=1}^n (1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^{(1-\gamma)(\delta_{1j} + \delta_{2j}) - \gamma \delta_{1j} \delta_{2j}} \\ &\times \prod_{j=1}^n [1 - \gamma + (1 + \rho_1 F_{01}(y_{1j}|\boldsymbol{\eta}_1) + \rho_2 F_{02}(y_{2j}|\boldsymbol{\eta}_2))^\gamma]^{\delta_{1j} \delta_{2j}}, \end{aligned} \quad (\text{D.2})$$

onde $f_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ e $F_{0k}(y_{kj}|\boldsymbol{\eta}_k)$ são as funções de densidade e distribuição acumulada de base, respectivamente, para $j = 1, \dots, n$ e $k = 1, 2$. \square

