

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Classe de modelos de fragilidade com efeito do acúmulo de reparos em múltiplos sistemas reparáveis

Amanda Morales Eudes D'Andrea

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

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

AMANDA MORALES EUDES D'ANDREA

**CLASSE DE MODELOS DE FRAGILIDADE COM EFEITO DO ACÚMULO DE REPAROS EM
MÚLTIPLOS SISTEMAS REPARÁVEIS**

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre ou Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientadora: Profa. Dra. Vera Lucia Damasceno Tomazella

**São Carlos
Janeiro de 2020**

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

D178c D'Andrea, Amanda Morales Eudes
Classe de modelos de fragilidade com efeito do
acúmulo de reparos em múltiplos sistemas reparáveis
/ Amanda Morales Eudes D'Andrea; orientadora Vera
Lucia Damasceno Tomazella. -- São Carlos, 2020.
132 p.

Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2020.

1. Sistemas reparáveis. 2. reparo mínimo. 3.
efeito de acumulação. 4. fragilidade. I. Tomazella,
Vera Lucia Damasceno, orient. II. Título.

Amanda Morales Eudes D'Andrea

**Class of frailty models with repair accumulation effect on
multiple repairable systems**

Doctoral dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Doctorate Interagency Program Graduate in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Vera Lucia
Damasceno Tomazella

USP – São Carlos
January 2020



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado da candidata Amanda Morales Eudes D'Andrea, realizada em 13/12/2019:

Profa. Dra. Vera Lucia Damasceno Tomazella
UFSCar

Prof. Dr. Francisco Louzada Neto
USP

Prof. Dr. Paulo Henrique Ferreira da Silva
UFBA

Prof. Dr. Pedro Luiz Ramos
UFSCar

Profa. Dra. Daniele Cristina Tita Granzotto
UFSCar

AGRADECIMENTOS

Sou muito grata à minha orientadora, Vera L. D. Tomazella, por sua orientação, conhecimento, ensinamentos, apoio, amizade e incentivo durante a preparação desta tese e durante a minha vida acadêmica. Muito obrigada, Vera, por toda confiança e paciência que você teve comigo, sem você eu não teria chegado tão longe.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e, por isto, sou muito grata a esta fundação.

A todos os professores e funcionários do Departamento de Estatística da UFSCar e do Departamento de Estatística da USP-ICMC.

Aos colegas Marco Pollo e Pedro Ramos e ao professor Francisco Louzada pela valiosa contribuição neste trabalho.

À Bruna Verssani por disponibilizar parte dos dados de sua dissertação de mestrado para motivação e aplicação da minha pesquisa.

Aos colegas da UFSCar e da USP pela amizade e por me apoiarem nos momentos difíceis, em especial a Vanessa Rufino, George Lucas e Cirdêmia Costa.

Agradeço à minha família, especialmente meu pai Antônio Eudes, minha mãe Márcia Morales e minha irmã Fernanda Eudes pelo constante apoio.

Finalmente, minha mais profunda gratidão é para meu esposo Thiago Lima D'Andrea, por seu amor, companhia, paciência, força, por sempre acreditar em mim e por seu apoio que, sem dúvida, foram essenciais nessa jornada.

RESUMO

D'ANDREA, A. M. E. **Classe de modelos de fragilidade com efeito do acúmulo de reparos em múltiplos sistemas reparáveis**. 2020. 132 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, 2020.

Em sistemas reparáveis, um aspecto fundamental a ser considerado é prever a confiabilidade dos sistemas em estudo. No entanto, os métodos padrões utilizados para analisar dados de sistemas reparáveis ignoram o efeito de acumulação de ocorrências dos tempos de parada para o reparo e a heterogeneidade dos sistemas. Assim, neste trabalho será explorada uma classe geral de modelos de fragilidade em sistemas reparáveis, a qual utiliza um termo que permite verificar o efeito da acumulação da ocorrência de reparos e um termo de fragilidade, caracterizado pelo uso de um efeito aleatório, ou seja, uma variável aleatória não-observável que representa a informação que não pôde ou não foi observada. O método inferencial clássico e Bayesiano para estimar os parâmetros será mostrado para modelos em sistemas reparáveis sob a suposição de reparo mínimo. Para a metodologia proposta, serão considerados estudos de simulação e aplicações a um conjunto de dados de colheitadeiras de cana-de-açúcar e a um conjunto de dados de caminhões basculantes.

Palavras-chave: Sistemas reparáveis, reparo mínimo, efeito de acumulação, fragilidade.

ABSTRACT

D'ANDREA, A. M. E. **Class of frailty models with repair accumulation effect on multiple repairable systems**. 2020. 132 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, 2020.

In repairable systems, a fundamental aspect to be considered is predict the reliability of the systems being studied. However, the standard methods used to analyze repairable system data ignore the cumulative effect of stopping time to the repair and the heterogeneity of systems. Thus, this work explores a general class of frailty models in repairable systems that uses a term that allows to verify the effect of the accumulation of the occurrence of repairs and a term of frailty, characterized by the use of a random effect, that is, an unobservable random variable that represents information that could not or was not observed. The inferential method to estimate the parameters will be shown to models in repairable systems under minimal repair. For the proposed methodology, it will be considered simulation studies and applications to a sugarcane harvesters data set and a dump truck data set.

Keywords: Repairable systems, minimal repair, effect of accumulating, frailty.

LISTA DE ILUSTRAÇÕES

Figura 1 – Tempos de falha (em dias) para cada colheitadeira de cana-de-açúcar.	24
Figura 2 – Tempos de falha (em dias) para cada caminhão basculante.	25
Figura 3 – Representação gráfica da variável T e do processo de contagem $N(t)$	31
Figura 4 – Função de intensidade do PLP com $\eta = 1$ e diferentes valores de β	36
Figura 5 – Exemplo de tempos de falha truncados por tempo (falhas são representadas pelo símbolo “x”).	38
Figura 6 – Função de intensidade do modelo de RM.	40
Figura 7 – Função de intensidade do modelo de RP.	41
Figura 8 – Funções de intensidade dos modelos ARA e ARI considerando o PLP.	44
Figura 9 – Histórico de falhas de um grupo de 5 sistemas reparáveis.	54
Figura 10 – Vício, EQM para estimativa dos parâmetros para os diferentes métodos de estimação considerando o cenário com $\beta = 3, \delta = 5$	64
Figura 11 – PC das estimativas dos parâmetros para os diferentes métodos de estimação para o cenário com $\beta = 3, \delta = 5$	64
Figura 12 – Vício, EQM das estimativas dos parâmetros para os diferentes métodos de estimação para o cenário com $\beta = 0.5, \delta = 10$	65
Figura 13 – PC das estimativas dos parâmetros para os diferentes métodos de estimação considerando o cenário com $\beta = 0.5, \delta = 10$	65
Figura 14 – Estimativa da MCF para os dados de colheitadeiras de cana-de-açúcar.	66
Figura 15 – Duane plot para as colheitadeiras de cana-de-açúcar.	67
Figura 16 – Estimativa da MCF para os dados dos caminhões basculantes.	68
Figura 17 – Duane plot para os caminhões basculantes.	69
Figura 18 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$	80
Figura 19 – Vício, EQM das estimativas com diferentes tamanhos amostrais para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 0,5$ e $\delta = 5$	82
Figura 20 – Função de intensidade do modelo de fragilidade ajustado aos dados das colheitadeiras.	84
Figura 21 – Função de intensidade do modelo de fragilidade ajustado aos dados dos caminhões.	86
Figura 22 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais e para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$ e $\gamma = 0,5$	98

Figura 23 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais e para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$ e $\gamma = 1,5$	100
Figura 24 – Comparação do KM para eventos recorrentes com a função de confiabilidade estimada dos modelos para os dados de colheitadeiras de cana-de-açúcar. . .	102
Figura 25 – Comparação do KM para eventos recorrentes com a função de confiabilidade estimada dos modelos para os dados de caminhões basculantes.	104
Figura 26 – Função de intensidade do modelo de fragilidade não-paramétrico ajustado aos dados das colheitadeiras.	115
Figura 27 – Função de intensidade do modelo de fragilidade não-paramétrico ajustado aos dados dos caminhões.	117

LISTA DE TABELAS

Tabela 1 – Dados do tempo de quebra da lâmina <i>Chopper</i> localizada no sistema de corte de colheitadeiras de cana-de-açúcar.	23
Tabela 2 – Esboço dos tempos de falha dos caminhões basculantes	24
Tabela 3 – Vício e EQM das estimativas considerando os três métodos de Estimação.	62
Tabela 4 – Probabilidades de cobertura para os parâmetros considerando os três métodos de Estimação.	63
Tabela 5 – Estimativas para os parâmetros do modelo.	67
Tabela 6 – Estimativas para os parâmetros do modelo.	70
Tabela 7 – Cenário com $\beta = 3$ e $\delta = 5$	79
Tabela 8 – Cenário com $\beta = 0.5$ e $\delta = 5$	81
Tabela 9 – Estimativas dos parâmetros do modelo proposto	83
Tabela 10 – Estimativas dos parâmetros para o modelo proposto	84
Tabela 11 – Simulação com o cenário: $\beta = 1,5$, $\delta = 5$ e $\gamma = 0,5$	97
Tabela 12 – Simulação com o cenário: $\beta = 1,5$, $\delta = 5$ e $\gamma = 1,5$	99
Tabela 13 – Estimativas dos modelos para os dados das colheitadeiras.	101
Tabela 14 – Critérios AIC, BIC e log-verossimilhança dos modelos para os dados das colheitadeiras.	102
Tabela 15 – Estimativas dos modelos para os dados dos caminhões basculantes.	103
Tabela 16 – Critérios AIC, BIC e log-verossimilhança dos modelos para os dados dos caminhões basculantes.	104
Tabela 17 – Estimativas dos parâmetros do modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica para os dados de colheitadeiras	114
Tabela 18 – Estimativas dos parâmetros da CGMF sob RM e fragilidade não-paramétrica para os dados de colheitadeiras	115
Tabela 19 – Estimativas dos parâmetros do modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica para os dados de caminhões basculantes.	116
Tabela 20 – Estimativas dos parâmetros da CGMF sob RM e fragilidade não-paramétrica para os dados de caminhões basculantes.	117

LISTA DE ABREVIATURAS E SIGLAS

ABAO	tão ruim quanto velho, do inglês <i>As Bad as Old</i>
AGAN	tão bom quanto novo, do inglês <i>As Good as New</i>
ARA	Redução Aritmética de Idade, do inglês <i>Arithmetic Reduction of Age</i>
ARI	Redução Aritmética de Intensidade, do inglês <i>Arithmetic Reduction of Intensity</i>
CGMF	Classe Geral de Modelos de Fragilidade com efeito da idade virtual
EMV	estimador de máxima verossimilhança
EP	erro padrão
EQM	erro quadrático médio
fda	função de distribuição acumulada
fdp	função de densidade de probabilidade
IC	intervalo de confiança
iid	independente e identicamente distribuídos
ISO	Organização Internacional de Padronização, do inglês <i>International Organization for Standardization</i>
KM	Kaplan-Meier
MAP	máximo à posteriori
MCF	função média acumulada, do inglês <i>mean cumulative function</i>
MCMC	Monte Carlo via Cadeias de Markov
PC	probabilidade de cobertura
PLP	processo de lei de potência
PPH	processo de Poisson homogêneo
PPNH	processo de Poisson não-homogêneo
RI	Reparo Imperfeito
RM	Reparo Mínimo
ROCOF	taxa de ocorrência de falhas, do inglês <i>rate of occurrence of failures</i>
RP	Reparo Perfeito
TRP	processo de renovação de tendência, do inglês <i>trend renewal process</i>

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Objetivo	22
1.2	Motivação	23
1.2.1	<i>Dados de colheitadeiras de cana-de-açúcar</i>	23
1.2.2	<i>Dados de caminhões basculantes</i>	24
1.3	Organização dos capítulos	25
1.4	Produtos da tese	25
2	CONCEITOS BÁSICOS	27
2.1	Análise de confiabilidade	27
2.2	Processos de contagem	30
2.2.1	<i>Processo de renovação</i>	32
2.2.2	<i>Processo de Poisson</i>	33
2.2.3	<i>Processo lei de potência</i>	35
2.2.3.1	<i>Uma útil reparametrização para o processo de lei de potência</i>	37
2.3	Sistemas reparáveis	37
2.3.1	<i>Modelo de reparo mínimo</i>	39
2.3.2	<i>Modelo de reparo perfeito</i>	40
2.3.3	<i>Modelo de reparo imperfeito</i>	42
2.4	Modelo de fragilidade	44
2.4.1	<i>Distribuição de fragilidade</i>	46
2.5	Inferência Bayesiana objetiva	47
2.5.1	<i>Priori de Jeffreys</i>	48
2.5.2	<i>Priori de referência</i>	49
2.5.3	<i>Matching priors</i>	51
3	ANÁLISE BAYESIANA OBJETIVA PARA MÚLTIPLOS SISTEMAS REPARÁVEIS	53
3.1	Múltiplos sistemas reparáveis	53
3.1.1	<i>Função de verossimilhança do modelo de RM</i>	55
3.2	Estimação Bayesiana objetiva para o modelo	56
3.2.1	<i>Priori de Jeffreys</i>	57
3.2.2	<i>Priori de referência</i>	58

3.2.3	<i>Matching priors</i>	60
3.3	Estudo de simulação	60
3.4	Aplicação - Dados de colheitadeiras	66
3.5	Aplicação - Dados de caminhões basculantes	68
3.6	Considerações finais	70
4	MÚLTIPLOS SISTEMAS REPARÁVEIS COM UM TERMO DE FRAGILIDADE	73
4.1	Múltiplos sistemas reparáveis com fragilidade gama	73
4.1.1	<i>Função de sobrevivência e risco não-condicional usando transformação de Laplace</i>	75
4.2	Estudo de simulação	77
4.3	Aplicação - Dados de colheitadeiras	82
4.4	Aplicação - Dados de caminhões basculantes	84
4.5	Considerações finais	86
5	UMA CLASSE GERAL DE MODELOS DE FRAGILIDADE EM SISTEMAS REPARÁVEIS	87
5.1	Formulação do modelo	88
5.1.1	<i>Modelo de classe geral sem fragilidade</i>	90
5.2	Modelo de classe geral sob reparo mínimo e fragilidade gama	90
5.2.1	<i>Modelo de classe geral sob reparo mínimo sem fragilidade</i>	93
5.3	Estudo de simulação	95
5.4	Aplicação - Dados de colheitadeiras	101
5.5	Aplicação - Dados de caminhões basculantes	103
5.6	Considerações finais	105
6	MODELOS DE FRAGILIDADE NÃO-PARAMÉTRICA PARA MÚLTIPLOS SISTEMAS REPARÁVEIS	107
6.1	Múltiplos sistemas reparáveis com fragilidade não-paramétrica	108
6.2	Classe geral de modelos de fragilidade sob reparo mínimo e fragilidade não-paramétrica	110
6.3	Aplicação - Dados de colheitadeiras	114
6.4	Aplicação - Dados de caminhões basculantes	116
6.5	Considerações finais	118
7	CONCLUSÕES E PROPOSTAS FUTURAS	119
	REFERÊNCIAS	123
	ANEXO A ESTIMADOR NÃO-PARAMÉTRICO DE WANG E CHANG	131

INTRODUÇÃO

Tendo em vista a oportunidade de aumentar a produção e reduzir custos, com foco no cenário industrial atual, as falhas que aparecem nos processos e equipamentos devem ser analisadas, ações preventivas devem ser tomadas e o controle deve ser constante. A análise e prevenção de falhas se tornam fatores cruciais dentro de uma demanda mais produtiva que se concentra na redução de custos operacionais e no aumento da qualidade, além de ser essencial que uma ação de reparo ocorra o mais rápido possível após cada falha. Para estudar a análise das falhas e ações de reparo, surgiu uma área de confiabilidade denominada sistemas reparáveis.

Nesta área, o objeto de estudo é o sistema, que pode ser máquina, software ou equipamento eletrônico, e a definição de sistemas reparáveis estabelecida por [Ascher e Feingold \(1984\)](#) afirma que um sistema é considerado reparável se sua atividade puder ser satisfatoriamente retomada através de reparo após uma falha sem a necessidade de substituir todos os componentes do sistema, ou seja, dados de sistemas reparáveis são dados de eventos recorrentes. É muito importante notar a diferença entre sistemas reparáveis e não-reparáveis, porque é possível tirar conclusões errôneas se essa diferença não for percebida.

Modelos probabilísticos e estatísticos para analisar e otimizar o desempenho de sistemas reparáveis têm sido amplamente discutidos na literatura por vários autores. O primeiro livro dedicado exclusivamente à confiabilidade de sistemas reparáveis foi [Ascher e Feingold \(1984\)](#) e, por um longo tempo, esta foi a principal referência desta área. [Rigdon e Basu \(2000\)](#) descreveram detalhadamente a escolha de um modelo para sistemas reparáveis a partir da influência da degradação, das falhas e dos reparos.

Os modelos em sistemas reparáveis devem ser capazes de descrever a ocorrência de eventos (falhas) ao longo do tempo, bem como o efeito da manutenção corretiva (reparos). Assim, qualquer estudo destinado a determinar uma política adequada para sistemas reparáveis deve basear-se em tais modelos. Vários modelos surgiram com a ideia de modelar o efeito da ação de reparo realizada nos sistemas. O primeiro autor a desenvolver modelos estatísticos com base nos

processos de contagem de eventos recorrentes foi Aalen (1978). Tomazella (2003) descreveu a ampla literatura de modelos com base no uso de processos de contagem para dados de eventos recorrentes. Cox e Isham (1980) e Cox e Lewis (1996) são exemplos desses modelos em que covariáveis estão presentes.

Um ponto crítico a se considerar em modelos para sistemas reparáveis é como explicar o efeito das ações de reparo realizadas após falhas. Nesse sentido, os modelos mais conhecidos são Reparo Mínimo (RM), que retorna o sistema à mesma condição logo antes da falha, conhecido na literatura como “tão ruim quanto velho” (ABAO, do inglês *As Bad as Old*), Reparo Perfeito (RP), que deixa o sistema como se fosse novo, condição conhecida na literatura como “tão bom quanto novo” (AGAN, do inglês *As Good as New*) e Reparo Imperfeito (RI), que deixa o sistema em uma condição intermediária. O RM é a suposição mais analisada na literatura, que descreve situações em que a ação de reparo é apenas para o propósito do sistema retornar à sua funcionalidade; portanto, após o reparo, o sistema está na mesma condição imediatamente antes da falha. Phelps (1981), Barlow e Proschan (1987), Zhao e Xie (1996), Park, Jung e Yum (2000), Wang (2002), entre outros, discutiram essa suposição.

O RM pressupõe que o reparo mantém o sistema na mesma condição anterior e é razoável para sistemas constituídos por vários componentes, cada um com seu próprio modo de falha, de acordo com Kijima (1989). A falha seguida de um reparo ocorre várias vezes durante o período do estudo e, portanto, a ocorrência de falhas possui um processo de contagem associado que pode ser caracterizado por um processo de Poisson não-homogêneo, no qual a probabilidade de falha em pouco tempo depende apenas da idade do sistema, não do histórico de falhas (MURALIDHARAN, 2008).

As estimativas de parâmetros de modelos de sistemas reparáveis foram amplamente discutidas com base no estimador de máxima verossimilhança (EMV) e em suas propriedades assintóticas, como em Berman e Turner (1992), Zhao e Xie (1996) e Rigdon e Basu (2000). Existem algumas desvantagens no contexto frequentista, uma vez que os EMVs são tendenciosos para amostras pequenas e os intervalos de confiança se apoiam na teoria assintótica, retornando resultados não confiáveis. Para superar esse problema, a inferência Bayesiana pode ser utilizada. Vários trabalhos consideraram essa abordagem, como Guida, Calabria e Pulcini (1989), que usa várias diferentes escolhas de priores, tanto informativas quanto não-informativas; Sen (2002), Yu, Tian e Tang (2007) e de Oliveira, Colosimo e Gilardoni (2012), que usam distribuições à priori não-informativas; e como Kim, Park e Ahn (2008) e Huang (2001), que usam distribuição à priori conjugada. Na área de análise de sobrevivência também há estudos utilizando inferências Bayesianas, como Moala e Garcia (2013), que também usam distribuições à priori não-informativas. Por outro lado, os trabalhos existentes não exploram quais priores levam a estimativas imparciais, nem se elas têm bons desempenhos frequentistas. É importante ressaltar que, em muitos casos, a distribuição à priori pode diferir de acordo com a escolha da ordenação dos parâmetros, por outro lado, estamos interessados em estimar todos os parâmetros simultaneamente.

Muitos trabalhos têm aparecido na literatura com diferentes modelagens. Alguns exemplos em que covariáveis não estão presentes nesses modelos são [Cox e Isham \(1980\)](#) e [Cox e Lewis \(1996\)](#). [Hussain e Naikan \(2010\)](#) apresentam alguns modelos de escala de tempo alternativos, que não usam o tempo de calendário ou o tempo global, por exemplo, a quilometragem percorrida por um automóvel pode ser usada como uma escada de tempo. [Lindqvist \(1993\)](#) propôs um novo modelo denominado processo de renovação de tendência (TRP, do inglês *trend renewal process*). A inferência paramétrica clássica do modelo TRP é estudada por [Lindqvist, Elvebakk e Heggland \(2003\)](#). [Krivtsov \(2007\)](#) estendeu o processo de Poisson não-homogêneo ao processo de renovação não-homogêneo. [Pulcini \(2001\)](#) também propôs um modelo para analisar sistemas reparáveis mais complexos, que pode exibir o comportamento da curva de banheira da função de intensidade. [Leung e Fong \(2000\)](#) propôs um modelo chamado de Processo Geométrico, para representar a idade que causa a diminuição dos tempos de operação devido ao aumento do tempo de reparo.

De acordo com [Lehmann \(2002\)](#), sempre que um sistema é submetido a um reparo após cada falha, o nível de degradação do sistema é repassado para um ponto que se situa entre os níveis do novo sistema e do sistema imediatamente antes da falha. A este respeito, um modelo de degradação do sistema é desenvolvido com base em um processo pontual. Usando o processo Weiner, [Lindqvist e Skogsrud \(2008\)](#) propuseram a modelagem de um processo de degradação para os sistemas.

Na análise de dados de sistemas reparáveis, é comum que muitos componentes do mesmo tipo sejam estudados e, nesses casos, é relevante verificar a heterogeneidade entre os sistemas. [Proschan \(1963\)](#) apontou que a heterogeneidade não-observada pode explicar taxas de falha crescentes, que muitas vezes são encontradas na análise de confiabilidade. A heterogeneidade não-observada pode ser estimada a partir de modelos de fragilidade. Este tipo de modelo é caracterizado pelo uso de um efeito aleatório, ou seja, uma variável aleatória não-observável que representa a informação que não pôde ou não foi observada. Segundo [Vaupel, Manton e Stallard \(1979\)](#), os métodos padrões na análise de dados de sistemas reparáveis ignoram a heterogeneidade não-observada. A fim de modelar esta estrutura de dependência, surge o modelo de fragilidade multiplicativo amplamente utilizado na literatura; podemos citar [Wienke \(2010\)](#) e [Hougaard \(2012\)](#) como boas referências nesta área. [D'Andrea et al. \(2017\)](#) e [Feitosa \(2015\)](#) consideraram a fragilidade Gama para o modelo de reparo mínimo.

[Bain e Wright \(1982\)](#) aplicou o modelo com processo binomial negativo para lidar com múltiplos sistemas tendo intensidades diferentes, embora sejam de natureza similar, em que essas diferentes intensidades refletem a heterogeneidade entre os sistemas. Alguma heterogeneidade não-observada entre os múltiplos sistemas foi tratada por [Lindqvist, Elvebakk e Heggland \(2003\)](#) no modelo de processo de renovação com tendência heterogêneo, onde a heterogeneidade não-observada é introduzida como um fator multiplicativo, que é uma variável aleatória dependente na intensidade.

Combinando os efeitos de manutenção preventiva e as variáveis de predição, um modelo denominado modelo de intensidade proporcional generalizada foi proposto por [Percy e Alkali \(2006\)](#). Como os tempos de falha de sistemas dependem de muitos fatores, com base no modelo de regressão proposto por [Cox \(1972a\)](#), vários trabalhos propuseram modelagem com covariáveis. Para os sistemas com reparos gerais, com base no histórico de reparos prévios, um modelo de intensidade proporcional com intensidade de base log-linear foi proposto por [Guo et al. \(2007\)](#), que considera simultaneamente as tendências do tempo e os efeitos de reparo.

É intuitivo pensar que uma vez reparado o equipamento, suas chances de apresentar uma nova falha são maiores do que antes de apresentar o primeiro problema, especialmente dependendo do tipo de reparo que foi realizado. Assim, o número de reparos realizados pode influenciar o próximo reparo e, então, o estudo deste efeito é de grande interesse para que o planejamento da manutenção do equipamento seja preventivo ou corretivo. [Peña e Hollander \(2004\)](#), no contexto da análise de confiabilidade e sobrevivência, propôs uma nova classe de modelos a dados recorrentes que generaliza a maioria dos modelos de confiabilidade e de Cox existentes. É uma classe de modelos geral e flexível que incorpora simultaneamente os efeitos das covariáveis, o impacto na unidade de ocorrências de eventos acumulados, o efeito de variáveis latentes ou não-observadas, bem como o efeito de intervenções realizadas após cada ocorrência do evento.

1.1 Objetivo

Essa tese tem por objetivo estudar diversos aspectos da modelagem para múltiplos sistemas reparáveis, supondo que o RM é adequado aos dados e considerar a forma paramétrica do processo de lei de potência (PLP) para a função de intensidade do sistema. Assim, podemos definir os objetivos específicos:

- considerar uma parametrização ortogonal para os parâmetros do modelo PLP e derivar prioris objetivas na abordagem Bayesiana objetiva. Sob essa abordagem Bayesiana objetiva, as informações dominantes na distribuição à posteriori são fornecidas pelos dados;
- considerar a heterogeneidade não-observada entre os sistemas através da incorporação de um termo de fragilidade multiplicativo, que será abordado de forma paramétrica e não-paramétrica;
- propor uma classe geral e flexível de modelos de fragilidade para descrever e analisar dados de sistemas reparáveis, no sentido de que incorpora simultaneamente a heterogeneidade não-observada, o efeito do acúmulo de recorrências dos reparos no sistema e covariáveis observadas.

O desenvolvimento dessa modelagem visa subsidiar a elaboração de uma futura política ótima de manutenção de sistemas reparáveis, aportando novos resultados e complementando alguns resultados da literatura.

1.2 Motivação

O trabalho teve como principal motivação dois problemas práticos. O primeiro foi retirado de uma usina brasileira produtora de açúcar e álcool. Um segundo conjunto de dados sobre caminhões basculantes, já analisado na literatura, também foi considerado.

1.2.1 Dados de colheitadeiras de cana-de-açúcar

Denominaremos por “Dados de colheitadeiras” o conjunto de dados relacionado ao tempo de quebra de uma peça localizada no sistema de corte da cana-de-açúcar e o evento de interesse é a quebra da lâmina *Chopper* de colheitadeiras de cana-de-açúcar de usinas produtoras de açúcar e álcool nos estados de São Paulo e Paraná durante a safra de 2014/2015. Tal conjunto de dados é parte dos dados coletados por Verssani em sua dissertação (VERSSANI, 2018). A lâmina *Chopper* é responsável por cortar a cana em pedaços pequenos, que são posteriormente levados até a usina. Sempre que há uma quebra nessa lâmina, a colheitadeira vai para o conserto e a colheita de cana é prejudicada, causando prejuízos para a usina. Portanto, as falhas desse sistema devem ser evitadas, a fim de minimizar os prejuízos para a empresa.

Este conjunto de dados descreve os tempos de falha, em dias, de três colheitadeiras de cana-de-açúcar, em que cada colheitadeira de cana tem um número diferente de reparos, variando de 11 a 14 falhas por colheitadeira, com um total de 38 falhas na lâmina *Chopper*, cada uma seguida de um reparo. Cada um desses reparos é considerado mínimo, ou seja, consertou-se somente o necessário para que a colheitadeira voltasse rapidamente ao serviço. A observação dos dados terminou em 195 dias. O objetivo para esses dados é estudar a confiabilidade das colheitadeiras de cana-de-açúcar.

O conjunto de dados é apresentado na Tabela 1 e sua representação gráfica é apresentada na Figura 1, em que cada linha representa uma colheitadeira e cada ponto representa um tempo de falha.

Tabela 1 – Dados do tempo de quebra da lâmina *Chopper* localizada no sistema de corte de colheitadeiras de cana-de-açúcar.

colheitadeira	Tempos de falha (em dias)													
1	10	42	51	68	85	110	120	146	157	167	194			
2	8	9	25	31	40	62	73	88	107	118	124	154	158	178
3	1	29	31	38	60	82	83	101	102	128	129	153	182	

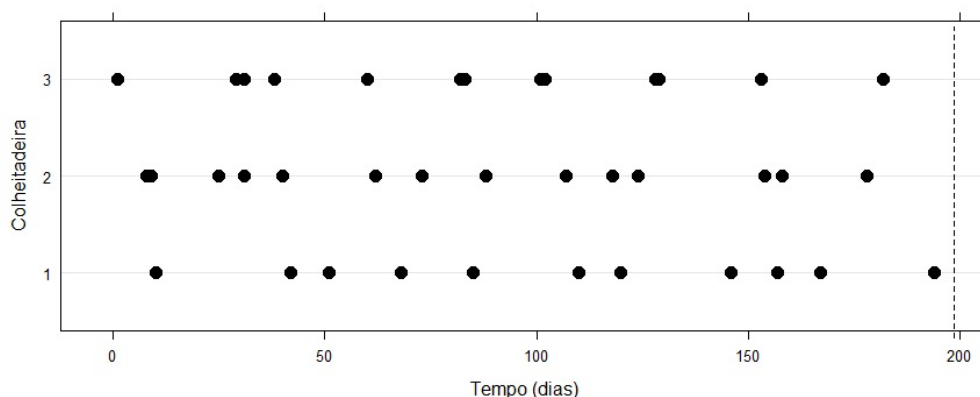


Figura 1 – Tempos de falha (em dias) para cada colheitadeira de cana-de-açúcar.

1.2.2 Dados de caminhões basculantes

Este conjunto de dados refere-se a uma situação real em relação a falhas de caminhões basculantes usados por uma empresa de mineração brasileira, disponibilizados em Toledo (2014). Os caminhões basculantes que auxiliam no processo de mineração, não foram projetados para operar em condições extremas como as encontradas em locais de mineração. Quando esses caminhões apresentam falhas, seu reparo geralmente é imediato, pois seu objetivo é prestar assistência a caminhões grandes projetados para o transporte de minério. A ocorrência de uma falha nestes caminhões é muito custosa e prejudicial à mineradora por problemas como segurança, atrasos na produção, entres outros. Por isso, estudar a confiabilidade desses caminhões é fundamental para reduzir custos.

Assim, consideramos os tempos de falha de cinco caminhões, em que foram observadas 129 falhas no período de observação que foi de julho a outubro de 2012. Cada falha foi seguida de um reparo. Para este trabalho, consideramos que o estudo foi truncado em 107 dias. Os tempos de falhas foram computados em dias, no qual foram registradas 129 falhas no total. Um esboço dos dados segue na Tabela 2.

Tabela 2 – Esboço dos tempos de falha dos caminhões basculantes

Unidade	Tempos de falha (em dias)									
1	9	15	25	...	94	98	106			
2	3	5	11	...	66	69	72	...	100	103
⋮				⋮						
5	3	9	26	...	91	93	99			

Fonte: Toledo (2014).

A Figura 2 apresenta uma representação gráfica para o conjunto de dados, em que cada linha corresponde a um caminhão e cada ponto representa um tempo de falha. O objetivo para

esses dados também é estudar a confiabilidade dos caminhões basculantes.

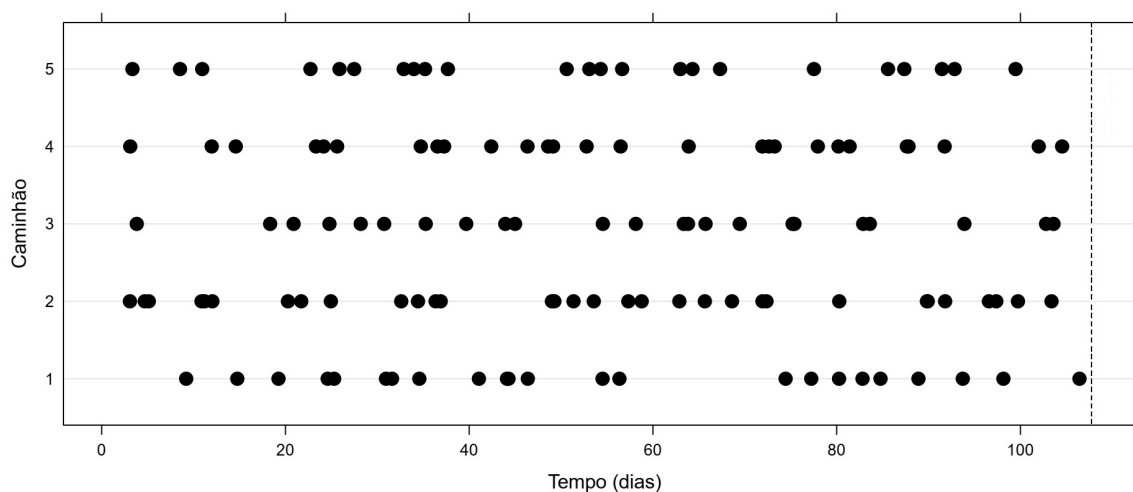


Figura 2 – Tempos de falha (em dias) para cada caminhão basculante.

1.3 Organização dos capítulos

Este trabalho está organizado da seguinte forma. Apresentaremos no Capítulo 2 todos os conceitos básicos necessários para a sequência do trabalho. No Capítulo 3 será considerada a abordagem Bayesiana objetiva para modelo de RM. No Capítulo 4 apresentaremos a inserção de um efeito aleatório chamado fragilidade na função de intensidade dos sistemas reparáveis, supondo a distribuição Gama, em que realizaremos um estudo de simulação, bem como uma aplicação para dois conjuntos de dados reais. Propomos no Capítulo 5 uma classe geral de modelos de fragilidade para sistemas reparáveis, que usa o termo de fragilidade e o acúmulo das ocorrências de reparos na modelagem. Estudo de simulação e aplicação para dados reais serão realizados. No Capítulo 6 apresentaremos uma abordagem não-paramétrica para a variável de fragilidade, tanto no modelo de fragilidade para sistemas reparáveis quanto para a classe geral de modelos de fragilidade, em que estimaremos as fragilidades individuais dos sistemas nas aplicações. O trabalho será concluído no Capítulo 7 com as considerações finais e propostas futuras.

1.4 Produtos da tese

Alguns capítulos da tese estão sendo preparados para submissão de artigos, sendo que um deles já está submetido e sob análise:

- D'Andrea, A. M. E., Tomazella, V. L. D. T., Ramos, P. L., Almeida, M. P., Louzada, F. e Verssani, B. A. W.. *Objective Bayesian Analysis for Multiple Repairable Systems*.

Este artigo está apresentado no Capítulo 3 da tese, foi submetido para publicação, revisado uma vez e está atualmente sob análise (Quality Engineering).

CONCEITOS BÁSICOS

Neste capítulo apresentamos as metodologias utilizadas neste trabalho a fim de dar suporte para a construção da classe geral de modelos de fragilidade em sistemas reparáveis. Para tal, introduzimos a análise de confiabilidade, mostramos os processos de contagem, descrevendo o processo de Poisson e o de renovação, importantes processos em sistemas reparáveis. Em seguida, descrevemos sistemas reparáveis, em que apresentamos suas principais características e modelos, introduzimos a modelagem de fragilidade e mostramos metodologias para inferência Bayesiana.

2.1 Análise de confiabilidade

No mundo atual, com a tecnologia avançando em larga escala, preocupa-se cada vez mais com a qualidade de sistemas. Nessa tese, adotaremos o termo “sistema” em um sentido amplo, podendo significar desde um único componente até sistemas complexos. Caso estes sistemas falhem, podem causar diversos tipos de danos, como descontentamento de um cliente até graves danos ao meio ambiente e à população. Neste contexto, tem-se desejado estudar cada vez mais a confiabilidade destes sistemas, mas como medir essa confiabilidade? Ou seja, como podemos dizer o quão confiável é um sistema?

Segundo [Hamada et al. \(2008\)](#), a Organização Internacional de Padronização (ISO, do inglês *International Organization for Standardization*) define confiabilidade como “a capacidade de um item desempenhar uma função necessária, sob determinadas condições ambientais e operacionais e por um período de tempo determinado”. Essa definição de confiabilidade nos permite perceber que uma análise deve não só levar em conta se um sistema teve sucesso ou fracasso, mas também o período em que o sistema funcionou e, assim, diferentes modelos estatísticos podem ser pensados para realizar a análise de confiabilidade. Com tais modelos, podemos estimar a probabilidade de um sistema falhar após um tempo específico. Por exemplo, em um estudo da confiabilidade de uma lâmpada, podemos estar interessados na probabilidade

de uma lâmpada queimar após um tempo específico, como 1000 horas.

A confiabilidade pode ser descrita através de uma função, mas antes de defini-la, precisamos estabelecer alguns conceitos. Definimos uma variável aleatória contínua T como sendo o tempo até a falha de um sistema em um experimento, em que T assume valores no conjunto dos números reais positivos e o espaço amostral desse experimento é qualquer tempo positivo.

Nos estudos de confiabilidade, um dos principais focos é na modelagem da distribuição do tempo de falha de um sistema e, a partir dela, podemos definir algumas funções de interesse, entre elas, a função de confiabilidade. Primeiro, definimos a função de densidade de probabilidade (fdp), que satisfaz

$$f(t) \geq 0, \quad (2.1)$$

em que $t \geq 0$, e

$$\int_{-\infty}^{\infty} f(t) dt = 1. \quad (2.2)$$

A partir da fdp, podemos definir a função de distribuição acumulada (fda), que indica a probabilidade da variável aleatória assumir valores menores ou iguais a t , isto é,

$$F(t) = P(T \leq t) = \int_{-\infty}^t f(u) du. \quad (2.3)$$

Podemos então definir a função de confiabilidade, também chamada de função de sobrevivência em estudos de análise de sobrevivência. A função de confiabilidade indica a probabilidade do tempo de falha ser superior a um tempo t , ou seja,

$$R(t) = P(T > t) = 1 - F(t), \quad (2.4)$$

em que podemos perceber que esta função assume valores no intervalo $[0, 1]$.

Uma última função de interesse em estudos de confiabilidade é a função de risco. Ela nos fornece a taxa de falha instantânea, ou seja, a tendência que um sistema falhe em um pequeno intervalo de tempo, sabendo que ele sobreviveu até o tempo t , e é definida como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{R(t)}. \quad (2.5)$$

Estudos tradicionais de análise de confiabilidade têm a característica de se observar o tempo até uma única falha de um ou mais sistemas, ou seja, o evento de interesse é terminal por poder acontecer apenas uma vez; esses sistemas são chamados de sistemas não-reparáveis. Neste tipo de estudo, pode haver a presença de censura, ou seja, quando o tempo exato de falha de algum sistema é desconhecido. Na literatura são descritos diversos mecanismos de censura (ver, por exemplo, [Colosimo e Giolo \(2006\)](#) e [Lawless \(1982\)](#)).

Existem estudos em sobrevivência e confiabilidade em que o evento de interesse não é terminal, podendo acontecer mais do que uma vez durante o período de observação em um mesmo sistema. Existem diversas aplicações desse tipo de estudo, como recorrências de uma doença, ex-presidiário cometer crimes novamente, um casal ter filhos, entre outras. Podemos notar que nos exemplos citados podem ter múltiplas ocorrências do mesmo evento de interesse em um mesmo indivíduo/sistema ao longo do tempo. Para esse tipo de estudo, técnicas tradicionais de análise de sobrevivência e confiabilidade não são adequadas. A área da análise de sobrevivência e confiabilidade que estuda esse tipo de situação é chamada de análise de eventos recorrentes ou análise de eventos históricos. Em estudos de eventos recorrentes, o interesse principal é estudar o processo que gera os diversos eventos de interesse ao longo do tempo.

Tomazella (2003) e Cook e Lawless (2007) descrevem alguns pontos importantes sobre eventos recorrentes, como caracterizar a dinâmica de ocorrência de eventos individual para cada processo; descrever a variação entre um conjunto de processos; comparar grupos de processos e conhecer a relação entre a ocorrência do evento com covariáveis observadas.

Voltando ao contexto de confiabilidade, o evento de interesse é a falha de um sistema. Existem sistemas que, após apresentar a falha, podem ser recuperados através de um reparo e voltam a ser estudados, podendo falhar novamente diversas vezes. Nesse tipo de estudo, em cada sistema podem ser observadas diversas falhas até o final da observação e, assim, trata-se de eventos recorrentes. Mas, nesse contexto, dá-se o nome de análise de Sistemas Reparáveis. Análise de Sistemas Reparáveis é, portanto, uma área da confiabilidade em que um dos principais interesses é estudar o processo que gera as diversas falhas nos sistemas, em que esses sistemas são ditos reparáveis.

Um conceito muito importante em sistemas reparáveis é o truncamento. É comum que o truncamento seja confundido com a censura. Em sistemas não-reparáveis, o conceito de censura é devido à falta de informação sobre a variável de interesse. Censura ocorre quando você tem alguma informação sobre a variável aleatória de interesse, mas esta informação está incompleta. Existem basicamente três tipos de censura: à direita, à esquerda e intervalar; e três mecanismos de censura: censura tipo I, censura tipo II e censura aleatória.

Truncamento é um conceito utilizado na área de confiabilidade e existem basicamente duas formas de observar os dados. Na primeira, o conjunto de dados termina em um limite de tempo pré-estabelecido T , chamado por truncamento por tempo e, na segunda, o conjunto de dados termina após um número de falhas pré-determinado n , denotado por truncamento por falha.

Seja $t_{i,j}$, para $i = (1, 2, \dots, k)$ e $j = (1, 2, \dots, n_i)$ observações da variável aleatória T que representa os tempos de falha para o i -ésimo sistema, registrado como o tempo desde o início do sistema ($0 < t_{i,1} < t_{i,2} < \dots < t_{i,n_i}$). Se o i -ésimo sistema for truncado por tempo, ele é observado até o pré-determinado tempo T_i , em que $0 < t_{i,1} < t_{i,2} < \dots < t_{i,n_i} < T_i$ e T_i é fixado, n_i é aleatório e $t_{i,n_i} < T_i$; ou se o i -ésimo sistema é truncado por falha, o número de falhas n_i é pré-determinado,

em que $(0 < t_{i,1} < t_{i,2} < \dots < t_{i,n_i})$ e, assim, n_i é fixo e $T_i = t_{i,n_i}$ é aleatório.

Portanto, no caso em que as amostras apresentaram o fenômeno da censura, todos os casos são registrados, mesmo aqueles que são considerados como censurados. Por outro lado, nas amostras truncadas, os dados registrados são apenas aqueles que ocorrem dentro do limite de truncamento e os outros são omitidos ou descartados.

Vimos que em sistemas reparáveis um ponto importante é conhecer o processo gerador das diversas falhas de um sistema. Apresentamos então, na Seção 2.2, alguns conceitos importantes sobre processos de contagem e, em seguida, na Seção 2.3, definimos os conceitos principais em sistemas reparáveis.

2.2 Processos de contagem

Aalen (1978) inicialmente desenvolveu modelos estatísticos baseados em processos de contagem para analisar dados de eventos recorrentes. Recentemente, muitos modelos têm sido desenvolvidos utilizando a teoria de processos de contagem.

Suponha que T seja a variável aleatória que representa o tempo de falha e que o experimento começa em $t = 0$. Os eventos ocorrem aleatoriamente, onde T_1 ocorre primeiro, T_2 ocorre posteriormente e assim por diante; então, T_j indica o tempo que o j -ésimo evento ocorre. Se $N(t)$ representa o número total de eventos que ocorrem no intervalo $(0, t)$, este é um processo de contagem, ou seja, $N(t)$ é um processo estocástico que assume os valores $0, 1, 2, \dots$, registrando um salto da observação $j - 1$ à observação j , quando o evento de interesse ocorre pela j -ésima vez, $j = 0, 1, 2, \dots$. Até a primeira ocorrência do evento, $N(t)$ é igual a zero, quando então, salta para o valor 1. O processo continua até uma nova ocorrência de evento, e assim, vão se registrando saltos sucessivos de tamanho 1 até o final do experimento. Uma representação dos tempos de falha e do processo de contagem $N(t)$ encontra-se na Figura 3.

Um processo de contagem pode ser entendido como um conjunto de variáveis aleatórias $\{N(t), t \geq 0\}$ e deve satisfazer as seguintes propriedades:

- $N(t) \geq 0$ e $N(0) = 0$;
- $N(t)$ é um inteiro;
- $t \mapsto N(t)$ é contínuo à direita;
- $\Delta N(t) = \lim_{h \rightarrow 0} N(t) - N(t - h) = 0$.

O caráter temporal de um processo de contagem requer o conhecimento dos sistemas em cada instante $t \in [0, \infty)$. Esse conhecimento acerca do que aconteceu aos sistemas até o instante t , é chamado de história ou filtragem do processo de contagem no instante t , e é representado por

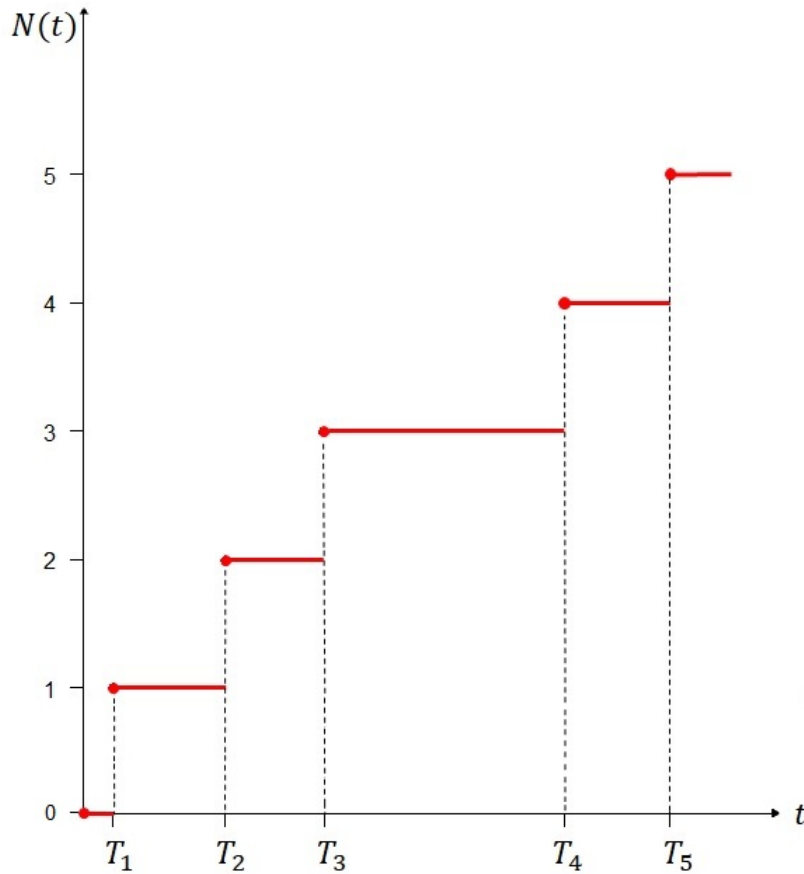


Figura 3 – Representação gráfica da variável T e do processo de contagem $N(t)$.

\mathcal{F}_t . A filtragem do processo estocástico até um instante imediatamente anterior a t , representa os dados referentes aos conjuntos de tempos de falhas até um tempo $t^- < t$, denotada por \mathcal{F}_{t^-} .

Assim, podemos definir a função de intensidade de falha. Assumindo que as falhas nos sistemas reparáveis são equivalentemente definidas pelo processo $\{N(t)\}_{t \geq 0}$ ou $\{T_{N(t)}\}_{t \geq 0}$, em que $N(t)$ indica o número de falhas até o tempo t e $T_{N(t)}$ corresponde ao tempo decorrido até a $N(t)$ -ésima falha, a função de intensidade de falha é dada por

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) - N(t) = 1 | \mathcal{F}_{t^-})}{\Delta t}, \quad (2.6)$$

para todo $t > 0$, em que \mathcal{F}_{t^-} é o histórico de todos os tempos de falha ocorrendo antes de t , com função média, que é a função de intensidade acumulada de (2.6), determinada por

$$\Lambda(t) = \mathbb{E}[N(t)] = \int_0^t \lambda(u) du. \quad (2.7)$$

A taxa de ocorrência de falhas (ROCOF, do inglês *rate of occurrence of failures*) é a intensidade de falha do sistema antes da primeira ação de reparo, dada por

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[N(t + \Delta t) - N(t)]}{\Delta t}, \quad (2.8)$$

em que percebemos que a ROCOF não é condicional à história \mathcal{F}_{t-} , enquanto que a função de intensidade de falha depende do histórico de falhas.

Existem dois processos de contagem amplamente utilizados em sistemas reparáveis, o processo de renovação e o processo de Poisson. Descreveremos ambos a seguir.

2.2.1 Processo de renovação

O processo de renovação depende do histórico através de $T_{N(t-)}$ que, por sua vez, representa o tempo até a ocorrência do evento que aconteceu anterior a t . Este processo é dito renovável porque a probabilidade de falha começa com cada processo de reparo, ou seja, o tempo até a próxima falha tem a mesma distribuição se o sistema é novo ou se ele foi reparado muitas vezes.

A caracterização mais simples é através dos intervalos entre tempos de ocorrência de falha (*gap times*), sendo independente e identicamente distribuídos (iid). Assim, se todas as variáveis aleatórias Y_j são iid, então $\{Y_n; n \geq 1\}$ é um processo de renovação, em que $Y_j = T_j - T_{j-1}$ e $T_0 = 0$ indicam o tempo entre os eventos $j - 1$ e j . Desta forma, imediatamente após a ocorrência de um evento o risco acumulado assume valor zero e cresce ou se estabiliza ao longo do tempo até a ocorrência de um novo evento, pois se pressupõe que o risco de ocorrência do $(j + 1)$ -ésimo evento somente se inicia após observação do j -ésimo. Desta forma, é possível dizer que o processo se renova após o acontecimento de um evento, pois o risco acumulado de ocorrência de um novo evento cai para zero após observação de um evento. Ao contrário do processo de Poisson, que imediatamente após a ocorrência de um evento mantém a intensidade acumulada $\lambda(t)$, de forma crescente, isto é, o número esperado de eventos sempre será maior após ocorrência de um evento, pois assume-se que todos os eventos começam a se desenvolver simultaneamente, de forma independente, no mesmo instante de tempo e a ordenação apenas reflete a ordem de seus acontecimentos no tempo. Isso justifica o fato de alguns autores mencionarem esse processo como o de análise de tempo total (ou tempo global) em estudo.

A função de intensidade que caracteriza o processo de renovação é dada por

$$\lambda(t|\mathcal{F}_{t-}) = \lambda[t - T_{N(t-)}], \quad (2.9)$$

sendo \mathcal{F}_{t-} o histórico do processo e $N(t-)$ o número de eventos que ocorreram antes de t , pois $t-$ representa um tempo infinitesimalmente menor que t .

Assim denotamos a fdp das variáveis aleatórias Y_j por $f(y)$ e sua fda por $F(y)$. Considerando um estudo truncado em um tempo τ , temos que $0 < T_1 < T_2 < \dots < T_n < \tau$ e $\tau - T_n$ é o tempo entre a última falha e o final do estudo. Os equivalentes tempos entre falhas Y_1, Y_2, \dots, Y_n são condicionalmente independentes e todos com fdp $f(y)$, então a função de verossimilhança de um sistema, de acordo com Hamada *et al.* (2008), é da forma

$$L(\theta|\mathbf{y}) = \left(\prod_{j=1}^n f(y_j) \right) [1 - F(\tau - t_n)], \quad (2.10)$$

em que y_1, y_2, \dots, y_n são tempos entre falhas observados, t_n o último tempo de falha observado e θ é o parâmetro (ou o vetor de parâmetros) relacionado a $f(y)$.

O principal motivo para o estudo dos processos de renovação é que em muitos processos complicados existem certos instantes em que o sistema retorna a um estado probabilisticamente equivalente ao estado inicial de forma aleatória. Esses tempos de renovação internos permitem separar o comportamento a longo prazo do processo de ocorrência de falhas, do comportamento do processo atual dentro de um período de renovação, que pode ser estudado através da teoria da renovação.

2.2.2 Processo de Poisson

Embora o processo de renovação seja uma classe muito importante de modelos, ele não acomoda situações em que o crescimento ou a diminuição da confiabilidade pode acontecer. Assim, trataremos aqui do processo de Poisson, que acomoda situações desse tipo.

O processo de Poisson é um modelo de processos de contagem amplamente utilizado na literatura. Um processo $N(t)$ é dito ser um processo de Poisson se:

- $N(0) = 0$;
- Para quaisquer $a < b \leq c < d$, as variáveis $N(a, b]$ e $N(c, d]$ são independentes, isto é, contagens em intervalos sem sobreposição são independentes;
- A função de intensidade $\lambda(t)$ é definida como

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(N(t, t + \Delta t] = 1)}{\Delta t}; \quad (2.11)$$

- Não há falhas simultâneas, ou seja,

$$\lim_{\Delta t \rightarrow 0} \frac{P(N(t, t + \Delta t] \geq 2)}{\Delta t} = 0. \quad (2.12)$$

A partir dessas condições, podemos definir que

$$P(N(t) = x) = \frac{\Lambda(t)^x \exp[-\Lambda(t)]}{x!}, \quad (2.13)$$

em que $\Lambda(t) = \int_0^t \lambda(u) du$.

Assim, através de 2.13, temos que $N(a, b]$ tem distribuição Poisson com parâmetro $\Lambda(t)$ e então

$$\mathbb{E}(N(a, b]) = \text{Var}(N(a, b]) = \int_a^b \lambda(u) du = \Lambda(b) - \Lambda(a). \quad (2.14)$$

Se o intervalo de observação do processo de Poisson for $(0, \tau]$, então $\mathbb{E}(N(0, \tau]) = \Lambda(\tau)$. Em situações em que $\Lambda(t)$ é diferenciável, podemos definir a ROCOF como $\frac{d}{dt} \Lambda(t)$.

Uma importante característica do processo de Poisson é que

$$\Lambda(T_j) - \Lambda(T_{j-1}) \sim \text{Exponencial}(1). \quad (2.15)$$

Assim, para $j = 1, \dots, n$ e $T_0 = 0$, $\Lambda(T_j) - \Lambda(T_{j-1})$ são variáveis aleatórias iid Exponencial(1) e então podemos expressar T_j da seguinte forma:

$$T_j \sim \lambda^{-1}[\Lambda(T_{j-1}) + \text{Exponencial}(1)], \quad (2.16)$$

em que $\lambda^{-1}(\cdot)$ indica função inversa de $\lambda(\cdot)$, e assim temos uma forma de simular tempos de falha de um processo de Poisson.

O processo de Poisson pode ser de dois tipos: o processo de Poisson homogêneo (PPH) e o processo de Poisson não-homogêneo (PPNH).

O PPNH é um processo de Poisson cuja função de intensidade $\lambda(t)$ não é constante, assim, ele segue as propriedades já descritas para processos de Poisson. Para o PPNH, $N(a, b]$ tem uma distribuição de Poisson com média e variância

$$\mathbb{E}(N(a, b]) = \text{Var}(N(a, b]) = \Lambda(b) - \Lambda(a). \quad (2.17)$$

O PPNH é um processo de contagem muito utilizado em sistemas reparáveis. Uma das principais importâncias do processo não-homogêneo de Poisson reside no fato de que ele não requer a condição de incrementos estacionários. Além disso, a propriedade de incrementos independentes do processo de Poisson exige que $\lambda(t)$ não dependa do histórico do processo até t . A função de verossimilhança do PPNH deve combinar a função de densidade de probabilidade conjunta dos tempos de falha de cada sistema. Sua construção será descrita na Seção 3.1.1.

O PPH é o processo de Poisson que apresenta função de intensidade constante, isto é, $\lambda(t) = \lambda$, e assim $\mathbb{E}(N(a, b]) = \text{Var}(N(a, b]) = \lambda(b - a)$, ou seja, o número médio de falhas em um intervalo é uma função linear de λ .

Além disso, os intervalos de tempo entre falhas $\{Y_n; n \geq 1\}$ de um PPH com intensidade λ são variáveis aleatórias iid com distribuição Exponencial(λ). Finalmente, o número de eventos que ocorrem em um intervalo de tempo de comprimento t é uma variável aleatória discreta de Poisson com taxa λ .

Considerando um estudo truncado em um tempo τ , temos que $0 < T_1 < T_2 < \dots < T_n < \tau$. Como $\{Y_n; n \geq 1\}$ são variáveis aleatórias iid com distribuição Exponencial(λ), então a função de verossimilhança dos tempos observados $0 < t_1 < t_2 < \dots < t_n < \tau$ de um sistema, segundo Hamada *et al.* (2008), é da forma

$$L(\lambda | \mathbf{t}) = \left(\prod_{j=1}^n \lambda \exp[-\lambda(t_j - t_{j-1})] \right) \exp[-\lambda(\tau - t_n)] = \lambda^n \exp(-\lambda \tau), \quad (2.18)$$

em que t_n é o último tempo de falha observado.

O PPH ocorre quando os incrementos são estacionários, de modo que a distribuição do número de eventos em um intervalo depende apenas do comprimento do intervalo, ou seja, a taxa de intensidade do processo sempre será λ para todos os tempos.

Portanto, o PPH deve ser aplicado com cautela, pois, uma vez que a função de intensidade é constante, ele não pode ser usado para modelar sistemas que estão se deteriorando ou em processo de melhoria.

O processo de Poisson pode ser expresso através de diversas formas paramétricas, uma das mais conhecidas é a lei de potência, que será descrita na Seção 2.2.3.

2.2.3 Processo lei de potência

O modelo paramétrico mais utilizado na literatura para o PPNH é o Processo de Lei de Potência (PLP). Quando nos referimos a um PLP, significa que temos um PPNH com função de intensidade lei de potência. A origem do PLP está ligada ao trabalho de [Duane \(1964\)](#) e [Crow \(1974\)](#) introduziu formalmente e investigou tal modelo. A simples formulação matemática e a riqueza em formas são provavelmente as principais razões para o vasto uso da intensidade lei de potência na análise de confiabilidade (veja, por exemplo, [Kumar e Klefsjö \(1992\)](#) e [Coetzee \(1997\)](#)).

Seja $N(t)$ o número de falhas desde o início do acompanhamento do sistema até um certo tempo t , então, se $N(t)$ segue um PLP, a função de intensidade e a função de intensidade acumulada, propostas por [Crow \(1974\)](#), são dadas, respectivamente, por

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \quad (2.19)$$

e

$$\Lambda(t) = \left(\frac{t}{\eta} \right)^{\beta}, \quad (2.20)$$

em que $\eta > 0$ é o parâmetro de escala e $\beta > 0$ é o parâmetro de forma. Segundo [Oliveira, Colosimo e Gilardoni \(2014\)](#), o parâmetro η pode ser interpretado como o tempo durante o qual se espera exatamente que uma falha ocorra, ou seja, $\mathbb{E}[N(\eta)] = 1$, enquanto que β representa a elasticidade do número esperado de falhas em relação ao tempo, então, quando $\beta > 1$, o sistema está deteriorando e quando $\beta < 1$, o sistema está melhorando. A Figura 4 apresenta uma representação gráfica da função de intensidade do PLP, em que fixamos o parâmetro de escala ($\eta = 1$) e variamos o parâmetro de forma β . Quando $\beta = 1$, a função de intensidade é constante, ou seja, o sistema mantém a mesma intensidade de falha ao longo do tempo; quando $\beta > 1$, a função de intensidade é crescente, indicando que o sistema está se deteriorando e, quando $\beta < 1$, a função de intensidade é decrescente, indicando que o sistema está melhorando ao longo do tempo, ou seja, a intensidade com que ocorrem as falhas diminui com o tempo.

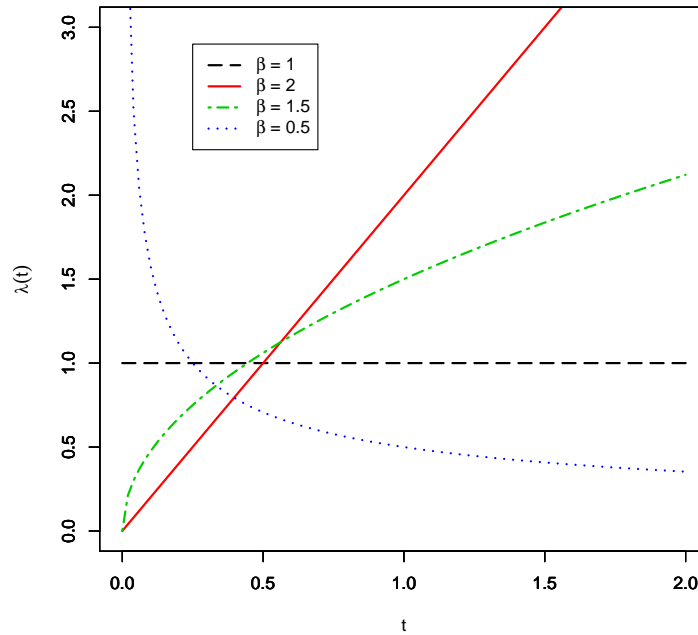


Figura 4 – Função de intensidade do PLP com $\eta = 1$ e diferentes valores de β .

Por muito tempo, o PLP foi referido como o processo Weibull, pois o tempo até a primeira falha do processo tem distribuição Weibull. Porém, de acordo com [Hamada et al. \(2008\)](#), o nome processo lei de potência é mais adequado, já que o leitor não faz confusão com o processo de renovação Weibull (tempos entre falhas iid seguindo a distribuição Weibull). Vários autores observam que essa nomenclatura pode gerar uma certa confusão, pois os tempos de falha não têm uma distribuição Weibull (exceto a primeira falha), nem os tempos entre falhas. A função de distribuição acumulada deste modelo é dada por

$$F(t) = \frac{\Lambda(t)}{\Lambda(\tau)} = \frac{\left(\frac{t}{\eta}\right)^\beta}{\left(\frac{\tau}{\eta}\right)^\beta} = \left(\frac{t}{\tau}\right)^\beta. \quad (2.21)$$

De acordo com [Crow \(1974\)](#), este processo é usual na literatura de sistemas reparáveis; e [Oliveira, Colosimo e Gilardoni \(2014\)](#) menciona que sua popularização é devida ao fato de sua função ser flexível. Mais precisamente, o PLP pode acomodar tanto intensidades de ocorrências crescentes, que ocorrem quando $\beta > 1$, quanto intensidades decrescentes, $\beta < 1$, ou ainda quando a intensidade de ocorrência do evento de interesse é constante, para os casos em que $\beta = 1$, caracterizando um PPH. A estimação pelo método de máxima verossimilhança dos parâmetros do modelo de RM com a forma paramétrica PLP já está estabelecida na literatura.

2.2.3.1 Uma útil reparametrização para o processo de lei de potência

Rigdon e Basu (2000) propuseram uma reparametrização dos parâmetros do PLP para obter uma melhor interpretação dos parâmetros. Como o parâmetro η significa o tempo até que o número esperado de falhas seja igual a 1, segundo Oliveira, Colosimo e Gilardoni (2012) essa é uma interpretação de difícil compreensão. Dessa forma, foi proposta a seguinte reparametrização:

$$\delta = \left(\frac{\tau}{\eta} \right)^\beta = \mathbb{E}(N(\tau)) = \Lambda(\tau). \quad (2.22)$$

Com esta reparametrização, a função de intensidade é dada por

$$\lambda(t) = \frac{\beta t^{\beta-1} \delta}{\tau^\beta}, \quad (2.23)$$

em que δ é o número esperado de falhas antes do tempo de truncamento τ e β é uma medida de melhoria ou deterioração do sistema, como na parametrização original. Por sua interpretação prática e por ter uma função de intensidade algebricamente mais tratável, muitos autores recentemente têm preferido utilizar esta reparametrização.

2.3 Sistemas reparáveis

Em sistemas reparáveis, após a falha ocorrer, o sistema é reparado e continua a ser observado, ou seja, um sistema é dito reparável se após uma falha o sistema é passível de reparo e após o reparo ele volta a operar. Lindqvist (2006) afirma que na modelagem a suposição implícita é que o sistema é reparado e reinicia imediatamente; isso significa que o tempo de reparo é considerado tão pequeno que pode ser descartado.

Os tempos de falha de um sistema reparável são descritos por um processo estocástico pontual e, segundo Crow (1974), um procedimento comum na prática é aproximar o processo estocástico complicado que descreve a ocorrência das falhas do sistema por um processo estocástico mais simples, que embora não seja exato, ainda produz resultados práticos úteis. Denotamos $0 < T_1 < T_2 < \dots$ os tempos de falha do sistema medido em tempo global, ou seja, tempos acumulados desde o início do funcionamento do sistema ($T_0 = 0$). Os tempos de falha geram um processo de contagem $\{N(t); t \geq 0\}$, em que $N(t)$ conta o número de falhas no intervalo $(0, t]$. O tempo é não necessariamente o tempo do calendário, mas pode ser o tempo de operação, a quilometragem total de um veículo, o número de ciclos, a duração de um componente, entre outros. Após cada falha, uma ação de reparo é feita imediatamente e o sistema volta a funcionar.

Como dito anteriormente, os dados podem ser observados até um tempo pré-determinado, indicando um truncamento por tempo; ou até que um número pré-determinado de falhas ocorra, indicando truncamento por falha. Nesta tese consideraremos apenas o truncamento por tempo, em que todos os sistemas serão observados até um tempo fixado τ . A Figura 5 apresenta uma representação de tempos de falha truncados por tempo, em que são observados k sistemas, todos

são truncados no tempo τ e as falhas são representadas pelo símbolo “x”. Com esta representação, podemos perceber que um estudo em que todos os sistemas são truncados no mesmo tempo, o tempo final do estudo é fixado e, portanto, cada sistema pode apresentar um número diferente de falhas.

Sistema

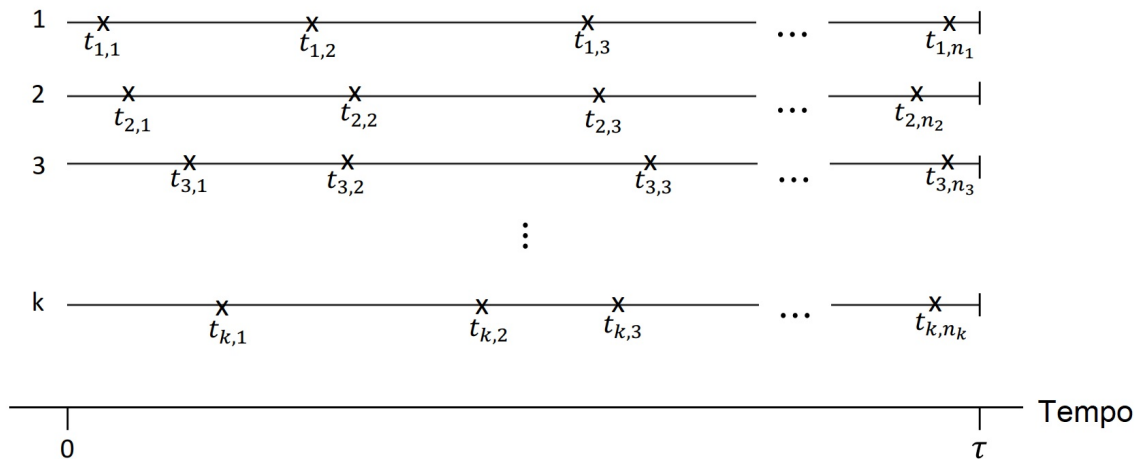


Figura 5 – Exemplo de tempos de falha truncados por tempo (falhas são representadas pelo símbolo “x”).

O tipo de reparo a ser feito depende do tipo de sistema e, claro, do tipo de falha observada. Existem vários níveis de reparos, desde os que corrigem apenas a causa originadora da falha até os que acabam melhorando o sistema deixando-o tão bom quanto um novo. Como já mencionamos, o efeito da ação do reparo no desempenho do sistema é o que caracteriza os diferentes modelos em sistemas reparáveis, pois cada tipo de reparo causa um efeito diferente na intensidade de falha, ou na confiabilidade. Assim como estabelecido em [Rigdon e Basu \(2000\)](#), dizemos que os sistemas podem estar se deteriorando ou em melhoria; um sistema reparável é dito estar se deteriorando se os tempos entre falhas tendem a ir diminuindo com o avanço da idade e, um sistema é dito estar em melhoria se os tempos entre falhas tendem a aumentar. Então, foram construídos modelos que levam em consideração esse tipo de reparo. Os modelos usuais da literatura são o RM, o RP e o RI.

Deve-se destacar que no contexto de sistemas reparáveis, especialmente, quando se analisa dados de sistemas que são passíveis de reparo após as falhas, o processo de Poisson e o de renovação são ditos, respectivamente, modelo de RM e RP. Isto devido ao fato de que no processo de Poisson, pressupõe que após ocorrência de um reparo, a ocorrência de um novo evento é eminente, pois o sistema retorna para o estado que estava exatamente antes da falha. Enquanto que no processo de renovação é assumido que, imediatamente após o reparo, uma nova falha é improvável de ocorrer, uma vez que o sistema volta às condições de funcionamento de um novo. Veremos na próxima seção essas definições de tipos de reparo mais detalhadamente.

2.3.1 Modelo de reparo mínimo

O RM concentra-se em corrigir apenas a causa originadora da falha, deixando o sistema na mesma condição em que estava antes, conhecida como *ABAO*. Usando o conceito de RM, é possível descrever de forma simples o fato de que muitos reparos na vida real trazem o sistema a uma condição que é basicamente a mesma antes da falha. Esse reparo pode ser usado para modelar sistemas onde um componente do sistema é substituído ou reparado. O objetivo da ação de reparo não é trazer o sistema exatamente na mesma condição e sim levar o sistema de volta à operação o mais rápido possível. Mas, observando a condição do sistema após o reparo, é uma suposição razoável dizer que o estado do sistema não mudou.

O processo associado à ocorrência de falha pode ser descrito por um PPNH, no qual a probabilidade de falha em um curto período de tempo não depende do histórico de falhas, mas apenas da falha anterior (MURALIDHARAN, 2008) e, assim, a função de intensidade dependerá apenas desse tempo. De acordo com Kijima (1989), esta suposição é plausível para sistemas que consistem em vários componentes, cada um com seu próprio modo de falha.

Isso significa que a intensidade condicional do processo de falha imediatamente após a falha é a mesma que antes da falha e, portanto, é exatamente como seria se nenhuma falha tivesse ocorrido. Assim, o modelo de RM é um processo com função de intensidade sendo a própria intensidade do PPNH. Na prática, um RM geralmente corresponde a reparar ou a substituir apenas uma parte menor do sistema.

Para ilustrar a função de intensidade de um sistema que é submetido a reparos mínimos, na Figura 6 temos uma representação gráfica da função de intensidade de um sistema que começou a ser observado em $T_0 = 0$ e teve falhas nos tempos $T_1 = 3$ e $T_2 = 4,5$, em que consideramos por fins didáticos $\lambda(t) = 3t^2$. Podemos observar que a cada falha, a intensidade de falha é exatamente a mesma antes da falha, ou seja, a cada RM realizado, a função de intensidade do sistema não se altera, mantendo o sistema na mesma condição em que estava antes da falha.

Para o caso particular em que a função de intensidade assume uma forma constante, ou seja, $\lambda(t) = \lambda$, temos um PPH, caso contrário, é um PPNH. Como a função de intensidade é constante, esse caso particular não é adequado para sistemas que melhoram ou se deterioram ao longo do tempo. De acordo com Fernandes (2015), o PPH com função de intensidade λ é um caso particular dos processos de renovação, em que os *gap times* são variáveis aleatórias iid com distribuição exponencial e média $1/\lambda$. Nesse caso, o tempo global até a n -ésima falha ($T_n = \sum_{j=1}^n Y_j$) segue uma distribuição Gama, em que o parâmetro de forma é n e o parâmetro de escala é $1/\lambda$.

O RM é a suposição mais explorada na literatura e muitos casos especiais do modelo de RM básico foram abordados na literatura.

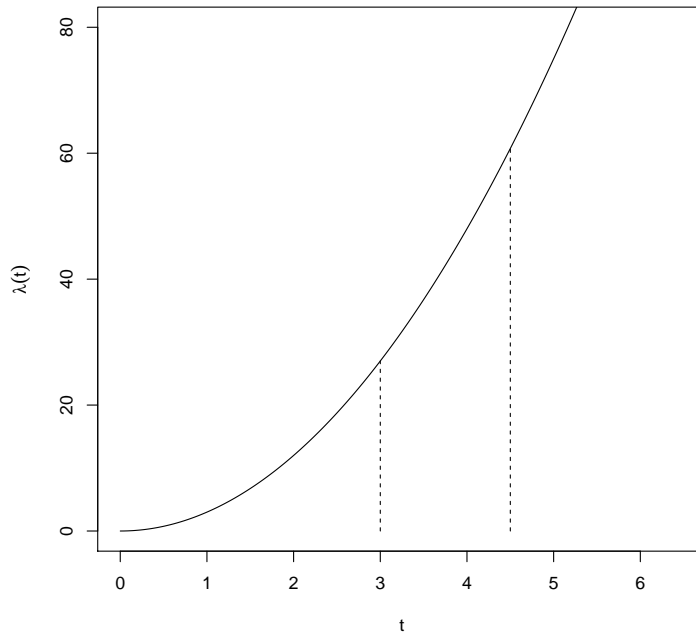


Figura 6 – Função de intensidade do modelo de RM.

2.3.2 Modelo de reparo perfeito

O RP é um tipo de reparo em que geralmente é substituído um componente importante para a operação de um sistema; assim, o sistema pode ser considerado novo, ou seja, o reparo deixou este sistema em melhor condição do que antes da falha, o que caracteriza um tipo de reparo perfeito, já que deixou o sistema em uma condição conhecida como *AGAN*. De acordo com [Kijima \(1989\)](#), assumir RP pode ser razoável para sistemas com uma unidade que é estruturalmente simples. O processo de falha é modelado por um processo de renovação, descrito no Seção 2.2.1, que é caracterizado por usar *gap times*, isto é, tempos entre falhas. Segundo [Doyen e Gaudoin \(2004\)](#), a função de intensidade de falha do modelo de RP é definida por

$$\lambda_P(t) = \lambda(t - T_{N(t-)}). \quad (2.24)$$

em que $t - T_{N(t-)}$ é o tempo desde a última falha estritamente antes do tempo t , $N(t)$ é o número de falhas no tempo t e $\lambda(t)$ é a função de intensidade do processo antes da primeira falha.

Para ilustrar esse modelo, suponhamos um sistema reparável com função de intensidade do processo antes da primeira falha $\lambda(t) = 3t^2$. Se o sistema começou a ser observado em $T_0 = 0$ e teve falhas nos tempos $T_1 = 3$ e $T_2 = 4,5$, sua função de intensidade, de acordo com 2.24, é

dada por

$$\lambda_P(t) = \begin{cases} \lambda(t) = 3t^2, & \text{se } 0 \leq t < 3, \\ \lambda(t-3) = 3(t-3)^2, & \text{se } 3 \leq t < 4,5, \\ \lambda(t-4,5) = 3(t-4,5)^2, & \text{se } 4,5 \leq t < T_3. \end{cases} \quad (2.25)$$

Na Figura 7 temos uma representação gráfica da função de intensidade 2.25, em que podemos observar que a cada falha, a função de intensidade retorna ao valor de zero e tem a mesma forma que tinha antes da primeira falha, ou seja, a cada RP realizado, a função de intensidade do sistema passa a ser exatamente igual à de quando o sistema era novo.

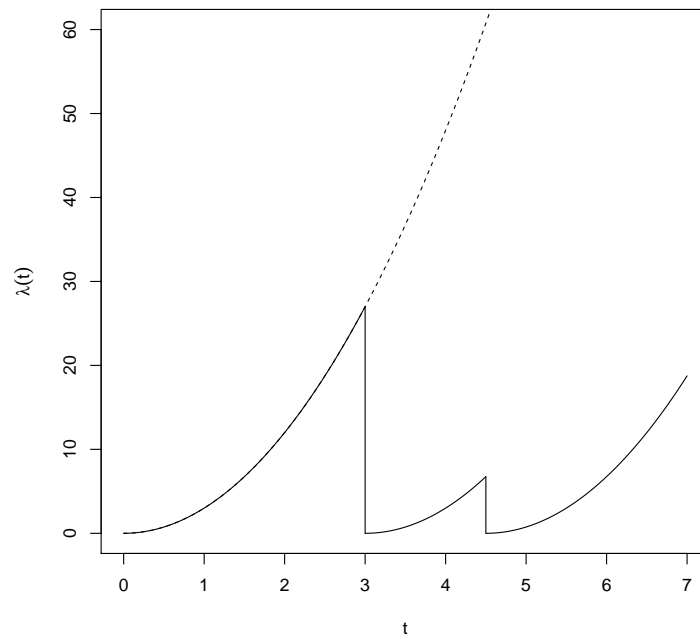


Figura 7 – Função de intensidade do modelo de RP.

A distribuição do processo de contagem $N(t)$ no modelo de RP, isto é, em processos de renovação não é matematicamente tratável, com exceção do caso em que os *gap times* Y_j são variáveis aleatórias seguindo distribuição exponencial, caracterizando um PPH e do caso apresentado por Cook e Lawless (2007) cuja distribuição de $N(t)$ pode ser encontrada através de

$$P(N(t) \geq n) = P(T_n \leq t) = P\left(\sum_{j=1}^n Y_j \leq t\right). \quad (2.26)$$

Mais detalhes dessa distribuição podem ser encontrados em Nakagawa (2006).

Rigdon e Basu (2000) enunciam um resultado interessante para processos de renovação, que diz que a taxa de ocorrência de falhas converge para $1/\mathbb{E}(Y_j)$. Esse resultado encontra-se no Teorema 1.

Teorema 1. Para um processo de renovação Y_1, Y_2, \dots , com $\mathbb{E}(Y_j) = \kappa$ e $V(Y_j) = \sigma^2$,

$$\lim_{t \rightarrow \infty} \frac{\Lambda(t)}{t} = \frac{1}{\kappa}.$$

Na literatura existem diversas extensões do processo de renovação. O modelo Exponencial por Partes (ou *Piecewise Exponential Model*), em que os *gap times* são variáveis aleatórias independentes com distribuição exponencial, mas que não são idênticas, foi proposto por [Rigdon e Basu \(2000\)](#). Os autores propuseram também o Processo Modulado, que é uma generalização do PPNH e do processo de renovação. [Cook e Lawless \(2007\)](#) propuseram algumas extensões, como a inclusão de covariáveis e modelos com efeitos aleatórios. Podemos encontrar ainda outras extensões em [Nakagawa \(2006\)](#).

2.3.3 Modelo de reparo imperfeito

Em muitas situações, as ações de reparo deixam o sistema em um nível intermediário, entre o *AGAN* e o *ABAO*. Segundo [Kijima, Morimura e Suzuki \(1988\)](#), para sistemas compostos de apenas alguns componentes vulneráveis, é mais apropriado considerar o reparo que faz com que o sistema tenha falhado em um nível intermediário. Portanto, as ações de reparo são mais direcionadas para a manutenção do sistema como um todo do que para o fator que origina a falha.

[Brown e Proschan \(1983\)](#) investigaram o processo de falha, em que para cada tempo de falha, ocorre um RP com probabilidade p e ocorre um RM com probabilidade $1 - p$, independente do histórico anterior da falha.

A ideia de idade virtual de um sistema foi introduzida por [Kijima \(1989\)](#), que é uma função positiva de sua idade real e sua história de falhas. O modelo de idade virtual proposto insere um parâmetro, denotado por θ ($0 \leq \theta \leq 1$), o qual representa o grau de eficiência do reparo. Um sistema com idade virtual $v > 0$ é assumido como se comportasse exatamente como um novo sistema que atingiu a idade v sem falhar.

Existem diversas extensões do modelo de Brown e Proschan na literatura. Podemos citar, por exemplo, [Block, Borges e Savits \(1985\)](#), [Whitaker e Samaniego \(1989\)](#), [Doyen e Gaudoin \(2004\)](#) e [Doyen \(2011\)](#). Os modelos propostos por [Doyen e Gaudoin \(2004\)](#) estão entre os mais conhecidos e estudados em RI, e são os que apresentamos aqui.

Duas classes de modelos para o RI foram propostas por [Doyen e Gaudoin \(2004\)](#), a Redução Aritmética de Intensidade (ARI, do inglês *Arithmetic Reduction of Intensity*) e a Redução Aritmética de Idade (ARA, do inglês *Arithmetic Reduction of Age*). Essas classes de modelo são definidas pela sua memória m , onde m refere-se ao número máximo de falhas anteriores envolvidas no cálculo da função de intensidade. Um caso específico é de $m = 1$, o que indica

que cada ação de reparo reduz o desgaste do sistema que ocorre apenas após a última falha. Ou seja, a função de intensidade do sistema é recalculada em cada falha, considerando o tempo de falha mais recente.

Na classe de modelos ARA, o efeito do reparo é expresso por uma redução na idade virtual do sistema, e sua função de intensidade é dada por

$$\lambda_{ARA}(t) = \lambda(t - (1 - \theta)T_{N(t)}). \quad (2.27)$$

Este modelo tem como casos especiais o RM e o RP. Quando $\theta = 1$, o modelo é

$$\lambda_{ARA}(t) = \lambda(t),$$

sendo visto que esta função é igual à função de intensidade do modelo de RM.

O outro caso especial ocorre quando $\theta = 0$, no qual o modelo é

$$\lambda_{ARA}(t) = \lambda(t - T_{N(t)}),$$

sendo essa a função de intensidade de um RP (2.24).

Na classe de modelos ARI, cada ação de reparo reduz a função de intensidade de falha do sistema e, entre duas falhas consecutivas, sua função de intensidade é verticalmente paralela à sua força original. Assim, a função de intensidade é expressa por

$$\lambda_{ARI}(t) = \lambda(t) - (1 - \theta)\lambda(T_{N(t)}). \quad (2.28)$$

Este modelo também tem dois casos especiais. Quando $\theta = 0$, $\lambda_{ARI}(t) = \lambda(t - T_{N(t)})$, que é a função de intensidade do modelo de RP e quando $\theta = 1$, $\lambda_{ARI}(t) = \lambda(t)$, que é a função de intensidade do modelo de RM.

A Figura 8 apresenta o esboço das funções de intensidade referentes aos modelos ARA e ARI para ilustrar a diferença entre esses modelos, em que podemos observar como a função de intensidade diminui após as falhas de cada modelo. No caso do modelo ARA, no tempo $T_1 = 2$, a idade é reduzida de 2 para 0,7. Para o modelo ARI, vemos que no tempo $T_1 = 2$ a intensidade foi reduzida de 32 para 9,9.

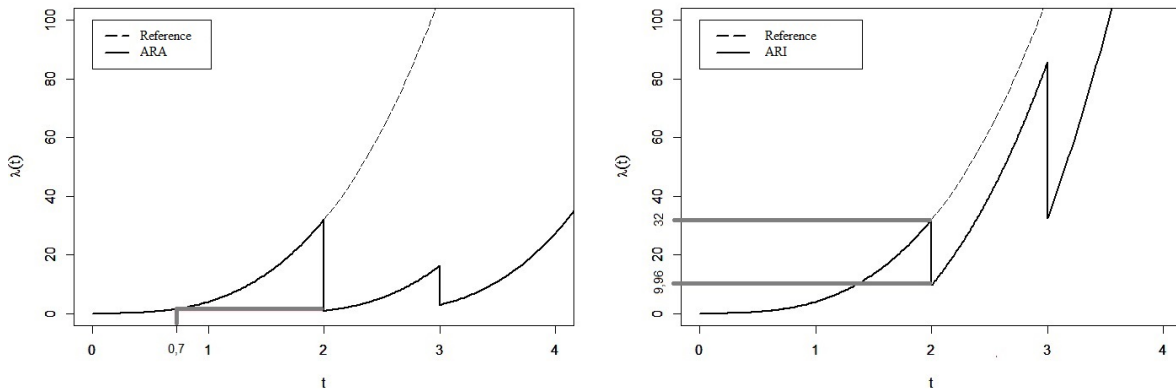


Figura 8 – Funções de intensidade dos modelos ARA e ARI considerando o PLP.

2.4 Modelo de fragilidade

Os métodos tradicionais de análise de confiabilidade assumem que não existe heterogeneidade não-observada e que as observações são independentes. Como resultado, eles não levam em conta o problema da dependência causada pela heterogeneidade não-observada (WIENKE, 2010). O modelo de Cox básico pode ser limitado para um contexto em que os dados de confiabilidade são correlacionados, como ocorre frequentemente em estudos multiníveis. Além disso, a falta de covariáveis importantes limita a capacidade explicativa do modelo, pois a confiabilidade pode ter uma variabilidade que é relativamente maior do que a explicada pelo modelo.

Na literatura relativa aos modelos de fragilidade, vários autores estudaram o uso de modelos multiplicativos de fragilidade que representam uma extensão do modelo de Cox (COX, 1972b). Do ponto de vista clássico, Hougaard (1995), Wienke (2010) e Andersen *et al.* (2012) apresentaram uma revisão desses modelos, enquanto que Sinha e Dey (1997) prepararam uma revisão sob a abordagem Bayesiana. Em dados univariados, o termo fragilidade foi inserido por Vaupel, Manton e Stallard (1979) e Clayton (1978), enquanto que Oakes (1982) trabalhou com modelos para casos multivariados.

O modelo multiplicativo de fragilidade introduz um efeito aleatório, chamado de fragilidade, na função de intensidade com o objetivo de descrever a possível associação entre as unidades e para controlar a heterogeneidade não-observável das unidades em estudo. Assim, a intensidade de falha de um sistema depende de uma variável aleatória não-observável, não-negativa, que atua de forma multiplicativa na função de intensidade.

Considerando uma fonte não-observada de heterogeneidade e covariáveis observadas, a função de intensidade de falha do modelo de fragilidade para o sistema i , com $i = 1, \dots, k$, é definida por

$$\lambda(t_{i,j}|z_i) = z_i \lambda_0(t_{i,j}) \exp(\mathbf{x}'_i \boldsymbol{\zeta}), \quad (2.29)$$

em que $\lambda_0(t_{i,j})$ ($j = 1, \dots, n_i$) é a função de intensidade comum aos sistemas, \mathbf{x}'_i é o vetor das covariáveis observadas do i -ésimo sistema, $\boldsymbol{\zeta}$ é o vetor de parâmetros associados às covariáveis e z_i é a variável de fragilidade iid com fdp $f(z_i)$.

Considerando que a modelagem não captura a fonte de variabilidade de variáveis observáveis, o modelo de fragilidade 2.29 pode ser reescrito por

$$\lambda(t_{i,j}|z_i) = z_i \lambda_0(t_{i,j}). \quad (2.30)$$

A função de confiabilidade condicional à variável de fragilidade é dada por

$$R(t_{i,j}|z_i) = [R_0(t_{i,j})]^{z_i} = e^{[-\Lambda_0(\tau)z_i]}, \quad (2.31)$$

em que z_i é a fragilidade do i -ésimo sistema, $\Lambda_0(t)$ é a função de intensidade acumulada comum a todas as unidades e τ é o tempo de truncamento.

Como z_i representa um valor da variável aleatória não-observável Z , $z_i > 1$ indica que a intensidade de falha do i -ésimo sistema está aumentando e $z_i < 1$ indica que a intensidade de falha está diminuindo, enquanto que para $z_i = 1$ o modelo de fragilidade (2.30) é reduzido para a função de intensidade comum dos sistemas. Neste modelo, como a variável de fragilidade atua de forma multiplicativa, quanto maior o valor desta variável, maiores as chances de ocorrência de falha, ou seja, quanto maior o valor de z_i , mais “frágeis” as unidades serão, daí o nome da fragilidade. Portanto, para essas unidades mais frágeis, espera-se que o evento de interesse ocorra com mais frequência. A variável aleatória de fragilidade Z é considerada iid.

A função de verossimilhança completa que contém os dados observados $(t_{i,j})$ e não-observados (z_i) é dada por (LAWLESS, 1987):

$$L(\boldsymbol{\mu}, z_i | \mathbf{t}) = \prod_{i=1}^k \prod_{j=1}^{n_i} L(\boldsymbol{\mu} | t_{i,j}, z_i) f(z_i), \quad (2.32)$$

em que a variável aleatória z_i tem distribuição $f(z_i)$, $\boldsymbol{\mu}$ é o vetor de parâmetros e n_i é o número total de falhas observadas no i -ésimo sistema.

Para obter a função de confiabilidade não-condicional à variável de fragilidade precisamos integrar o termo de fragilidade, ou seja,

$$R(t) = \int_0^{\infty} R(t|z) f(z) dz. \quad (2.33)$$

A fim de obter a função de intensidade acumulada não-condicional à variável de fragilidade, partimos da confiabilidade não-condicional à variável de fragilidade, definida por

$$\Lambda(t) = -\log(R(t)) = -\log\left(\int_0^{\infty} R(t|z) f(z) dz\right). \quad (2.34)$$

Muitas vezes essa integração não é trivial; nessas situações, uma ferramenta muito útil na análise de fragilidade é a transformada de Laplace (WIENKE, 2010). Dada uma função $f(x)$, a

transformada de Laplace considerado como função do verdadeiro argumento s é definida como

$$Q(s) = \int_0^{\infty} e^{sx} f(x) dx, \quad (2.35)$$

em que Q denota a transformada de Laplace.

A transformada de Laplace é útil em nosso contexto pois tem exatamente a mesma forma da função de confiabilidade não-condicional à variável de fragilidade. Sendo $f(x)$ a fdp da variável de fragilidade e fazendo $s = \Lambda_0(t)$, temos que

$$R(t) = \int_0^{\infty} e^{\Lambda_0(t)z} f(z) dz = Q(\Lambda_0(t)) \quad (2.36)$$

e

$$\Lambda(t) = -\log Q(\Lambda_0(t)), \quad (2.37)$$

e então teremos, de uma forma muito mais fácil, as funções de confiabilidade e de intensidade acumulada não-condicionais à variável de fragilidade, permitindo o cálculo da função de verossimilhança.

2.4.1 Distribuição de fragilidade

Em geral, presume-se que as fragilidades são iid para cada sistema. Se supormos que a variável de fragilidade segue uma distribuição de probabilidade, a escolha dessa distribuição é um ponto importante a ser abordado. Como esse termo de fragilidade atua de forma multiplicativa na função de intensidade, os candidatos à distribuição da fragilidade são não-negativos, geralmente contínuos e não dependentes do tempo. Exemplos de distribuição que se adequam a essas características são as distribuições gama, log-normal, Gaussiana inversa e Weibull. O primeiro modelo de fragilidade para dados de sobrevivência usando distribuição multivariada foi considerado por Clayton (1978) e Oakes (1982).

Devido ao fato de que a distribuição gama tem um tratamento algébrico fácil, ela tem sido amplamente utilizada para modelar a fragilidade e, como a variável de fragilidade não pode ser negativa, a distribuição gama pode atuar de forma direta na função de intensidade de falha. Hougaard (1984) sugeriu para a variável de fragilidade a distribuição Gaussiana inversa, uma vez que a distribuição se torna uma população mais homogênea ao longo do tempo, o que é compatível com a ideia de que o evento de interesse ocorre principalmente às unidades mais fracas. Mais recentemente, Leão *et al.* (2017) propuseram a distribuição de fragilidade Birnbaum-Saunders aplicada à área médica.

Podemos relacionar a distribuição de fragilidade com a heterogeneidade não-observada ao longo do tempo através do coeficiente de variação da distribuição de fragilidade. Assim, de acordo com Tomazella (2003), é razoável avaliar o grau de heterogeneidade não-observada em populações heterogêneas a partir das medidas de dispersão da fragilidade. Então, para um

menor grau de dispersão, a distribuição de fragilidade é mais concentrada e, assim, as unidades pertencem a uma população mais homogênea.

Elbers e Ridder (1982) apresentaram condições para que as distribuições de fragilidade possam ser identificáveis nos modelos de fragilidade com riscos proporcionais. Uma delas é que, se existir a esperança da variável de fragilidade e se pelo menos uma covariável for contínua, a fragilidade Z é identificável se $\mathbb{E}(Z) = 1$.

De acordo com Slimacek e Lindqvist (2016), a escolha da distribuição de fragilidade é uma parte problemática do uso desses modelos, semelhante à escolha da distribuição à priori nos modelos Bayesianos. Ainda segundo Slimacek e Lindqvist (2016), é necessário fazer a integração em relação à variável de fragilidade no uso de distribuições paramétricas de fragilidade e em muitos casos essa integração é feita numericamente, o que em modelos mais complexos podem gerar problemas. Também podem ocorrer problemas quando o modelo não se ajusta bem aos dados causando uma grande variância estimada, e esse problema pode ser mascarado pois uma alta variância em modelos de fragilidade indica que existe alta heterogeneidade não-observada. Para contornar estes problemas, surgiram diversos trabalhos que utilizam a fragilidade de forma não-paramétrica, em que as fragilidades individuais de cada sistema são diretamente consideradas como parâmetros e estimadas diretamente, o que já traz uma outra vantagem desse modelo que é conseguir estimar de forma fácil as fragilidades individuais e conseguir identificar quais sistemas são mais frágeis.

2.5 Inferência Bayesiana objetiva

Os problemas de inferência estatística são normalmente resolvidos condicionados à suposição de que um determinado modelo estatístico é uma descrição adequada do mecanismo probabilístico que gerou os dados, bem como a escolha do modelo que naturalmente envolve um elemento de subjetividade. Isto tornou-se uma prática habitual, no entanto, descrevemos como objetiva qualquer análise estatística que depende apenas do modelo assumido e dos dados observados. Na análise de modelos Bayesianos, a especificação de distribuições à priori para todos os parâmetros desconhecidos é fundamental. Argumentos fundacionais (BERNARDO; SMITH, 1994) ditam que os cientistas devem elicitar uma única distribuição à priori conjunta para todos os elementos desconhecidos do problema, com base em informações disponíveis, e usar o Teorema de Bayes para combinar a priori com as informações fornecidas pelos dados, encapsulado na função de verossimilhança.

É de conhecimento que sob a perspectiva Bayesiana, a distribuição à posteriori para a quantidade de interesse representa a mais completa inferência que podemos fazer a respeito desta quantidade. A distribuição a posteriori combina a informação contida nos dados com a informação prévia que possuíamos sobre a quantidade de interesse. Neste contexto, os parâmetros de interesse são considerados variáveis aleatórias e é possível expressar a incerteza sobre esses

parâmetros antes de observar os dados, utilizando a distribuição à priori para eles, enquanto que a distribuição à posteriori liga a informação contida na amostra com a distribuição à priori e, basicamente, as estimativas Bayesianas podem ser obtidas a partir desta distribuição à posteriori.

Assim, dado os tempos de falha \mathbf{t} e o vetor de parâmetros $\boldsymbol{\mu}$, o procedimento de inferência Bayesiana é baseado na distribuição à posteriori, determinada a partir do teorema de Bayes, que é dada por

$$\pi(\boldsymbol{\mu}|t_{i,j}) \propto L(\boldsymbol{\mu}|t_{i,j})\pi(\boldsymbol{\mu}), \quad (2.38)$$

em que $L(\boldsymbol{\mu}|t_{i,j})$ é a função de verossimilhança, $t_{i,j}$ é o j -ésimo tempo de falha do i -ésimo sistema, para $i = 1, \dots, k$ e $j = 1, \dots, n_i$, em que n_i é o número total de falhas do i -ésimo sistema e $\pi(\boldsymbol{\mu})$ é a distribuição à priori do vetor de parâmetros $\boldsymbol{\mu}$.

Para encontrar os estimadores Bayesianos, eles são na forma de esperanças à posteriori, obtidos através de

$$\mathbb{E}(h(\boldsymbol{\mu})|\mathbf{t}) = \int h(\boldsymbol{\mu})\pi(\boldsymbol{\mu}|\mathbf{t})d\boldsymbol{\mu} \quad (2.39)$$

para uma escolha de função de perda $h(\boldsymbol{\mu})$ de interesse. Mais detalhes podem ser encontrados em O'Hagan (1994) e em Schervish (2012). Ao considerar a função de perda quadrática, o estimador Bayesiano é a média à posteriori, ou seja,

$$\hat{\boldsymbol{\mu}}^{\text{MÉDIA}} = \mathbb{E}(\pi(\boldsymbol{\mu}|\mathbf{t})), \quad (2.40)$$

e, se a função de perda empregada for a 0-1, temos que o estimador Bayesiano é o máximo à posteriori (MAP), que é encontrado maximizando a distribuição à posteriori, ou seja,

$$\hat{\boldsymbol{\mu}}^{\text{MAP}} = \arg \max_{\boldsymbol{\mu}} \pi(\boldsymbol{\mu}|\mathbf{t}), \quad (2.41)$$

e, para encontrar a solução desse tipo de problema, alguns métodos numéricos baseados em amostragem têm sido desenvolvidos, como o método de Monte Carlo via Cadeias de Markov (MCMC).

A função à priori pode ser subjetiva, obtida através da opinião do pesquisador ou especialista na área, conhecida como priori informativa. Há funções à priori que são não-informativas, obtidas por métodos objetivos, em situações onde pouca ou nenhuma informação está disponível. Existem diversas maneiras de encontrar funções à priori não-informativas, uma delas é a priori de Jeffreys (JEFFREYS, 1967). Uma outra função à priori não-informativa é a priori de referência, e as descreveremos a seguir.

2.5.1 *Priori de Jeffreys*

Embora muitos novos métodos tenham surgido, o método de Jeffreys permanece como o mais usado na obtenção de prioris objetivas. Essa preferência pelo método de Jeffreys deve-se,

em grande parte, ao fato dele ser invariante a transformações paramétricas um-a-um, garantindo que a inferência não dependa da particular escolha da parametrização do modelo (JEFFREYS, 1946). Em geral, essa propriedade é considerada como requisito fundamental para uma priori objetiva.

Jeffreys considerou diferentes cenários objetivando-se a construção de distribuições à priori não-informativas (veja Tiao e Box (1973)). Imaginando que o espaço paramétrico fosse um intervalo limitado ou também um intervalo $(-\infty, \infty)$, Jeffreys argumenta que, nestes casos, a distribuição à priori seja considerada uma constante.

A matriz de informação esperada de Fisher resume, com relação às quantidades de interesse, o total de informação decorrente dos dados. Existem muitas aplicações da matriz de informação esperada de Fisher em modelagem e estimação. Dentre elas, destacamos o cálculo da priori de Jeffreys.

Portanto, a priori de Jeffreys é obtida a partir da raiz quadrada do determinante da matriz de informação de Fisher $H(\boldsymbol{\theta})$, ou seja,

$$\pi_J(\boldsymbol{\theta}) \propto |H(\boldsymbol{\theta})|^{1/2}, \quad (2.42)$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros a ser estimado.

Os estimadores obtidos utilizando a priori de Jeffreys costumam ter excelentes propriedades. Essa priori apresenta um bom desempenho em casos uniparamétricos, mas em casos multiparamétricos ela pode não apresentar um desempenho razoável.

2.5.2 Priori de referência

O método de Análise de referência Bayesiana objetiva foi inicialmente proposto por Bernardo (1979) e, posteriormente, desenvolvido por Berger e Bernardo (1989) e Berger, Bernardo *et al.* (1992). O método de análise de referência generaliza muitos dos métodos alternativos propostos e incluem, como casos especiais, as soluções apropriadas que tais métodos apresentam em contextos específicos. A priori de referência é minimamente informativa no sentido da teoria da informação. Além disso, as informações a partir dos dados dominam a informação à priori, refletindo a natureza vaga do conhecimento prévio dos parâmetros de interesse. Tal priori é obtida através da maximização da informação esperada da divergência de Kullback-Leibler entre as distribuições à posteriori e à priori.

Para a construção da priori de referência no caso multiparamétrico, proposta por Bernardo e Ramón (1998), devemos escolher um parâmetro de interesse, que denominaremos de θ , e os outros serão perturbadores, que denominaremos de $\boldsymbol{\lambda} = \lambda_1, \dots, \lambda_m$.

Teorema 2. (Função a priori de referência sob fatoração)

Seja $f(t|\boldsymbol{\theta}, \boldsymbol{\lambda})$ o modelo com $m + 1$ parâmetros, $\boldsymbol{\theta}$ o parâmetro de interesse e seja $h_{i,i}$ o elemento (i, i) da matriz de informação de Fisher $H(\boldsymbol{\theta}, \boldsymbol{\lambda})$ e $s_{i,i}$ o elemento (i, i) de S , em que S é a inversa da matriz de informação de Fisher

Então, se o espaço de parâmetros perturbadores $\Lambda_i(\boldsymbol{\theta}, \lambda_1, \dots, \lambda_{i-1}) = \Lambda_i$ são independentes de ambos $\boldsymbol{\theta}$ e λ_i 's e as funções $s_{1,1}, h_{1,1}, \dots, h_{m,m}$ se fatoram da forma

$$s_{1,1}^{-1/2} = f_0(\boldsymbol{\theta})g_0(\lambda_1, \dots, \lambda_m)$$

e

$$h_{i+1,i+1}^{1/2} = f_i(\lambda_i)g_1(\boldsymbol{\theta}, \lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_m),$$

com $i = 1, \dots, m$, então

$$\pi(\boldsymbol{\theta}) \propto f_0(\boldsymbol{\theta}) \quad (2.43)$$

e

$$\pi(\lambda_i|\boldsymbol{\theta}, \lambda_1, \dots, \lambda_{i-1}) \propto f_i(\lambda_i). \quad (2.44)$$

[Berger et al. \(2015\)](#) propuseram métodos para selecionar uma única priori objetiva comum para todos os parâmetros de um modelo multiparamétrico, em que ela é única no sentido de ser a mesma, independentemente de quais parâmetros do modelo foram tomados para ser de interesse ou de importância inferencial. Neste caso, uma única priori de referência é construída e, assim, temos uma priori de referência geral (chamada de *overall*). Aqui, se estuda particularmente o caso quando a matriz de informação de Fisher é diagonal.

Seja a família $f(x|\boldsymbol{\theta})$ e o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Seja também $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$.

Teorema 3. (Função a priori de referência *overall* (BERGER et al., 2015))

Suponha que a matriz de informação de Fisher de $\boldsymbol{\theta}$ é da forma

$$H(\boldsymbol{\theta}) = \text{diag}(f_1(\theta_1)g_1(\boldsymbol{\theta}_{-1}), \dots, f_k(\theta_k)g_k(\boldsymbol{\theta}_{-k})), \quad (2.45)$$

em que f_i é uma função positiva de θ_i e g_i é uma função positiva de $\boldsymbol{\theta}_{-i}$, para $i = 1, \dots, k$. Então a priori de referência única, para qualquer escolha de parâmetro de interesse e qualquer ordem dos parâmetros perturbadores na derivação, é dada por

$$\pi^R(\boldsymbol{\theta}) \propto \sqrt{f_1(\theta_1) \dots f_k(\theta_k)}. \quad (2.46)$$

A distribuição à posteriori de referência possui boas propriedades teóricas, como invariância sob transformações um-a-um dos parâmetros, consistência sob marginalização e propriedades de amostragem consistentes.

2.5.3 Matching priors

Distribuições à posteriori que combinam probabilidades preditivas posterioris com as probabilidades frequentistas correspondentes são atraentes quando um objetivo principal de uma análise estatística é a construção de regiões de predição. Tais prioris fornecem calibração da previsão Bayesiana ou podem ser vistas como um mecanismo Bayesiano para produzir intervalos de previsão frequentista. Métodos frequentistas, em geral, encontram intervalos de confiança através da teoria assintótica, que frequentemente não garante a probabilidade de cobertura igual à desejada para pequenos e moderados tamanhos amostrais. Para contornar este problema, regras formais foram apresentadas para derivar estimadores intervalares Bayesianos, a fim de garantir a probabilidade de erro de cobertura com $O(n^{-1})$ no sentido frequentista, isto é, seja $\theta_1^{1-\alpha}(\boldsymbol{\pi}, \boldsymbol{t}) | (\theta_1, \theta_2)$ o $(1 - \alpha)$ -ésimo quantil da distribuição à posteriori de θ_1 ,

$$P[\theta_1 \leq \theta_1^{1-\alpha}(\boldsymbol{\pi}, \boldsymbol{t}) | (\theta_1, \theta_2)] = 1 - \alpha - O(n^{-1}). \quad (2.47)$$

A classe de prioris não-informativas $\pi(\theta_1, \theta_2)$ em que o intervalo de credibilidade para o parâmetro de interesse θ_1 tem um erro de cobertura no sentido frequentista é conhecida como *matching priors*; veja mais em [Datta e Mukerjee \(2012\)](#). De acordo com [Tibshirani \(1989\)](#), para encontrar estas prioris é necessário encontrar uma parametrização do modelo a fim de ter parâmetros ortogonais $(\boldsymbol{v}, \boldsymbol{\varphi})$, em que \boldsymbol{v} é o parâmetro de interesse e $\boldsymbol{\varphi}$ é o parâmetro perturbador.

Assim, as *matching priors* são todas as prioris da forma

$$\pi(\boldsymbol{v}, \boldsymbol{\varphi}) = g(\boldsymbol{\varphi}) \sqrt{H_{vv}(\boldsymbol{v}, \boldsymbol{\varphi})}, \quad (2.48)$$

em que $g(\boldsymbol{\varphi}) > 0$ é uma função arbitrária e $H_{vv}(\boldsymbol{v}, \boldsymbol{\varphi})$ é a v -ésima entrada diagonal da matriz de informação de Fisher. Obter prioris com um vetor de parâmetros perturbadores segue a mesma ideia.

ANÁLISE BAYESIANA OBJETIVA PARA MÚLTIPLOS SISTEMAS REPARÁVEIS

Este capítulo tem por objetivo a análise da confiabilidade de diversos sistemas que podem ter várias falhas ao longo do tempo. Sob RM, consideramos que as falhas tenham uma intensidade PLP e a abordagem Bayesiana objetiva é utilizada para estimar os parâmetros do modelo. Os estimadores Bayesianos são obtidos considerando a priori de Jeffreys e priori de referência. Também enfatizamos o uso da forma paramétrica PLP para a função de intensidade do PPNH. Em particular, apresentamos ao longo desta tese uma versão reparametrizada do modelo PLP, conforme Seção [2.2.3.1](#).

3.1 Múltiplos sistemas reparáveis

Em problemas de confiabilidade, assim como em outras áreas, é comum observar um grupo de sistemas similares, porém, não idênticos. O problema estatístico está em combinar a informação dos diferentes sistemas para entender as mudanças no processo sob estudo. Se todos os sistemas observados são considerados ter a mesma intensidade, então a análise de múltiplos sistemas é semelhante à análise de um único sistema. Desta forma, os sistemas são considerados uma amostra aleatória do mesmo processo, e uma única intensidade é estimada. Caso haja evidência de que os sistemas realmente diferem, então uma análise individual de cada sistema será mais adequada. A Figura [9](#) mostra, por exemplo, o cenário para os 5 caminhões dos dados de caminhões basculantes apresentados na Seção [1.2](#).

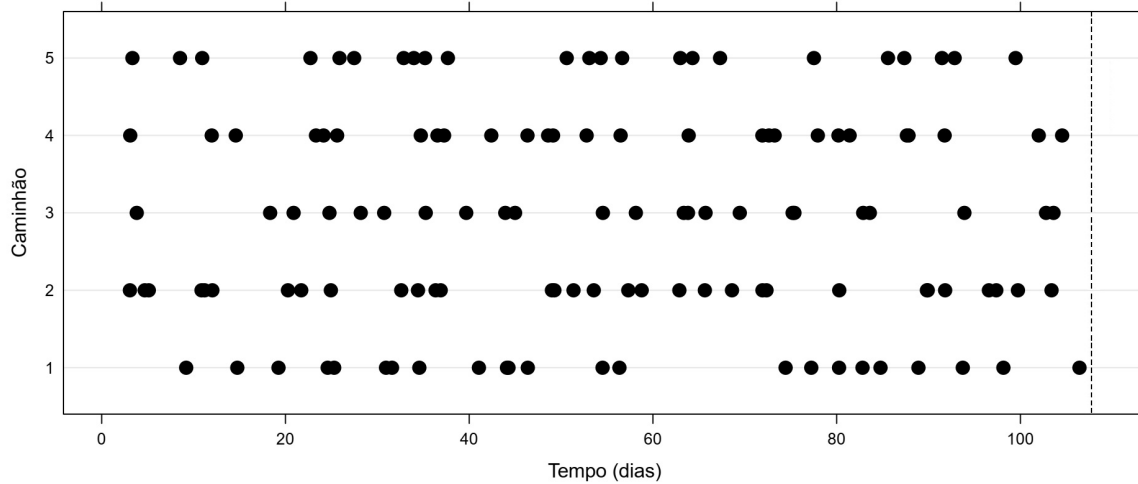


Figura 9 – Histórico de falhas de um grupo de 5 sistemas reparáveis.

Neste contexto destacamos duas suposições importantes. A primeira é assumir que todos os sistemas são idênticos ou diferentes. A segunda é assumir que todos os sistemas têm o mesmo tempo de truncamento τ ou, caso contrário, têm truncamentos diferentes em t_j . Contudo, por uma questão de simplicidade e por estar de acordo com os dados utilizados nesse trabalho, assumimos que o tempo de truncamento τ de cada sistema seja igual. Além disso, nesta tese, assumimos que todos os sistemas são idênticos, ou seja, os sistemas são especificados como k realizações independentes do mesmo processo, com função de intensidade λ . O cenário geral apresentado a seguir formaliza tal situação.

Considere um grupo de k sistemas que geram, de forma independente, falhas de acordo com um PLP. Se os processos de contagem $N_1(t), \dots, N_k(t)$ são todos observados no mesmo tempo τ o PPNH resultante da superposição de PPNHs é dado por $N(t) = \sum_{j=1}^k N_j(t)$ e tem uma função de intensidade dada por $\lambda(t) = k\lambda(t)$, isto é, realizações sobrepostas de um PLP.

Considere $N(t)$ o número de falhas desde o início do acompanhamento do sistema até certo tempo t , então, se $N(t)$ segue um PLP e sua função de intensidade é dada em (2.19) e função de intensidade acumulada (2.20), as quais repetiremos aqui por fins didáticos:

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \quad (3.1)$$

e

$$\Lambda(t) = \left(\frac{t}{\eta} \right)^{\beta}. \quad (3.2)$$

3.1.1 Função de verossimilhança do modelo de RM

Para a construção da função de verossimilhança, precisamos fazer algumas suposições:

- Considere k sistemas reparáveis, para $k = 1, 2, \dots$, em que os sistemas são independentes;
- O i -ésimo sistema é truncado por tempo, sendo observado até um tempo pré-determinado τ , em que $0 < t_{i,1} < t_{i,2} < \dots < t_{i,n_i} < \tau$;
- n_i falhas são observadas no i -ésimo sistema, $i = 1, 2, \dots, k$;
- $N = \sum_{i=1}^k n_i$ é o número total de falhas observadas nos sistemas;
- Seja $t_{i,j}$, $i = (1, 2, \dots, k)$ e $j = (1, 2, \dots, n_i)$, observações da variável aleatória T que representa os tempos de falha para o i -ésimo sistema, registrados como os tempos desde o início do experimento ($t_{i,1} < t_{i,2} < \dots < t_{i,n_i}$);
- Seja $\boldsymbol{\mu} = (\beta, \eta)$ o vetor de parâmetros a ser estimado.

A função de verossimilhança para este processo deve combinar a fdp conjunta dos tempos de falha dos k sistemas. Como os dados são truncados por tempo (τ), o número de falhas $N(t)$ é aleatório, e então ele precisa ser considerado na construção da função de verossimilhança. Assim, temos o seguinte:

$$\begin{aligned}
 L(\boldsymbol{\mu}|\mathbf{t}) &= \prod_{i=1}^k f(t_{i,1}, t_{i,2}, \dots, t_{i,n_i}, n) \\
 &= \prod_{i=1}^k f(t_{i,1})f(t_{i,2}|t_{i,1})f(t_{i,3}|t_{i,1}, t_{i,2}), \dots, f(t_{i,n_i}|t_{i,1}, t_{i,2}, \dots, t_{i,n_i-1}) \\
 &\quad \times P[N(t) = n|t_{i,1}, t_{i,2}, \dots, t_{i,n_i}].
 \end{aligned} \tag{3.3}$$

Com algumas manipulações (ver [Rigdon e Basu \(2000\)](#)), chegamos que a função de verossimilhança para os k sistemas é dada por

$$\begin{aligned}
 L(\boldsymbol{\mu}|\mathbf{t}) &= \prod_{i=1}^k \lambda(t_{i,1})\lambda(t_{i,2})\dots\lambda(t_{i,n_i})e^{-\Lambda(\tau)} \\
 &= \prod_{i=1}^k \prod_{j=1}^{n_i} [\lambda(t_{i,j})] e^{-\Lambda(\tau)}.
 \end{aligned} \tag{3.4}$$

Usando a forma paramétrica do PLP (2.19), a função de verossimilhança para truncamento por tempo sob a suposição de RM é

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left[\frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right] \exp \left[- \left(\frac{\tau}{\eta} \right)^\beta \right] \right\}, \tag{3.5}$$

e o logaritmo da função de verossimilhança é dado por

$$\ell(\boldsymbol{\mu}|\mathbf{t}) = N \log(\beta) - \beta N \log(\eta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) - \eta^{-\beta} \tau^\beta k. \quad (3.6)$$

Como os k sistemas são truncados no mesmo tempo τ , chegamos diretamente nos seguintes EMV para β e η :

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{i,j})} \quad \text{e} \quad \hat{\eta} = \tau \left(\frac{k}{N} \right)^{1/\hat{\beta}}. \quad (3.7)$$

Considerando agora a forma reparametrizada na subseção (2.2.3.1), em que a função de intensidade e a função de intensidade acumulada são dadas, respectivamente, em (2.22) e (2.23), as quais repetiremos aqui por fins didáticos:

$$\lambda(t) = \frac{\beta t^{\beta-1} \delta}{\tau^\beta}, \quad (3.8)$$

e

$$\Lambda(\tau) = \delta. \quad (3.9)$$

Para a construção da função de verossimilhança, basta substituir as funções (3.8) e (3.9) na função de verossimilhança (3.4), obtendo assim

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left[\frac{\beta t_{i,j}^{\beta-1} \delta}{\tau^\beta} \right] e^{-\delta} \right\}, \quad (3.10)$$

e o logaritmo da função de verossimilhança do modelo é dado por

$$\ell(\boldsymbol{\mu}|\mathbf{t}) = N \log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) + N \log(\delta) - \beta N \log(\tau) - k\delta. \quad (3.11)$$

Os EMVs podem ser facilmente encontrados, e são dados por

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{i,j})} \quad \text{e} \quad \hat{\delta} = \frac{N}{k}. \quad (3.12)$$

Caso seja de interesse obter o estimador de η , basta fazer

$$\hat{\eta} = \frac{\tau}{\hat{\delta}^{1/\hat{\beta}}}. \quad (3.13)$$

3.2 Estimação Bayesiana objetiva para o modelo

A inferência Bayesiana começa com a escolha de uma distribuição à priori. Essa distribuição deve retratar o conhecimento prévio do pesquisador sobre o assunto analisado, no entanto, na

prática, nem sempre temos algum conhecimento prévio sobre o assunto. Assim, surgiram prioris não-informativas, que visam demonstrar a falta de conhecimento prévio sobre os parâmetros. Existem vários tipos de prioris que são não-informativas, mas vamos nos concentrar na priori objetiva por ter várias boas propriedades.

Portanto, nesta seção, nós encontraremos a priori de Jeffreys e de referência para o modelo de RM, encontraremos as distribuições à posteriori, mostraremos que são distribuições próprias e mostraremos as estimativas Bayesianas dos parâmetros para cada priori utilizada. Utilizaremos aqui o modelo PLP reparametrizado dado em (2.22) por sua fácil interpretação e por facilidades algébricas.

3.2.1 Priori de Jeffreys

A priori de Jeffreys definida na 2.5.1, é uma priori não-informativa e é obtida a partir da raiz quadrada do determinante da matriz de informação de Fisher. Assim, devemos encontrar a matriz de informação de Fisher para o modelo e, para encontrá-la, devemos calcular a primeira e segunda derivadas da função de log-verossimilhança 3.11, dadas por

$$\frac{\partial \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \beta} = \frac{N}{\beta} + \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) - \sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau),$$

$$\frac{\partial \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \delta} = \frac{N}{\delta} - k,$$

$$\frac{\partial^2 \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \delta \partial \beta} = \frac{\partial^2 \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \beta \partial \delta} = 0,$$

$$\frac{\partial^2 \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \beta^2} = -\frac{N}{\beta^2},$$

e

$$\frac{\partial^2 \ell(\boldsymbol{\mu}|\mathbf{t})}{\partial \delta^2} = -\frac{N}{\delta^2}.$$

Então, a matriz de informação de Fisher é então dada por

$$H(\beta, \delta) = -\mathbb{E} \begin{bmatrix} -\frac{N}{\beta^2} & 0 \\ 0 & -\frac{N}{\delta^2} \end{bmatrix} = k \begin{bmatrix} \frac{\delta}{\beta^2} & 0 \\ 0 & \frac{1}{\delta} \end{bmatrix},$$

pois $\mathbb{E}[N(t)] = \delta$. Encontrada a matriz de informação de Fisher, segue de (2.42) que a priori de Jeffreys para β e δ é dada por

$$\pi_J(\beta, \delta) \propto \sqrt{\frac{\delta}{\beta^2} \frac{1}{\delta}} = \frac{1}{\beta}. \quad (3.14)$$

Propriedades dessa priori com o PLP reparametrizado e também para a parametrização original do PLP foram discutidas em Oliveira, Colosimo e Gilardoni (2012).

Combinando a priori dada em (3.14) e a função de verossimilhança dada em (3.10), podemos encontrar a distribuição à posteriori, dada por

$$\pi(\beta, \delta | \mathbf{t}) \propto \exp \left\{ -\beta \left[\sum_{i=1}^k \sum_{j=1}^{n_i} \log \left(\frac{\tau}{t_{i,j}} \right) \right] \right\} \frac{\beta^N}{\beta} \delta^N \exp(-k\delta).$$

Chamamos $\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log \left(\frac{\tau}{t_{i,j}} \right)}$, ou seja, $\hat{\beta}$ é o EMV de β . Assim, a distribuição à posteriori pode ser reescrita da seguinte forma:

$$\pi(\beta, \delta | \mathbf{t}) \propto \beta^{N-1} \exp \left(\frac{-N\beta}{\hat{\beta}} \right) \delta^N \exp(-k\delta).$$

Conseguimos então encontrar as distribuições à posteriori marginais, que são dadas por

$$\pi(\beta | \mathbf{t}) \propto \text{Gama} \left(\beta \mid N, \frac{N}{\hat{\beta}} \right) \quad \text{e} \quad \pi(\delta | \mathbf{t}) \propto \text{Gama} \left(\delta \mid N+1, k \right), \quad (3.15)$$

ou seja, ambas as posteriors têm distribuição gama, o que facilita o processo de encontrar os estimadores à posteriori.

Das distribuições marginais acima, podemos facilmente provar que a posteriori é própria e, portanto, podemos encontrar os estimadores dos parâmetros. Como visto na Seção 2.5, os estimadores podem ser a média à posteriori ou a moda à posteriori. Os estimadores Bayesianos usando a média à posteriori são dados por

$$\hat{\beta}_J^{MÉDIA} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log \left(\frac{\tau}{t_{i,j}} \right)} \quad \text{e} \quad \hat{\delta}_J^{MÉDIA} = \frac{N+1}{k}. \quad (3.16)$$

Os estimadores Bayesianos MAP são dados por

$$\hat{\beta}_J^{MAP} = \frac{N-1}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log \left(\frac{\tau}{t_{i,j}} \right)} \quad \text{e} \quad \hat{\delta}_J^{MAP} = \frac{N}{k}, \quad (3.17)$$

para $N \geq 2$.

É importante ressaltar que, para o estimador MAP, temos que

$$\mathbb{E}[\hat{\beta}_J^{MAP}] = \mathbb{E} \left[\frac{N-1}{N} \hat{\beta} \right] = \beta \quad \text{e} \quad \mathbb{E}[\hat{\delta}_J^{MAP}] = \mathbb{E} \left[\frac{N}{k} \right] = \delta,$$

em que $\hat{\beta}$ é o estimador de máxima verossimilhança de β . Ou seja, os estimadores Bayesianos MAP são não-viciados, enquanto que os estimadores usando a média à posteriori são viciados.

3.2.2 Priori de referência

A matriz de informação de Fisher encontrada na 3.2.1 é na forma diagonal, ou seja,

$$H(\beta, \delta) = k \begin{bmatrix} \frac{\delta}{\beta^2} & 0 \\ 0 & \frac{1}{\delta} \end{bmatrix},$$

e assim, temos que os parâmetros β e δ são ortogonais. Portanto, podemos usar o Teorema (3) para encontrar a priori de referência *overall*. Então, do teorema temos que a priori de referência conjunta para β e δ é

$$\pi_R(\beta, \delta) \propto \sqrt{f_1(\beta)f_2(\delta)} = \frac{1}{\beta\sqrt{\delta}}. \quad (3.18)$$

Com a priori (3.18) e a função de verossimilhança dada em (3.10), podemos encontrar a distribuição à posteriori, dada por

$$\pi(\beta, \delta|\mathbf{t}) \propto \exp\left\{-\beta \left[\sum_{i=1}^k \sum_{j=1}^{n_i} \log\left(\frac{\tau}{t_{i,j}}\right)\right]\right\} \frac{\beta^N \delta^N \exp(-k\delta)}{\beta \sqrt{\delta}}.$$

Chamemos novamente $\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log\left(\frac{\tau}{t_{i,j}}\right)}$, então conseguimos reescrever a distribuição à posteriori da seguinte forma:

$$\pi(\beta, \delta|\mathbf{t}) \propto \beta^{N-1} \exp\left(\frac{-N\beta}{\hat{\beta}}\right) \delta^{N-1/2} \exp(-k\delta),$$

em que é possível observar que a posteriori apresenta o núcleo de duas distribuições gama, ou seja,

$$\pi(\beta, \delta|\mathbf{t}) \propto \text{Gamma}\left(\beta|N, \frac{N}{\hat{\beta}}\right) \text{Gamma}\left(\delta|N + \frac{1}{2}, k\right).$$

Então, as distribuições à posteriori marginais para β e δ são dadas, respectivamente, por

$$\pi(\beta|\mathbf{t}) \propto \text{Gamma}\left(\beta|N, \frac{N}{\hat{\beta}}\right) \quad \text{e} \quad \pi(\delta|\mathbf{t}) \propto \text{Gamma}\left(\delta|N + \frac{1}{2}, k\right), \quad (3.19)$$

e, similar à priori de Jeffreys, baseado nas distribuições à posteriori marginais encontradas, podemos facilmente provar que a posteriori é própria. Assim, os estimadores Bayesianos MAP são dados por

$$\hat{\beta}_R^{MAP} = \frac{N-1}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log\left(\frac{\tau}{t_{i,j}}\right)} \quad \text{e} \quad \hat{\delta}_R^{MAP} = \frac{N-1/2}{k}. \quad (3.20)$$

E os estimadores de acordo com a média à posteriori são dados por

$$\hat{\beta}_R^{MÉDIA} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log\left(\frac{\tau}{t_{i,j}}\right)} \quad \text{e} \quad \hat{\delta}_R^{MÉDIA} = \frac{N+1/2}{k}. \quad (3.21)$$

Repare que o estimador MAP para o parâmetro β utilizando a priori de referência é não-viciado, enquanto que o estimador MAP para o parâmetro δ e os estimadores dos dois parâmetros utilizando a média à posteriori são viciados.

3.2.3 Matching priors

A classe de prioris não-informativas $\pi(\theta_1, \theta_2)$ em que o intervalo de credibilidade para o parâmetro de interesse θ_1 tem um erro de cobertura no sentido frequentista é conhecida como *matching priors*. A parametrização utilizada nesta seção é ortogonal, o que nos permite encontrar *matching priors*.

Proposição 1. A priori de Jeffreys (3.14) é uma *matching prior* para β .

Prova. Seja $v = \beta$ o parâmetro de interesse e denote por $\varphi = \delta$ o parâmetro perturbador. Então, $\sqrt{H_{vv}(v, \varphi)} = \frac{1}{\beta}$ e $g(\varphi) = 1$, assim sendo, a priori de Jeffreys pode ser escrita como (2.48) e a prova está completa.

Proposição 2. A priori de referência *overall* (3.18) é uma *matching prior* para todos os parâmetros.

Prova. Se β é o parâmetro de interesse e $\varphi = \delta$, então a prova é análoga à priori de Jeffreys quando consideramos $g(\varphi) = \frac{1}{\sqrt{\delta}}$. Se δ é o parâmetro de interesse e $\varphi = \beta$ é o parâmetro de perturbação, então $\sqrt{H_{\delta\delta}(\delta, \beta)} = \frac{1}{\sqrt{\delta}}$ e $g(\beta) = \frac{1}{\beta}$. Portanto, a priori de referência *overall* pode ser escrita como (2.48) e a prova está completa.

Através dos resultados encontrados nas últimas seções, tanto utilizando a priori de Jeffreys quanto utilizando a priori de referência, com essa reparametrização do PLP chegamos em resultados bastante satisfatórios, pois uma das principais dificuldades em fazermos inferência Bayesiana é encontrar distribuições à posteriori próprias e nestes dois casos chegamos em distribuições à posteriori próprias. Outra vantagem foi que chegamos nos dois casos em núcleos de distribuição gama, o que permitiu encontrar de forma analítica os estimadores dos parâmetros, tornando essa modelagem muito atrativa para se utilizar na prática.

A principal vantagem de usar o modelo com a priori de Jeffreys é que os estimadores são não-viciados, enquanto que a principal vantagem do modelo com a priori de referência é ela ser *matching prior* para os seus dois parâmetros, o que significa que o intervalo de credibilidade dos dois parâmetros apresenta uma boa probabilidade de cobertura.

3.3 Estudo de simulação

Nesta seção apresentamos os procedimentos utilizados para a simulação dos dados, assim como os diferentes cenários considerados, os resultados alcançados e a discussão dos resultados. A partir de d réplicas, o objetivo deste estudo de simulação é examinar numericamente as propriedades dos estimadores propostos neste capítulo. Mais especificamente, avaliaremos o impacto do tamanho da amostra sobre as propriedades dos estimadores através de métricas como vício, erro quadrático médio e probabilidade de cobertura.

Neste estudo, os diferentes cenários escolhidos são para avaliar os seguintes aspectos: tamanhos de amostra (entenda tamanho da amostra como número de sistemas) pequenos e grandes, função de intensidade crescente e decrescente, mais e menos falhas em cada sistema. Consideramos amostras de diferentes tamanhos, ou seja, $k = 2, 5, 10, 20$ sistemas. Em todos os casos, assumimos que os sistemas são observados em um período de tempo fixado, ou seja, são truncados em um tempo $\tau = 50$.

Utilizamos dois cenários para os diferentes parâmetros: Para o primeiro cenário assumimos uma função de intensidade crescente $\beta = 3$ ou seja, sistemas se deteriorando e uma situação com menos falhas $\delta = 5$. Para o segundo cenário assumimos uma função de intensidade decrescente $\beta = 0,5$ ou seja, sistemas em processo de melhoria e uma situação com mais falhas $\delta = 10$. Em cada cenário, foram geradas $d = 50.000$ réplicas.

Considerando que as falhas seguem um PPNH, baseado em [Rigdon e Basu \(2000\)](#), seguimos o seguinte algoritmo de simulação de dados:

1. Fixamos os valores dos parâmetros;
2. Para cada sistema, geramos o número de falhas $n_i \sim \text{Poisson}(\delta)$, $i = 1, \dots, k$;
3. No i -ésimo sistema, os tempos de falha $t_{i1}, t_{i2}, \dots, t_{in_i}$ foram gerados através da inversa da função de intensidade, ou seja, $t_{ij} = \tau U_{ij}^{1/\beta}$, em que U_{ij} são números aleatórios da distribuição Uniforme(0,1);
4. Repetimos os itens 1 e 2 d vezes.

A comparação entre as diferentes abordagens é realizada usando o vício e o erro quadrático médio (EQM), que são calculados por:

$$\begin{aligned} \text{Vício}(\beta) &= \sum_{i=1}^d \frac{\hat{\beta}_i - \beta}{d} \quad \text{e} \quad \text{EQM}(\beta) = \sum_{i=1}^d \frac{(\hat{\beta}_i - \beta)^2}{d}, \\ \text{Vício}(\delta) &= \sum_{i=1}^d \frac{\hat{\delta}_i - \delta}{d} \quad \text{e} \quad \text{EQM}(\delta) = \sum_{i=1}^d \frac{(\hat{\delta}_i - \delta)^2}{d}, \end{aligned}$$

em que $d = 50.000$ é o número de réplicas consideradas.

Com base nessas métricas, espera-se que o melhor método de estimação retorne o vício e o EQM mais próximos de zero. A probabilidade de cobertura (PC) dos parâmetros também são calculadas assumindo um nível de confiança/credibilidade de 95%, e devem incluir o valor verdadeiro com a proporção de 0,95. Portanto, a PC é uma medida para avaliar a qualidade dos estimadores intervalares. Para o intervalo de confiança, consideramos as variâncias assintóticas obtidas a partir da matriz de informação de Fisher esperada para construir os intervalos. A partir da abordagem Bayesiana, obtivemos os valores diretamente da função quantílica da distribuição gama.

O *software* R (R Core Team, 2019) foi usado para calcular os resultados. Os EMVs são comparados com as estimativas Bayesianas que são computadas em expressões de forma fechada, portanto, não consideramos o uso de métodos de MCMC. Nas tabelas e figuras a seguir, são apresentados os resultados para os cenários descritos acima.

A Tabela 3 apresenta o vício e o erro quadrático médio para os dois cenários considerando os seguintes tamanhos amostrais, $k = 2, 5, 10, 20$ para os três diferentes métodos de estimação: EMV, estimador Bayesiano MAP usando priori de Jeffreys e estimador Bayesiano MAP usando priori de referência. Quando $\beta = 3$, o EMV apresenta os maiores vícios e os maiores EQM em todos os tamanhos amostrais, enquanto que o vício e o EQM dos estimadores Bayesianos já são pequenos com $k = 2$. Uma situação parecida ocorre quando $\beta = 0,5$ mas, nesse caso, o vício e principalmente o EQM em todos os métodos são menores do que no cenário anterior. Quando analisado o parâmetro δ , nos dois cenários vemos que o estimador utilizando a priori de referência apresenta os maiores vícios, porém com EQM parecido com os outros métodos e, conforme o tamanho amostral aumenta, o vício do estimador utilizando a priori de referência vai se aproximando dos outros métodos. Vemos também que os resultados obtidos para o vício usando a posteriori de Jeffreys são muito próximos de zero; estes achados são consistentes com os resultados teóricos que provaram que ambos os estimadores para β e δ são estimadores não viciados.

Tabela 3 – Vício e EQM das estimativas considerando os três métodos de Estimação.

Parâmetro	Método	$k = 2$		$k = 5$		$k = 10$		$k = 20$	
		Vício	EQM	Vício	EQM	Vício	EQM	Vício	EQM
$\beta = 3$	EMV	0,3825	2,0550	0,1275	0,4603	0,0618	0,2001	0,0307	0,0944
	Jeffreys	-0,0019	1,3454	-0,0020	0,4054	-0,0002	0,1882	0,0003	0,0916
	Referência	-0,0019	1,3454	-0,0020	0,4054	-0,0002	0,1882	0,0003	0,0916
$\delta = 5$	EMV	0,0296	2,4255	0,0386	0,9696	0,0318	0,4834	0,0352	0,2466
	Jeffreys	0,0296	2,4255	0,0386	0,9696	0,0318	0,4834	0,0352	0,2466
	Referência	-0,2204	2,4732	-0,0614	0,9719	-0,0182	0,4827	0,0102	0,2454
$\beta = 0,5$	EMV	0,0282	0,0177	0,0102	0,0056	0,0051	0,0026	0,0026	0,0013
	Jeffreys	0,0002	0,0149	-0,0002	0,0053	0,0000	0,0026	0,0001	0,0013
	Referência	0,0002	0,0149	-0,0002	0,0053	0,0000	0,0026	0,0001	0,0013
$\delta = 10$	EMV	0,0098	5,0054	0,0021	1,9975	-0,0054	0,9996	0,0055	0,5014
	Jeffreys	0,0098	5,0054	0,0021	1,9975	-0,0054	0,9996	0,0055	0,5014
	Referência	-0,2402	5,0630	-0,0979	2,0071	-0,0554	1,0027	-0,0195	0,5018

Na Tabela 4, encontramos os resultados da PC em cada um dos cenários com os três métodos de estimação. De forma geral, os três métodos alcançam a PC muito próxima à esperada mesmo com um tamanho de amostra pequeno e com poucas falhas em cada sistema. Podemos destacar que o estimador Bayesiano com a priori de referência atinge a PC próxima a 0,95 em todos os cenários; é o melhor método nesse sentido. Esse resultado vai de encontro com os resultados teóricos, que diziam que a priori de referência *overall* de β e δ é *matching prior*.

Tabela 4 – Probabilidades de cobertura para os parâmetros considerando os três métodos de Estimação.

Parâmetro	Método	$k = 2$	$k = 5$	$k = 10$	$k = 20$
$\beta = 3$	EMV	0,9545	0,9510	0,9515	0,9505
	Jeffreys	0,9497	0,9501	0,9514	0,9491
	Referência	0,9497	0,9501	0,9514	0,9491
$\delta = 5$	EMV	0,9353	0,9382	0,9477	0,9538
	Jeffreys	0,9606	0,9461	0,9477	0,9505
	Referência	0,9480	0,9599	0,9564	0,9561
$\beta = 0,5$	EMV	0,9521	0,9508	0,9501	0,9493
	Jeffreys	0,9500	0,9492	0,9503	0,9480
	Referência	0,9500	0,9492	0,9503	0,9480
$\delta = 10$	EMV	0,9262	0,9397	0,9480	0,9485
	Jeffreys	0,9394	0,9416	0,9475	0,9472
	Referência	0,9574	0,9521	0,9547	0,9517

As Figuras 10 e 11 apresentam o vício, o EQM e a PC no cenário $\beta = 3$ e $\delta = 5$ e as Figuras 12 e 13 apresentam o vício, o EQM e a PC no cenário $\beta = 0,5$, $\delta = 10$, respectivamente. Nas figuras, os tamanhos amostrais são de $k = 2, 4, 6, \dots, 30$. De uma forma geral, podemos destacar que o EMV apresenta maiores vícios para o parâmetro β , maiores EQM para o parâmetro β e é o método que precisa de um tamanho amostral maior para atingir a PC muito próxima à nominal. Podemos destacar também que o estimador Bayesiano utilizando a priori de referência apresenta maiores vícios para o parâmetro δ e é o estimador que alcança a PC próxima a 0,95 com menores tamanhos amostrais. O estimador Bayesiano utilizando a priori de Jeffreys se comporta bem de uma forma geral em todos os aspectos.

No geral, os estimadores Bayesianos com a priori de Jeffreys retornaram estimativas muito próximas às verdadeiras para ambos os parâmetros, enquanto que os estimadores Bayesianos com a priori de referência retornaram intervalos de credibilidade com PC muito próxima a 0,95 em ambos os parâmetros. Chegamos à conclusão que ambos os estimadores apresentaram boas propriedades, então sugerimos o uso de ambos. O uso da priori de Jeffreys é mais aconselhável portanto em situações em que se queira vício menor nos estimadores pontuais e, se a análise estiver mais interessada nas estimativas intervalares, a distribuição à posteriori com a priori de referência deve ser utilizada.

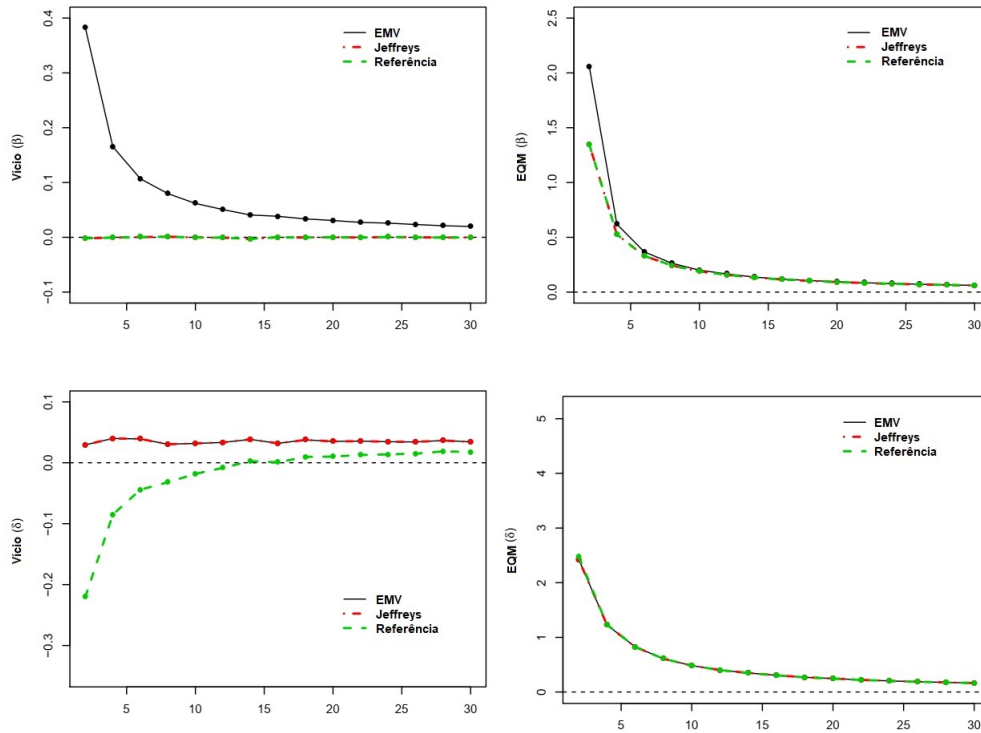


Figura 10 – Vício, EQM para estimativa dos parâmetros para os diferentes métodos de estimação considerando o cenário com $\beta = 3, \delta = 5$.

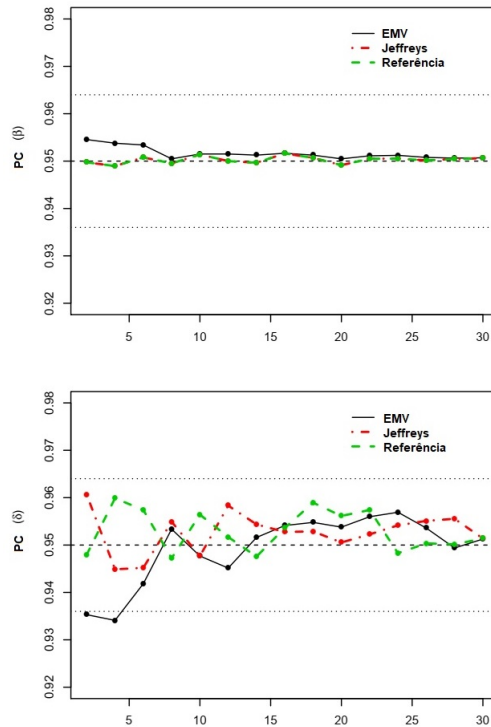


Figura 11 – PC das estimativas dos parâmetros para os diferentes métodos de estimação para o cenário com $\beta = 3, \delta = 5$.

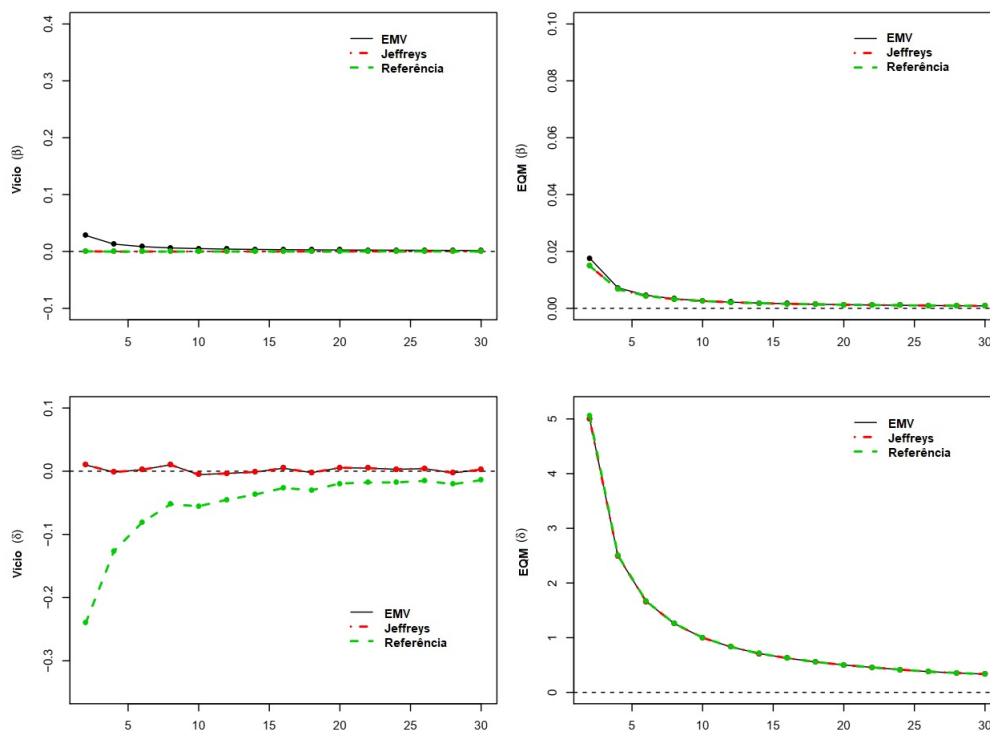


Figura 12 – Vício, EQM das estimativas dos parâmetros para os diferentes métodos de estimação para o cenário com $\beta = 0.5, \delta = 10$.

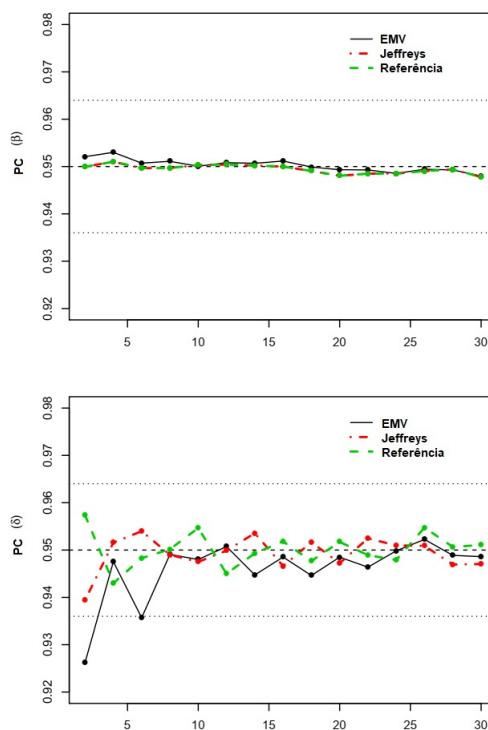


Figura 13 – PC das estimativas dos parâmetros para os diferentes métodos de estimação considerando o cenário com $\beta = 0.5, \delta = 10$.

3.4 Aplicação - Dados de colheitadeiras

Nesta seção analisamos o conjunto de dados de colheitadeiras de cana-de-açúcar, que são relacionados à quebra da lâmina Chopper de colheitadeiras de cana-de-açúcar, descritos na Seção 1.2.1.

O conjunto de dados descreve os tempos de falha de 3 colheitadeiras, com um total de 38 falhas e foi truncado em 195 dias. A Figura 14 apresenta a função média acumulada (MCF, do inglês *mean cumulative function*), ou seja, a estimativa não-paramétrica de Nelson-Aalen para a função $\Lambda(t)$. É possível observar que há uma leve indicação de que a função de intensidade está diminuindo.

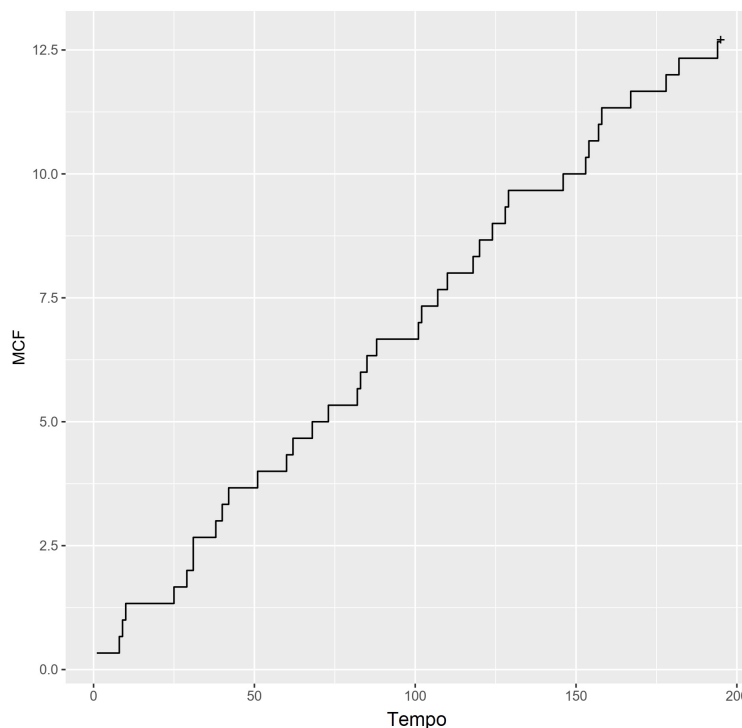


Figura 14 – Estimativa da MCF para os dados de colheitadeiras de cana-de-açúcar.

É importante verificar se o PLP é um modelo adequado aos dados antes de aplicá-lo. A adequação do PLP foi verificada através do gráfico *Duane plot* (RIGDON; BASU, 2000) apresentado na Figura 15, e podemos observar que o PLP é adequado para todas as três colheitadeiras de cana-de-açúcar por apresentarem curvas bastante próximas à de referência.

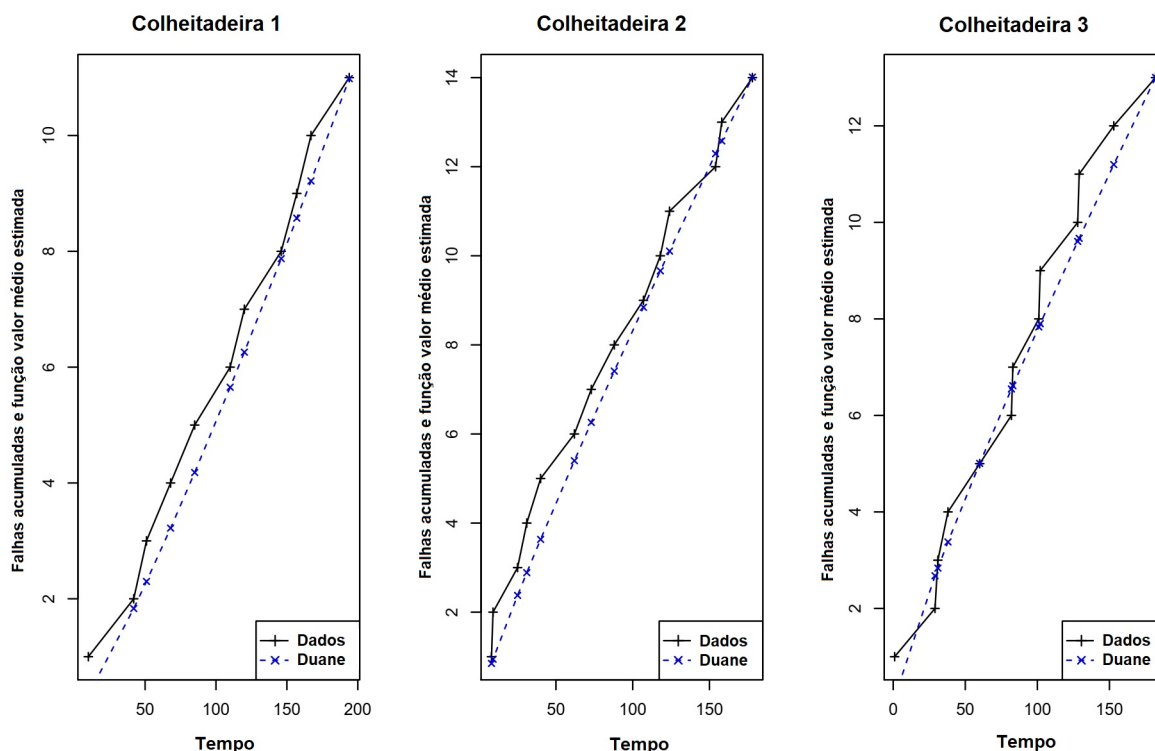


Figura 15 – Duane plot para as colheitadeiras de cana-de-açúcar.

A Tabela 5 apresenta as estimativas dos parâmetros do modelo para os três métodos apresentados no estudo: EMV, priori de Jeffreys e priori de referência. As estimativas foram encontradas usando os estimadores dados em (3.12), (3.17) e (3.20). O intervalo de confiança (IC) do EMV foi encontrado por meio da teoria assintótica, ou seja, encontramos a variância dos estimadores através da matriz de informação de Fisher e utilizamos os quantis da distribuição normal. Os intervalos de credibilidade usando as prioris de Jeffreys e de referência foram encontrados usando os quantis de sua distribuição à posteriori, dadas em (3.15) e (3.19), respectivamente.

Tabela 5 – Estimativas para os parâmetros do modelo.

Método	Parâmetro	Estimativa	IC
EMV	β	0,899	(0,614 ; 1,186)
	δ	12,667	(8,639 ; 16,694)
Jeffreys	β	0,876	(0,637 ; 1,208)
	δ	12,667	(8,964 ; 16,999)
Referência	β	0,876	(0,637 ; 1,208)
	δ	12,500	(9,104 ; 17,193)

As estimativas da Tabela 5 sugerem que a confiabilidade dos sistemas está aumentando ao longo do tempo, ou seja, a função de intensidade está diminuindo pois o parâmetro $\beta < 1$, o que está de acordo com o gráfico MCF. Além disso, espera-se que aconteçam 12,5 falhas em cada colheitadeira de cana no período de 195 dias. Essas conclusões podem ser tomadas

através dos três métodos, já que as estimativas são muito próximas em todos os três métodos utilizados. Também observamos que os IC obtidos por meio dos três métodos são bem próximos. Como provamos teoricamente e a partir de um estudo de simulação, as estimativas de Bayes obtidas com a posteriori de referência *overall*, ou as obtidas com a priori de Jeffreys, devem ser utilizadas.

3.5 Aplicação - Dados de caminhões basculantes

Analisaremos agora o conjunto de dados “Dados caminhões basculantes”, relacionados aos tempos de falha de 5 caminhões de uma mineradora, em um total de 129 falhas com um tempo de truncamento de 107 dias, descritos na Seção 1.2.2. A Figura 16 apresenta a função MCF, em que é possível observar que há uma indicação de que a função de intensidade está aproximadamente constante ao longo do tempo.

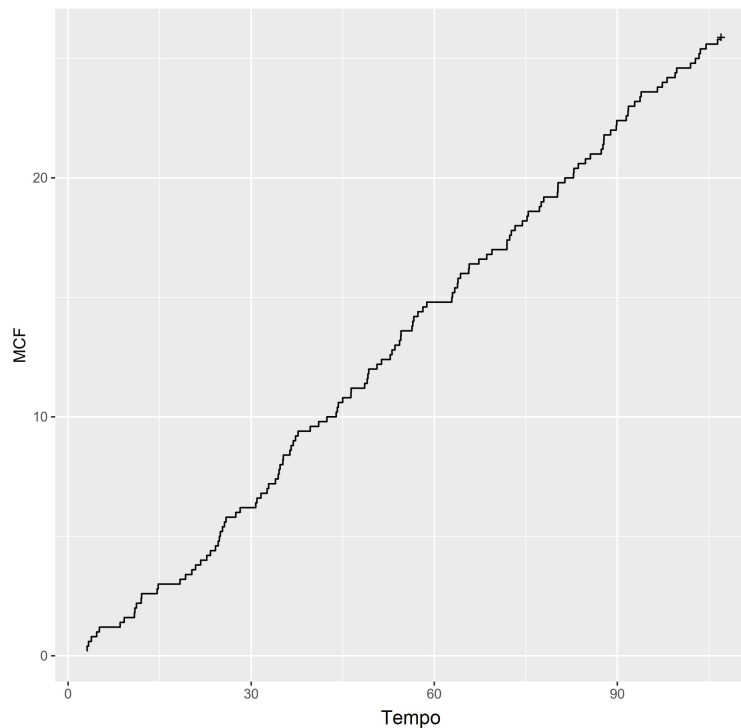


Figura 16 – Estimativa da MCF para os dados dos caminhões basculantes.

O gráfico *Duane plot* para os cinco caminhões é apresentado na Figura 17, e pode-se observar que o PLP é adequado para todos os cinco caminhões basculantes, uma vez que a curva dos cinco caminhões está bem próxima à de referência.

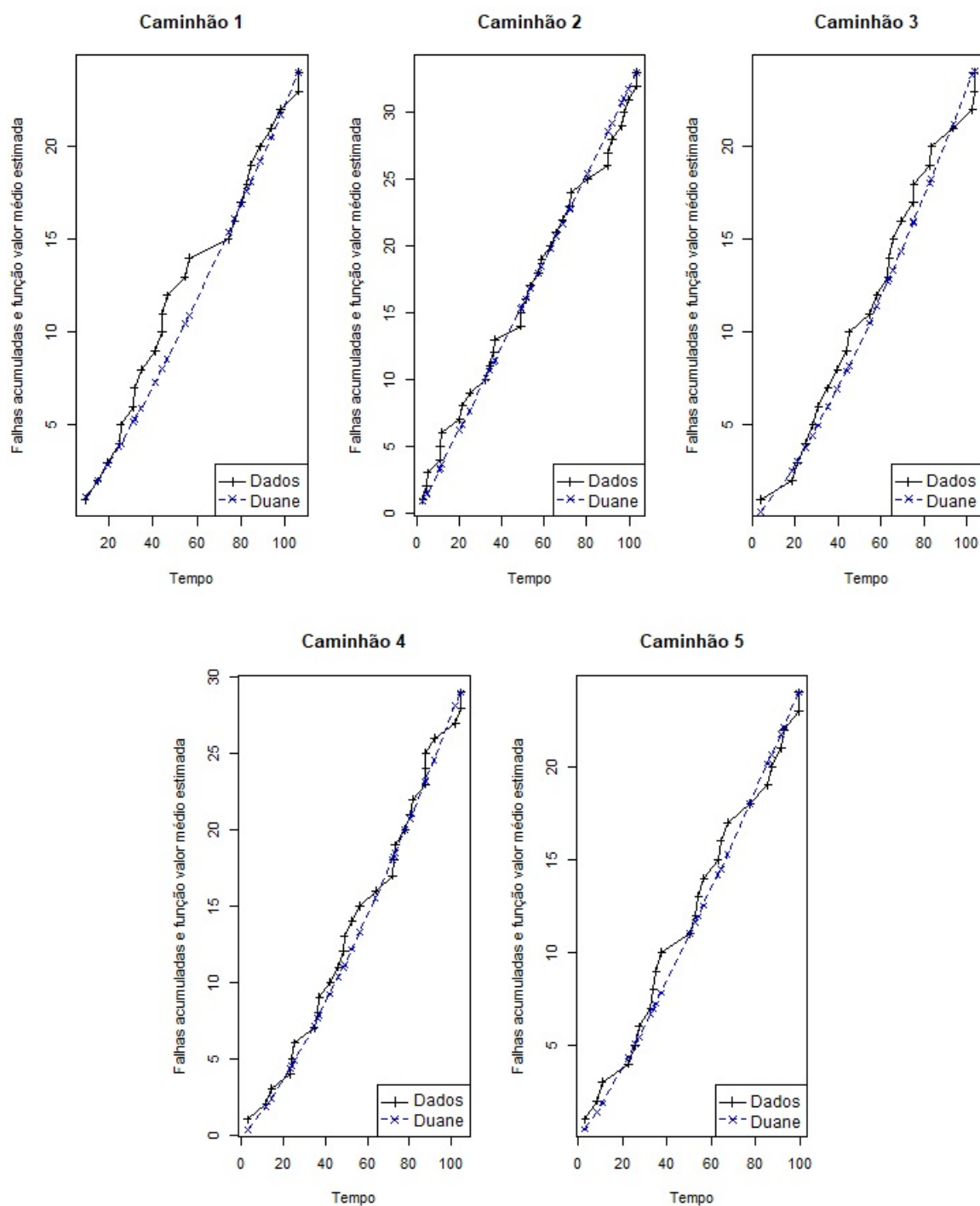


Figura 17 – Duane plot para os caminhões basculantes.

Ajustamos também para estes dados o modelo com os três métodos apresentados no estudo: EMV, priori de Jeffreys e priori de referência e as estimativas encontram-se na Tabela 6. Podemos notar que as estimativas são muito próximas em todos os três métodos utilizados. A partir das estimativas de β , temos um indicativo de que a confiabilidade dos sistemas está diminuindo muito pouco ao longo do tempo, ou seja, a função de intensidade está ligeiramente aumentada, o que está de acordo com o gráfico MCF, pois vemos que $\hat{\beta} > 1$ mas que é muito

próximo de 1. Observamos também que espera-se que aconteça 25,8 falhas em cada caminhão basculante no período de 107 dias. Nesta aplicação, os intervalos de confiança obtidos por meio dos três métodos estão muito próximos. Por fim, ressaltamos que os três métodos de estimação direcionam às mesmas conclusões.

Tabela 6 – Estimativas para os parâmetros do modelo.

Método	Parâmetro	Estimativa	IC
EMV	β	1,095	(0,906 ; 1,284)
	δ	25,800	(21,348 ; 30,252)
Jeffreys	β	1,086	(0,914 ; 1,292)
	δ	25,800	(21,540 ; 30,438)
Referência	β	1,086	(0,914 ; 1,292)
	δ	25,700	(21,631 ; 30,547)

3.6 Considerações finais

Neste capítulo, nos concentramos na análise da confiabilidade de vários sistemas que podem ter várias falhas ao longo do tempo. sob o pressuposto de RM, em que assumimos que as falhas seguem um PLP e, a fim de facilitar a interpretação e o processo de estimação, considerou-se uma reparametrização útil para obtenção dos estimadores Bayesianos.

Começamos o capítulo mostrando como encontrar a função de verossimilhança para múltiplos sistemas reparáveis, todos truncados em um mesmo tempo. Encontramos os EMV para os parâmetros supondo que a forma paramétrica do RM é o PLP na parametrização original e na reparametrização que vimos ter uma interpretação mais fácil.

Utilizando as abordagens Bayesianas, os estimadores de parâmetros do modelo PLP reparametrizado foram obtidos em expressões de forma fechada. Discutimos duas prioris objetivas conhecidas como priori de Jeffreys e priori de referência. As distribuições à posteriori resultantes levam a estimadores com forma fechada. Considerando a priori de Jeffreys, provamos que os estimadores resultantes levam a estimadores MAP não-viciados para ambos os parâmetros. Por outro lado, a priori de referência *overall* forneceu intervalos de credibilidade com cobertura frequentista precisa para ambos os parâmetros, isto é, a priori de referência é uma *matching prior*. As distribuições à posteriori obtidas são próprias e possuem a propriedade de invariância um-a-um. Um extenso estudo de simulação foi realizado confirmando nossos resultados teóricos. No geral, os estimadores Bayesianos com a priori de Jeffreys retornaram estimativas melhoradas para ambos os parâmetros, bem como bons intervalos de credibilidade e devem ser usados para obter as estimativas à posteriori para os parâmetros. Observe que, se a análise estiver mais interessada nas estimativas intervalares, a distribuição à posteriori com a priori de referência deve ser usada.

Um conjunto de dados reais relacionados à quebra da lâmina *Chopper* de colheitadeiras de cana-de-açúcar foi utilizado para confirmar a aplicabilidade da metodologia proposta. Pôde-se observar que as lâminas ao longo do tempo têm sua confiabilidade aumentada e durante os 195 dias do estudo elas têm 12,5 falhas como média. Um segundo conjunto de dados reais, relacionados a caminhões de uma mineradora também foi utilizado, em que percebemos que sua confiabilidade é aproximadamente constante ao longo do tempo, com ligeira tendência a aumentar a intensidade de falhas. Vimos também que em 107 dias, os caminhões falham em média 25,7 vezes. Por fim, concluímos que as abordagens Bayesianas retornam estimativas mais acuradas do que os EMVs, observadas pelas diferenças entre os vícios, EQM e intervalos de confiança.

MÚLTIPLOS SISTEMAS REPARÁVEIS COM UM TERMO DE FRAGILIDADE

Na prática, muitos sistemas reparáveis embora aparentemente idênticos, apresentam grande heterogeneidade entre si e, muitas vezes, essa heterogeneidade não pode ser descrita por covariáveis observadas. Na literatura de análise de sobrevivência, essa heterogeneidade não-observável é denominada fragilidade e, em modelos de sobrevivência, a fragilidade é modelada através da multiplicação (chamado de modelo de fragilidade multiplicativo) ou pela adição (chamado de modelo de fragilidade aditivo) da função de risco por uma variável aleatória positiva com média um e, portanto, a heterogeneidade é descrita por sua variância. Neste capítulo analisamos dados com múltiplos sistemas reparáveis supondo que existem fatores não-observáveis que causam heterogeneidade entre os sistemas. Na modelagem utilizamos um termo aleatório (variável de fragilidade) na função de intensidade de forma multiplicativa e consideramos para a variável de fragilidade a distribuição gama.

4.1 Múltiplos sistemas reparáveis com fragilidade gama

Considerando dados de múltiplos sistemas reparáveis, em que acredita-se que fatores não-observáveis causam heterogeneidade entre os sistemas, o modelo considerando apenas a função de intensidade do RM não é adequado. Dessa forma, partindo do pressuposto que existe heterogeneidade entre os sistemas causada por fatores externos, o modelo aqui proposto inclui um efeito aleatório que capta o efeito de fatores externos e, assim, condicionado a esse efeito os sistemas podem ser considerados independentes. Esse efeito aleatório é considerado na modelagem uma variável aleatória não-observável que representa a informação que não pôde ou não foi observada, em que essa variável é denominada fragilidade.

Modelos de fragilidade paramétricos são muito comuns na literatura de confiabilidade e sobrevivência. Como candidata à distribuição da variável de fragilidade consideramos a

distribuição gama, sendo esta uma das mais utilizadas na literatura. Um dos primeiros estudos de heterogeneidade gama foi em [Lawless \(1987\)](#). [Lindqvist \(2006\)](#) faz algumas discussões gerais envolvendo a modelagem de fragilidade. Mais recentemente, [Giorgio, Guida e Pulcini \(2014\)](#) apresentaram um modelo que mede o efeito de covariáveis observadas e da heterogeneidade não-observada para um sistema de ônibus urbanos e suburbanos. A variável de fragilidade irá atuar de forma multiplicativa na função de intensidade de falha, portanto estamos em um modelo chamado modelo de fragilidade multiplicativo.

Para determinarmos o modelo de fragilidade multiplicativo gama para dados de múltiplos sistemas reparáveis, considere a função de intensidade do modelo de fragilidade (2.30), cuja função base $\lambda_0(t)$ é a função de intensidade sob a hipótese de RM e a forma paramétrica do PLP (2.19). Assim, a função de intensidade de falha condicionada à variável de fragilidade $Z = z_i$ é dada por

$$\lambda(t_{i,j}|z_i) = z_i \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1}, \quad (4.1)$$

em que z_1, z_2, \dots, z_k são as variáveis de fragilidade.

Como visto na Seção 2.4.1, é necessário definir a distribuição da variável de fragilidade; boas referências nesse assunto são [Aalen, Borgan e Gjessing \(2008\)](#) e [Lawless \(1987\)](#). A escolha adequada dessa distribuição pode levar a resultados gerais interessantes. A variância é a quantidade de interesse dessa variável, pois com uma alta variância sabe-se que os sistemas são altamente heterogêneos, e com baixa variância os sistemas são pouco heterogêneos.

Neste trabalho, supomos que a variável de fragilidade tenha distribuição gama por ser de tratamento algébrico conveniente. A parametrização utilizada será Gama $(1/\alpha, 1/\alpha)$, com valor esperado $\mathbb{E}[Z] = 1$, evitando problemas de identificabilidade, e variância $\text{Var}[Z] = \alpha$, assim α é o parâmetro que quantifica a heterogeneidade não-observada e a dependência entre as unidades em estudo, ou seja, quanto maior for a variância de Z maior será o grau de heterogeneidade entre os sistemas. A fdp de Z é então dada por

$$f(z) = \frac{\frac{1}{\alpha}}{\Gamma\left(\frac{1}{\alpha}\right)} z^{\frac{1}{\alpha}-1} e^{-\frac{1}{\alpha}z}. \quad (4.2)$$

A função de confiabilidade condicionada à variável $Z = z_i$ é definida por

$$R(\tau|z_i) = [R_0(\tau)]^{z_i} = \left\{ \exp \left[- \left(\frac{\tau}{\eta} \right)^\beta \right] \right\}^{z_i}, \quad (4.3)$$

em que τ é o tempo de truncamento dos sistemas.

Usando as mesmas suposições para a construção da função de verossimilhança da Seção 3.1.1, mas aqui com o vetor de parâmetros a ser estimado $\boldsymbol{\mu} = (\beta, \eta, \alpha)$ e considerando a função de verossimilhança do modelo de RM (3.4), podemos reescrever a função de verossimilhança do

modelo de fragilidade (2.32) como segue:

$$L(\boldsymbol{\mu}, z_i | \mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \{z_i \lambda_0(t_{i,j})\} e^{-\Lambda_0(\tau) z_i} f(z_i) \right\}. \quad (4.4)$$

Integrando (4.4) com relação a z_i , em que $f(z)$ é encontrada em (4.2), a função de verossimilhança não-condicional à variável de fragilidade é dada por

$$L(\boldsymbol{\mu} | \mathbf{t}) = \int_0^\infty \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \{z_i \lambda_0(t_{i,j})\} e^{-\Lambda_0(\tau) z_i} \frac{1}{\Gamma(\frac{1}{\alpha})} z_i^{\frac{1}{\alpha}-1} e^{-\frac{1}{\alpha} z_i} \right\} dz_i. \quad (4.5)$$

Considerando o PLP, temos:

$$\begin{aligned} L(\boldsymbol{\mu} | \mathbf{t}) &= \int_0^\infty \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ z_i \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right\} e^{-\left(\frac{\tau}{\eta}\right)^\beta z_i} \frac{1}{\Gamma(\frac{1}{\alpha})} z_i^{\frac{1}{\alpha}-1} e^{-\frac{1}{\alpha} z_i} \right\} dz_i \\ &= \int_0^\infty \prod_{i=1}^k \left\{ z_i^{n_i} \frac{1}{\Gamma(\frac{1}{\alpha})} z_i^{\frac{1}{\alpha}-1} e^{-\frac{1}{\alpha} z_i} e^{-\left(\frac{\tau}{\eta}\right)^\beta z_i} \prod_{j=1}^{n_i} \left\{ \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right\} \right\} dz_i \\ &= \prod_{i=1}^k \left\{ \frac{1}{\Gamma(\frac{1}{\alpha})} \int_0^\infty z_i^{n_i + \frac{1}{\alpha} - 1} e^{-\left[\frac{1}{\alpha} + \left(\frac{\tau}{\eta}\right)^\beta\right] z_i} dz_i \right\} \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right\} \right\}, \end{aligned}$$

em que podemos notar que chegamos a uma integral com núcleo de uma distribuição gama: $\text{Gama}\left(n_i + \frac{1}{\alpha}, \frac{1}{\alpha} + \left(\frac{\tau}{\eta}\right)^\beta\right)$, e então

$$L(\boldsymbol{\mu} | \mathbf{t}) = \prod_{i=1}^k \left\{ \frac{1}{\Gamma(\frac{1}{\alpha})} \frac{\Gamma(n_i + \frac{1}{\alpha})}{\left[\frac{1}{\alpha} + \left(\frac{\tau}{\eta}\right)^\beta\right]^{(n_i + \frac{1}{\alpha})}} \right\} \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right\} \right\}. \quad (4.6)$$

Com a função de verossimilhança (4.6) não conseguimos encontrar os EMVs de forma analítica, então são necessários métodos numéricos para realizar a estimação dos parâmetros.

4.1.1 Função de sobrevivência e risco não-condicional usando transformada de Laplace

Uma ferramenta utilizada em modelos de fragilidade é a transformada de Laplace, dada em (2.35). Ela facilita a integração em relação à variável de fragilidade. Uma vantagem dela em relação à integração convencional é a possibilidade de encontrar diretamente a função de confiabilidade, de intensidade e de intensidade não-condicional a Z e, claro, é possível também encontrar a função de verossimilhança não-condicional a Z . Faremos nessa seção a integração via transformada de Laplace para encontrar a função de confiabilidade, de intensidade

e de intensidade não-condicional a Z , o que não foi encontrado diretamente com a integração convencional.

Para realizar a integração via transformada de Laplace, basta obter a função de confiabilidade não-condicional à variável de fragilidade, ou seja, é necessário integrá-la em relação ao termo de fragilidade (2.33). Considerando que $f(z)$ tenha distribuição Gama $(1/\alpha, 1/\alpha)$, então basta utilizar a transformada de Laplace da distribuição Gama $(1/\alpha, 1/\alpha)$, que é dada por

$$Q(s) = (1 + s\alpha)^{-\alpha^{-1}}. \quad (4.7)$$

Então, substituindo $s = \Lambda_0(t)$ em (4.7), teremos a função de confiabilidade não-condicional à variável de fragilidade, dada por

$$R(\tau) = (1 + \Lambda_0(\tau)\alpha)^{-\alpha^{-1}}, \quad (4.8)$$

e a partir de (2.37), encontramos também a função de intensidade acumulada não-condicional à variável de fragilidade, dada por

$$\Lambda(\tau) = -\log \left[(1 + \Lambda_0(\tau)\alpha)^{-\alpha^{-1}} \right], \quad (4.9)$$

chegando assim na função de intensidade não-condicional à variável de fragilidade:

$$\lambda(t) = \Lambda'(\tau) = \lambda_0(t) (1 + \Lambda_0(\tau)\alpha)^{-1}. \quad (4.10)$$

Dessa forma, sendo $t_{i,j}$ o j -ésimo tempo de falha do i -ésimo sistema, a função de verossimilhança é determinada por

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \{ \lambda_0(t_{i,j}) \} [(1 + \Lambda_0(\tau)\alpha)^{-(n_i + \alpha^{-1})}] \right\} \quad (4.11)$$

$$= \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \right\} \left[1 + \left(\frac{\tau}{\eta} \right)^\beta \alpha \right]^{-(n_i + \alpha^{-1})} \right\}. \quad (4.12)$$

O logaritmo da função de verossimilhança (4.12) é dado por

$$\begin{aligned} \ell(\boldsymbol{\mu}|\mathbf{t}) &= N \log(\beta) - N\beta \log(\eta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) \\ &\quad - \log \left[1 + \left(\frac{\tau}{\eta} \right)^\beta \alpha \right] (N + k\alpha^{-1}). \end{aligned} \quad (4.13)$$

Como não é possível obter uma solução analítica, os métodos numéricos são considerados para estimação dos parâmetros.

Considerando a reparametrização do PLP apresentada na Seção 2.2.3.1, o modelo para múltiplos sistemas reparáveis com fragilidade gama tem função de intensidade de falha e função de intensidade de falha acumulada, que são dadas, respectivamente, por

$$\lambda(t_{i,j}|z_i) = \frac{z_i \beta t_{i,j}^{\beta-1} \delta}{\tau \beta} \quad (4.14)$$

e

$$\Lambda(\tau|z_i) = z_i \delta. \quad (4.15)$$

A partir da função de verossimilhança dada em 4.11, basta substituir $\lambda(t)$ e $\Lambda(t)$ pelas funções utilizando a reparametrização. Assim, obtemos a seguinte função de verossimilhança

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ \frac{\beta t_{i,j}^{\beta-1} \delta}{\tau^\beta} \right\} (1 + \delta \alpha)^{-(n_i + \alpha^{-1})} \right\} \quad (4.16)$$

e o logaritmo da função de verossimilhança, dado por

$$\begin{aligned} \ell(\boldsymbol{\mu}|\mathbf{t}) &= N \log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) + N \log(\delta) - \beta N \log(\tau) \\ &\quad - (N + k\alpha^{-1}) \log(1 + \delta \alpha). \end{aligned} \quad (4.17)$$

Com a função de log-verossimilhança dada em 5.17, podemos obter analiticamente apenas o estimador do parâmetro β , que é exatamente o EMV de β no modelo de RM, ou seja,

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{i,j})}. \quad (4.18)$$

Para os outros parâmetros, novamente não é possível obter uma solução analítica, sendo necessário o uso dos métodos numéricos para estimação de parâmetros.

4.2 Estudo de simulação

Nessa seção apresentamos os procedimentos utilizados para o estudo de simulação para o modelo de múltiplos sistemas reparáveis com um termo de fragilidade, com função de intensidade dada em 4.14, a fim de verificar as propriedades frequentistas desse modelo. Mais especificamente, avaliaremos o impacto do tamanho da amostra sobre as propriedades dos estimadores através de métricas como vício, erro quadrático médio e probabilidade de cobertura.

Neste estudo, os diferentes cenários escolhidos são para avaliar os seguintes aspectos: tamanhos de amostra (entenda tamanho da amostra como número de sistemas) pequenos e grandes, função de intensidade crescente e decrescente e, principalmente, como os estimadores se comportam para sistemas pouco ou muito heterogêneos. Assim, consideramos amostras de diferentes tamanhos: $k = 5$, $k = 10$ e $k = 30$ sistemas. Em todos os casos, assumimos que os sistemas são observados em um período de tempo fixado, ou seja, são truncados em um tempo τ , em que consideramos $\tau = 50$. Fixamos o parâmetro de escala do PLP em $\delta = 5$ e utilizamos dois cenários para o parâmetro de forma β , que são os seguintes:

- $\beta = 3$: Considerando que o sistema está se deteriorando (função de intensidade crescente);

- $\beta = 0,5$: Considerando que o sistema está melhorando (função de intensidade decrescente).

Para todos os cenários, fixamos o parâmetro α , que representa a variância da variável de fragilidade em três casos: $\alpha = 0,5$, $\alpha = 1$ e $\alpha = 3$, representando baixa, média e alta variância, respectivamente. Lembrando que, quanto maior for a variância α , mais heterogêneos são os sistemas. Em cada cenário, foram geradas $d = 50.000$ réplicas.

Considerando que as falhas seguem um PPNH e as fragilidades $z_i \sim \text{Gamma}(1/\alpha, 1/\alpha)$, baseado em [Rigdon e Basu \(2000\)](#), seguimos o seguinte algoritmo de simulação de dados:

1. Fixamos os valores dos parâmetros;
2. Para cada sistema, geramos $z_i \sim \text{Gamma}(1/\alpha, 1/\alpha)$;
3. Para cada sistema, geramos o número de falhas $n_i \sim \text{Poisson}(z_i\delta)$, $i = 1, \dots, k$;
4. No i -ésimo sistema, os tempos de falha $t_{i1}, t_{i2}, \dots, t_{in_i}$ foram gerados através da inversa da função de intensidade, ou seja, $t_{ij} = \tau U_{ij}^{1/\beta}$, em que U_{ij} são números aleatórios da distribuição Uniforme(0,1);
5. Repetimos os itens anteriores d vezes.

A comparação entre os diferentes cenários é realizada usando o vício e o EQM para cada um dos parâmetros, que são calculados por:

$$\text{Vício}(\theta) = \sum_{i=1}^d \frac{\hat{\theta}_i - \theta}{d} \quad \text{e} \quad \text{EQM}(\theta) = \sum_{i=1}^d \frac{(\hat{\theta}_i - \theta)^2}{d}, \quad (4.19)$$

em que θ se refere aos parâmetros β , δ e α a serem estimados, $d = 50.000$ é o número de réplicas consideradas.

Com base nas propriedades frequentistas, espera-se que quanto maior for o tamanho da amostra, mais próximos de zero devem ser o vício e o EQM. As PC dos parâmetros também são calculadas assumindo um nível de confiança de 95%, e devem incluir o valor verdadeiro com a proporção de 0,95 das d réplicas. Portanto, a PC é uma medida para avaliar a qualidade dos estimadores intervalares. Consideramos as variâncias assintóticas obtidas a partir da matriz de informação de Fisher observada para construir os intervalos. O *software* R foi usado para calcular os resultados.

A Tabela 7 apresenta o vício, o EQM e a PC para o cenário em que o parâmetro α é fixado em 0,5, 1 e 3, para os diferentes tamanhos amostrais, $k = 5, 10, 30$, e para os parâmetros $\beta = 3$ e $\delta = 5$. Encontramos a representação gráfica desses resultados na Figura 18.

Tabela 7 – Cenário com $\beta = 3$ e $\delta = 5$.

Parâmetro α	k sistemas	Parâmetro	Vício	EQM	PC (95%)
0,5	5	β	0,011	0,073	0,945
		δ	0,042	0,519	0,939
		α	0,118	0,978	0,939
	10	β	-0,006	0,031	0,948
		δ	0,027	0,376	0,943
		α	0,053	0,594	0,940
	30	β	0,002	0,011	0,949
		δ	-0,005	0,063	0,947
		α	0,021	0,133	0,945
1	5	β	0,019	0,077	0,943
		δ	0,053	0,601	0,939
		α	0,168	0,931	0,938
	10	β	0,012	0,055	0,948
		δ	0,041	0,482	0,945
		α	0,059	0,868	0,939
	30	β	-0,009	0,017	0,951
		δ	0,012	0,059	0,946
		α	0,040	0,371	0,943
3	5	β	0,033	0,082	0,938
		δ	0,049	0,803	0,934
		α	0,178	1,115	0,937
	10	β	0,021	0,074	0,947
		δ	0,039	0,567	0,941
		α	0,114	0,931	0,941
	30	β	0,008	0,029	0,948
		δ	0,016	0,061	0,944
		α	0,071	0,492	0,942

A partir da Tabela 7 e da Figura 18, podemos perceber que, em geral, quanto maior o valor da variância α , maiores são os valores do vício e do EQM para os três parâmetros e, portanto, menos precisas são as suas estimativas, fato esse que pode ser explicado, em parte, devido à grandeza desse parâmetro. Entre os parâmetros, podemos notar que α apresenta os maiores vícios e EQM, e notamos também que α é o parâmetro que mais demora para alcançar a probabilidade nominal de cobertura. Enquanto que o parâmetro que atinge mais rápido a cobertura nominal é o parâmetro β , que é também o parâmetro mais preciso. Também notamos que, para todos os parâmetros, o vício e o EQM vão se aproximando de zero conforme o tamanho da amostra aumenta e a PC vai se aproximando de 0,95 conforme o tamanho da amostra aumenta.

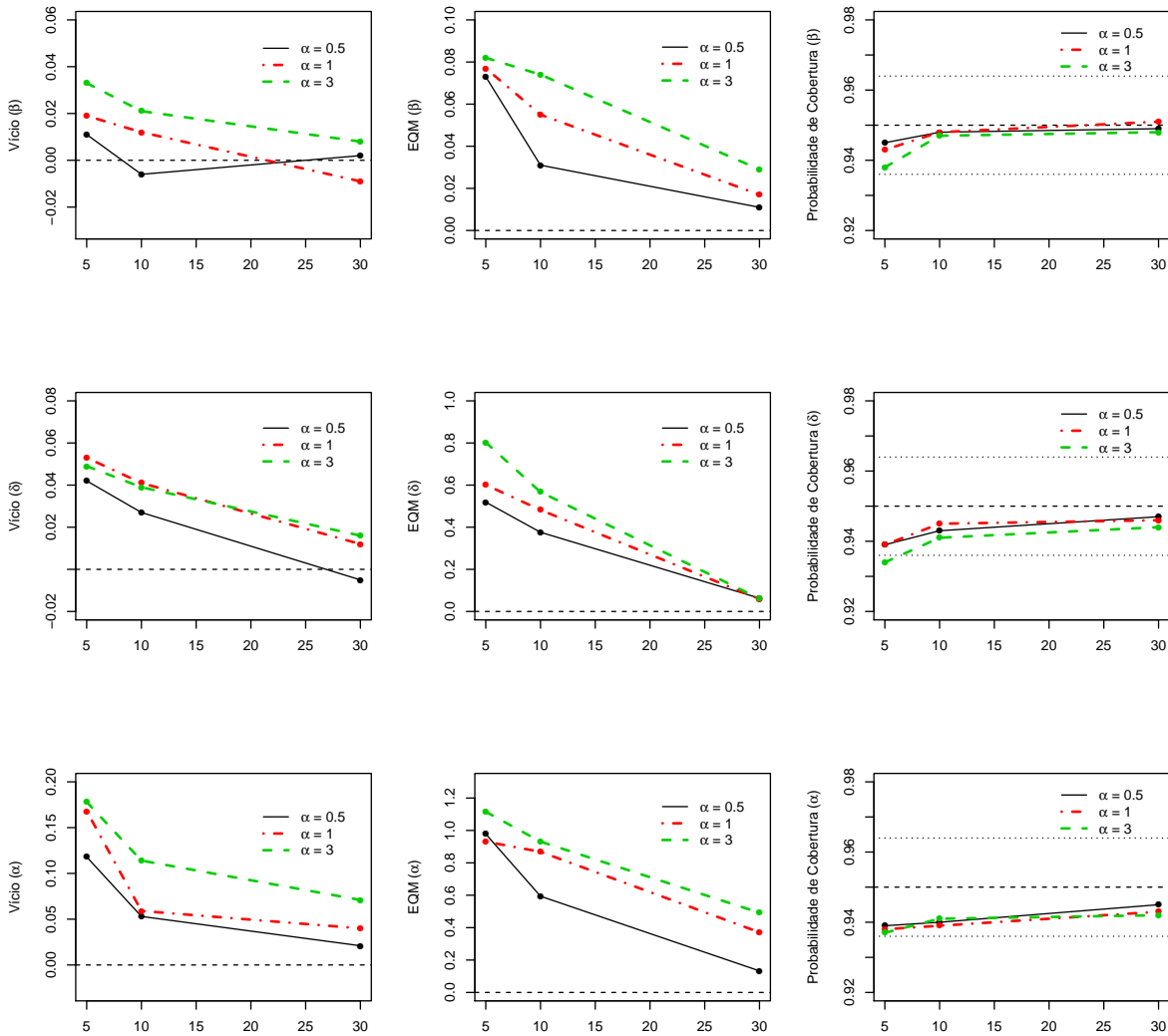


Figura 18 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$.

Na Tabela 8 também encontramos o vício, o EQM e a PC para as diferentes configurações do parâmetro α fixado, para os diferentes tamanhos amostrais, $k = 5, 10, 30$, mas agora para os parâmetros $\beta = 0,5$ e $\delta = 5$. Encontramos a representação gráfica desses resultados na Figura 19.

Tabela 8 – Cenário com $\beta = 0.5$ e $\delta = 5$.

Parâmetro α	k sistemas	Parâmetro	Vício	EQM	PC
0.5	5	β	0.008	0.065	0.943
		δ	0.041	0.998	0.939
		α	0.107	0.856	0.938
	10	β	0.005	0.042	0.947
		δ	0.034	0.463	0.942
		α	0.049	0.809	0.941
	30	β	0.002	0.024	0.951
		δ	-0.006	0.071	0.948
		α	0.017	0.098	0.944
1	5	β	0.015	0.072	0.941
		δ	0.044	1.015	0.938
		α	0.157	0.957	0.937
	10	β	0.011	0.059	0.946
		δ	0.043	0.538	0.944
		α	0.051	0.767	0.939
	30	β	-0.007	0.031	0.949
		δ	0.011	0.048	0.947
		α	0.029	0.283	0.943
3	5	β	0.021	0.080	0.940
		δ	0.047	1.014	0.939
		α	0.184	1.136	0.937
	10	β	0.012	0.079	0.946
		δ	0.037	0.773	0.941
		α	0.121	0.996	0.940
	30	β	0.009	0.031	0.947
		δ	0.017	0.063	0.945
		α	0.059	0.381	0.942

Nessa configuração, a partir da Tabela 8 e da Figura 19, temos, ao contrário do caso anterior, sistemas que estão em processo de melhoria pois o parâmetro β fixado é menor do que 1. Para esse cenário chegamos às mesmas conclusões encontradas no cenário anterior, ou seja, o estimador mais preciso é o parâmetro β e o estimador menos preciso é o parâmetro α ; conforme o valor de α aumenta, as estimativas em geral ficam menos precisas e, conforme o tamanho da amostra aumenta, o vício e o EQM vão se aproximando de zero e a PC vai se aproximando de 0,95.

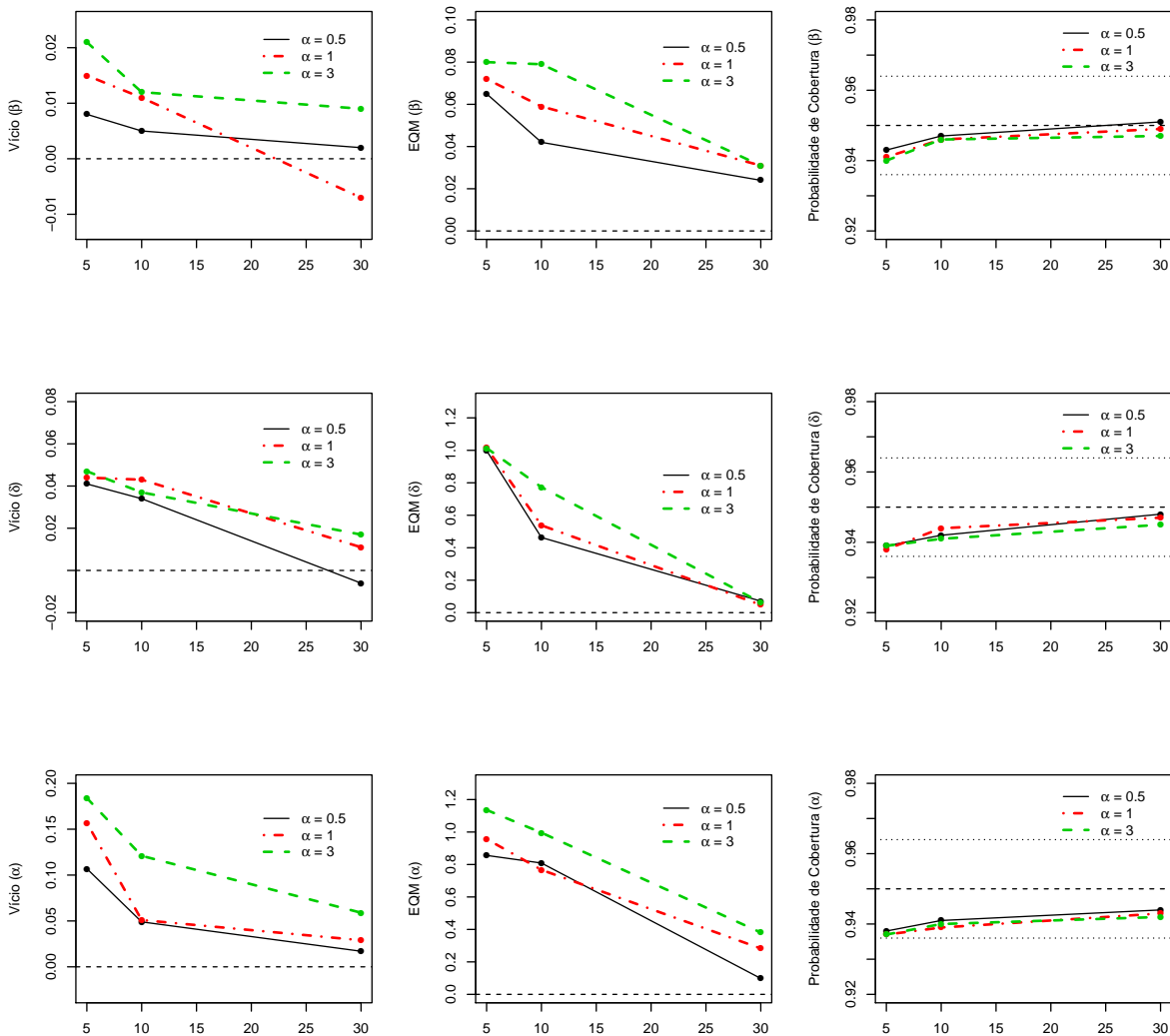


Figura 19 – Vício, EQM das estimativas com diferentes tamanhos amostrais para os valores de $\alpha = 0,5, 1$ e 3 considerando $\beta = 0,5$ e $\delta = 5$.

No geral, os estimadores retornaram estimativas próximas às verdadeiras para ambos os parâmetros e retornaram intervalos de credibilidade com PC próxima a 0,95 em todos os parâmetros, conforme o tamanho da amostra aumenta. Chegamos à conclusão que os estimadores apresentaram as propriedades frequentistas desejáveis.

4.3 Aplicação - Dados de colheitadeiras

Esses dados, descritos na Seção 1.2.1, já foram analisados na Seção 3.4, em que verificamos, através do gráfico *Duane plot*, que o PLP é adequado em cada uma das colheitadeiras de cana-de-açúcar. Observamos também, através das estimativas do modelo de RM, que temos um indicativo de que as colheitadeiras estão melhorando ao longo do tempo, ou seja, a função de

intensidade é decrescente, e espera-se que no período de 195 dias ocorram 12,5 falhas.

Nesta seção, a partir do modelo apresentado neste capítulo, o modelo de múltiplos sistemas reparáveis com um termo de fragilidade, iremos verificar se as três colheitadeiras de cana-de-açúcar analisadas são heterogêneas por efeito de fontes não-observadas. Utilizamos o *software* R e a função *optim* para obter as estimativas do modelo com função de intensidade dada em (4.14). As estimativas encontradas, bem como o erro padrão (EP) e o intervalo de 95% de confiança (IC (95%)) encontram-se na Tabela 9.

Tabela 9 – Estimativas dos parâmetros do modelo proposto

Parâmetros	EMV	EP	IC (95%)
β	0,889	0,144	(0,606 ; 1,172)
δ	12,232	2,355	(7,616 ; 16,849)
α	0,034	0,010	(0,015 ; 0,054)

Indo de encontro com os resultados da Seção 3.4, podemos notar com a Tabela 9 que os sistemas estão melhorando, ou seja, a confiabilidade está aumentando com o tempo, uma vez que $\hat{\beta} < 1$. Mas, ao analisar o intervalo de confiança do parâmetro β , vemos que o valor 1 pertence ao intervalo, indicando que a intensidade de falha das colheitadeiras pode ser aproximadamente constante ao longo do tempo. Vemos também, através da estimativa de δ que o número esperado de falhas em cada colheitadeira de cana-de-açúcar é de 12,2 falhas em 195 dias, resultado também muito parecido com o já encontrado.

O grande interesse dessa aplicação está em saber se as colheitadeiras são heterogêneas. Isso pode ser verificado a partir da estimativa de α , a variância da variável de fragilidade. De acordo com o resultado apresentado na Tabela 9, $\hat{\alpha} = 0,034$, observamos que a variância da fragilidade é pequena, porém significativa, como mostra o intervalo de confiança. Segue que as colheitadeiras são um pouco heterogêneas, nos levando a crer que existe algum fator externo que causa essa heterogeneidade, que pode ser estudado em pesquisas futuras.

Esse resultado encontrado está de acordo com a representação gráfica das falhas das colheitadeiras, encontrada na Figura 1 da Seção 1.2. Nesta figura, é possível perceber que o comportamento das falhas entre as três colheitadeiras é diferente, mas essa diferença não é muito grande, confirmando o resultado de que as colheitadeiras de cana-de-açúcar são pouco heterogêneas.

A função de intensidade de falha do modelo de fragilidade é dada em (4.10), então a função estimada para os dados das colheitadeiras é dada por

$$\hat{\lambda}(t_{i,j}) = \frac{\hat{\beta} t_{i,j}^{\hat{\beta}-1} \hat{\delta}}{\tau^{\hat{\beta}} (1 + \hat{\delta} \hat{\alpha})} = \frac{0,889 \cdot t_{i,j}^{0,889-1} \cdot 12,232}{195^{0,889} (1 + 12,232 \cdot 0,034)} = 0,0707 t_{i,j}^{-0,111}, \quad (4.20)$$

e, na Figura 21, temos sua representação gráfica. De acordo com os resultados analisados da Tabela 9, assim como o esperado, a função de intensidade de falha das colheitadeiras de cana-de-

açúcar é uma função decrescente, ou seja, a intensidade com que ocorrem as falhas diminui ao longo do tempo, indicando que a confiabilidade está aumentando com o tempo.

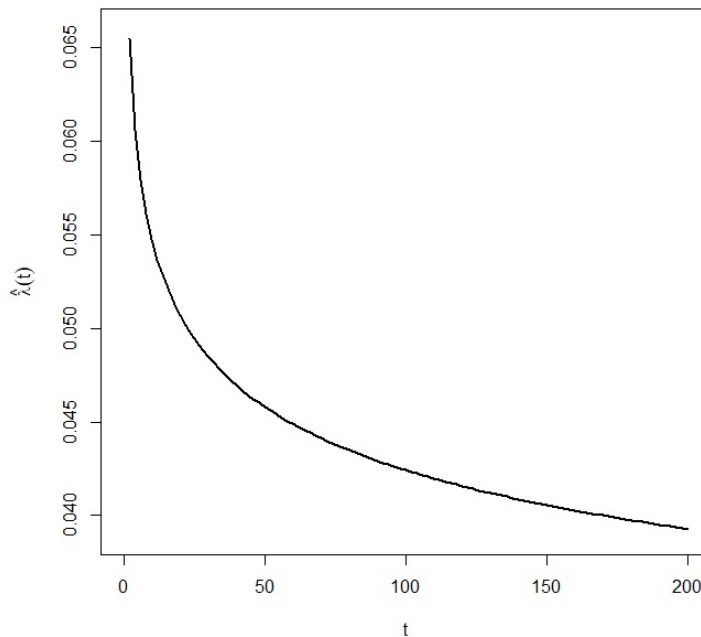


Figura 20 – Função de intensidade do modelo de fragilidade ajustado aos dados das colheitadeiras.

4.4 Aplicação - Dados de caminhões basculantes

Esses dados, descritos na Seção 1.2.2, também já foram analisados na Seção 3.5. O modelo PLP se mostrou adequado para cada um dos cinco caminhões, através do gráfico *Duane plot* encontrado na Figura 17. Através das estimativas do modelo de RM, ajustado na mesma referida seção, pudemos observar que há um indicativo de que os caminhões basculantes estão se deteriorando ao longo do tempo, ou seja, a função de intensidade é uma função crescente e espera-se que ocorram 25,8 falhas em cada caminhão basculante no período de 107 dias.

Através da aplicação do modelo de múltiplos sistemas reparáveis com um termo de fragilidade, queremos verificar se os caminhões basculantes são heterogêneos por efeito de fontes externas não-observadas. A partir da função *optim* do *software* R obtemos as estimativas do modelo com função de intensidade dada em (4.14). Os resultados encontrados nessa aplicação, ou seja, as estimativas encontradas, bem como o EP e o IC (95%), estão dispostos na Tabela 10.

Tabela 10 – Estimativas dos parâmetros para o modelo proposto

Parâmetros	EMV	DP	IC (95%)
β	1.115	0.098	(0.922 ; 1.307)
δ	24.602	2.622	(19.463 ; 29.742)
α	0.020	0.003	(0.014 ; 0.026)

Os resultados encontrados na Tabela 10 estão de acordo com a aplicação acima referida, ou seja, vemos que o número esperado de falhas é de 24,6 falhas em 107 dias, pois $\hat{\delta} = 24,602$ e, podemos notar que, como $\hat{\beta} = 1,115$, os sistemas estão se deteriorando. Porém, ao analisar o intervalo de confiança do parâmetro β , vemos que o valor 1 pertence ao intervalo, indicando que a intensidade de falha dos caminhões pode ser aproximadamente constante ao longo do tempo.

Como dito na aplicação anterior, o intuito dessa aplicação é verificar se existe alguma fonte de heterogeneidade não-observável, tornando os caminhões heterogêneos. Para tal, basta analisar a estimativa da variância da variável de fragilidade, dada por $\hat{\alpha}$. Através da Tabela 10, vemos que a estimativa da variância é de $\hat{\alpha} = 0,020$. Assim, vemos que essa variância é pequena e, através de seu intervalo de confiança, concluímos que existe uma pequena heterogeneidade entre os caminhões basculantes, ou seja, existem fontes não-observadas que causam essa heterogeneidade.

Assim como feito na aplicação das colheitadeiras, ao analisarmos a Figura 2 da Seção 1.2, que demonstra graficamente as falhas de cada caminhão, vemos que o comportamento das falhas entre os caminhões é um pouco diferente, sugerindo o que concluímos com a análise do parâmetro α , que os caminhões basculantes são pouco heterogêneos.

A função de intensidade de falha estimada pelo modelo de fragilidade para os dados dos caminhões é dada por

$$\hat{\lambda}(t_{i,j}) = \frac{\hat{\beta} t_{i,j}^{\hat{\beta}-1} \hat{\delta}}{\tau^{\hat{\beta}} (1 + \hat{\delta} \hat{\alpha})} = \frac{1,115 t_{i,j}^{1,115-1} 24,602}{107^{1,115} (1 + 24,602 \cdot 0,020)} = 0,1004 t_{i,j}^{0,115}, \quad (4.21)$$

e é apresentada na Figura 21. Confirmando os resultados analisados da Tabela 10, a função de intensidade de falha é uma função crescente, ou seja, a intensidade com que ocorrem as falhas nos caminhões basculantes aumenta ao longo do tempo.

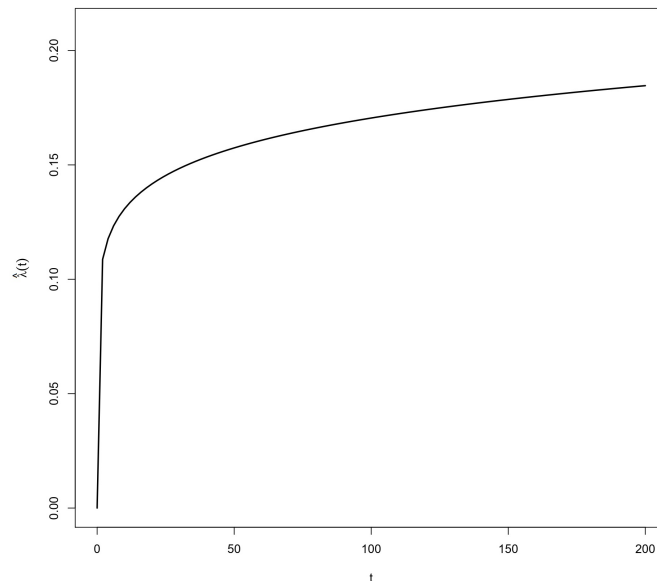


Figura 21 – Função de intensidade do modelo de fragilidade ajustado aos dados dos caminhões.

4.5 Considerações finais

Neste capítulo, focamos na análise de confiabilidade para múltiplos sistemas considerando que há um efeito aleatório não-observável causando heterogeneidade entre os sistemas. Sob o pressuposto de RM, assumimos que as sucessivas falhas seguem o PLP. O efeito aleatório não-observável foi modelado através de uma variável de fragilidade, introduzida de forma multiplicativa na função de intensidade de falha.

Apresentamos a modelagem supondo que a variável de fragilidade segue a distribuição gama, em que a vantagem dessa distribuição é sua tratabilidade algébrica. Para quantificar a heterogeneidade entre os sistemas, consideramos a variância da fragilidade gama. Para estimação dos parâmetros, consideramos o método de estimação de máxima verossimilhança. Realizamos um estudo de simulação, verificando as propriedades frequentistas, e os resultados estão de acordo com o esperado para EMVs. Duas aplicações foram feitas, em que foi possível observar através da estimação dos parâmetros, se os sistemas estão se deteriorando ou não, qual o número esperado de falhas dentro do período de estudo e foi possível quantificar a heterogeneidade entre os sistemas através da estimação da variância da variável de fragilidade.

UMA CLASSE GERAL DE MODELOS DE FRAGILIDADE EM SISTEMAS REPARÁVEIS

Neste capítulo propomos uma classe geral de modelos de fragilidade para análise de dados de sistemas reparáveis, baseada no modelo proposto por [Peña e Hollander \(2004\)](#). De acordo com os autores, esta classe de modelos incorpora simultaneamente: o efeito da intervenção após a ocorrência de um evento, se foi realizado um RM (processo de Poisson), RI ou RP (processo de renovação); o efeito de covariáveis ou variáveis concomitantes; o impacto das ocorrências de eventos acumulados e o efeito de variáveis latentes ou não-observadas que fornecem correlação entre os sistemas. Neste contexto, os modelos propostos por [Peña e Hollander \(2004\)](#) consideram uma abordagem não-paramétrica para a função de intensidade de base através de uma estrutura similar ao estimador de Nelson-Aalen (ou Aalen-Breslow). Na classe em que propomos, o efeito da variável latente será descrito pelo modelo de fragilidade multiplicativo gama e a função de intensidade de base induzida pelo PLP.

A maioria das extensões existentes do modelo de Cox lidam com boa parte desses efeitos. No entanto, esses modelos assumem que o efeito após cada reparo é sempre o mesmo (reparo mínimo ou perfeito). Em muitos estudos nem sempre o efeito da intervenção é o mesmo, neste caso, precisamos usar um modelo mais geral que nos permita incorporar diferentes efeitos que afetam as ocorrências de eventos. Então, esta classe pode modelar muitos tipos de dados em situações práticas. Segundo [Peña e Hollander \(2004\)](#), essa classe se faz necessária pois o efeito acumulativo do número de eventos pode acelerar ou desacelerar a ocorrência de novos eventos.

[González, Peña e Slate \(2005\)](#) demonstraram a aplicabilidade da classe geral na área de oncologia, em que acomodaram a heterogeneidade entre os pacientes, o efeito de covariáveis e o efeito do número de recaídas anteriores em recidivas de câncer. Segundo os autores, as ações de “reparo” são ações como quimioterapia, radioterapia, transplantes, em que acredita-se que cada ação desse diminua a recidiva de câncer. Portanto, o efeito do acúmulo dessas ações não pode ser desconsiderado em modelagens como essa, mas os modelos tradicionais de sobrevivência

não levam esse efeito em consideração.

A abordagem proposta por alguns autores, como em [Peña, Slate e González \(2007\)](#), propõe procedimentos de estimação para a classe geral estabelecida por [Peña e Hollander \(2004\)](#) de forma semiparamétrica, ou seja, a função de intensidade de falha é especificada de forma não-paramétrica. A modelagem de [Peña, Slate e González \(2007\)](#) foi aplicada a dados biomédicos e de saúde pública. Alguns outros autores, como em [Stocker IV e Peña \(2007\)](#), consideram a classe geral em que a função de intensidade de base do modelo é vista sob uma ótica totalmente paramétrica, porém sem considerar uma forma paramétrica específica. Isso é diferente do que é feito neste trabalho, em que consideramos o RM (ou o PPNH) sob a forma paramétrica do PLP, e assim teremos uma modelagem com uma maior aplicabilidade e de fácil interpretação.

[Gouvêa, Tomazella e Scalon \(2011\)](#) também consideram esta classe geral no contexto de eventos recorrentes aplicada a dados de insuficiência renal crônica, em que a função de intensidade assumida foi a distribuição Weibull e a variável de fragilidade foi assumida tendo distribuição Gama. Nestes trabalhos os autores consideraram métodos de estimação Bayesiana utilizando prioris informativas.

5.1 Formulação do modelo

Usaremos novamente a notação utilizada nos Capítulos 3 e 4 mas, por fins didáticos, ela será reafirmada aqui. Suponha que k sistemas são observados e cada um deles apresenta n_i falhas, seguidas de reparo ($i = 1, \dots, k$) e seja t_{ij} o tempo da j -ésima falha do i -ésimo sistema observado no período de tempo $(0, \tau]$, em que $j = 1, \dots, n_i$, de tal forma que $0 < t_{i1} < t_{i2} < \dots < t_{in_i} < \tau$. Suponha também que, para o i -ésimo sistema, possa ser observado o vetor de covariáveis $\mathbf{x}_i(t_{ij}) = (x_{i1}(t_{ij}), x_{i2}(t_{ij}), \dots, x_{ip}(t_{ij}))^\top$. A partir dessas definições, temos o processo de contagem definido por [Peña e Hollander \(2004\)](#) associado ao número acumulado de falhas gerado pelo processo $\{N_i(t_{ij}), t_{ij} \geq 0\}$, em que $N_i(t_{ij})$ é o número de falhas ocorridas no intervalo $[0, t_{ij}]$. Esse processo é contínuo à direita, em que t_{ij+} é um tempo infinitesimalmente maior que t_{ij} e t_{ij-} é um tempo infinitesimalmente menor que t_{ij} , assim $N_i(t_{ij}) = N_i(t_{ij+})$ e $N_i(t_{ij-})$ é o número de falhas que ocorreram antes de t_{ij} .

Para definirmos a classe geral, também precisamos definir a função $\varepsilon_i(t_{ij})$ que representa a idade virtual (ou idade efetiva) do sistema. Essa função é passiva de observação e, se depende de outros tempos, só dependerá dos tempos de falha anteriores a t_{ij} , portanto é uma função observável e previsível. O conceito de idade virtual foi introduzido por [Kijima \(1989\)](#), em que a ideia é distinguir entre a idade real do sistema, que é o tempo decorrido desde que o sistema era novo, geralmente no tempo $t = 0$, e a idade virtual do sistema, que descreve sua condição presente quando comparada a um novo sistema. A idade virtual é redefinida em cada falha de acordo com o tipo de reparo realizado e corre junto com o tempo real entre reparos, ou seja, ela se modifica de forma dinâmica conforme os reparos são realizados. Por exemplo,

se o reparo for mínimo (descrito por um processo de Poisson), a idade virtual é igual à idade observada do sistema, visto que o RM mantém o sistema na mesma condição logo antes da falha, e assim $\varepsilon_i(t_{ij}) = t_{ij}$. Se for realizado um RP (conforme o processo de renovação), a idade virtual é diferente da idade observada, pois o RP deixa o sistema em uma condição de novo, assim $\varepsilon_i(t_{ij}) = t_{ij} - t_{i, N_i(t_{ij}-)} = t_{ij} - t_{i, j-1}$. A idade virtual pode assumir ainda outras formas se for realizado um RI. A função $\varepsilon_i(t_{ij})$ deve ser não-negativa, ou seja, $\varepsilon_i(t_{ij}) \geq 0$ e deve ser monótona e quase certamente diferenciável no intervalo $(t_{i, j-1}, t_{ij}]$, com a primeira derivada $\varepsilon_i'(t_{ij})$ positiva.

Assim, a função de intensidade da Classe Geral de Modelos de Fragilidade com efeito da idade virtual (CGMF), proposta por [Peña e Hollander \(2004\)](#), pode ser obtida quando a função de intensidade condicional do processo é dada por

$$\lambda(t_{ij}|N_i(t_{ij}-), \mathbf{x}_i(t_{ij}), z_i) = z_i \lambda_0[\varepsilon_i(t_{ij})] \rho[N_i(t_{ij}-)|\gamma] \psi[\mathbf{x}_i^\top(t_{ij})\boldsymbol{\zeta}], \quad (5.1)$$

em que $\lambda_0(\cdot)$ é a função de intensidade de base, podendo assumir forma paramétrica ou não-paramétrica; z é a variável de fragilidade introduzida na função de intensidade para modelar a dependência entre as ocorrências dos reparos e a heterogeneidade entre as unidades. Neste contexto podemos observar que se $z_i > 1$ o i -ésimo sistema é mais frágil, pois sua função de intensidade de falha é aumentada e se $z_i < 1$ o i -ésimo sistema é menos frágil, pois sua função de intensidade é diminuída; $\varepsilon_i(t_{ij}) : 0 \leq t_{ij} \leq \tau, i = 1, 2, \dots, k$ é a idade virtual, que mede o efeito da performance dos reparos realizados; $\rho(\cdot|\gamma)$ é uma função não-decrescente ou não-crescente que depende de um único parâmetro, deve satisfazer $\rho[0|\gamma] = 1$ e mede o efeito do acúmulo de ocorrências de reparos; e $\psi(\mathbf{x}_i^\top(t_{ij})\boldsymbol{\zeta})$ é uma função de ligação não-negativa que mede os efeitos das covariáveis e deve ser especificada de tal forma que $\psi(0) = 1$ (consideraremos aqui a função exponencial, ou seja, $\psi(\mathbf{x}_i^\top(t_{ij})\boldsymbol{\zeta}) = \exp(\mathbf{x}_i^\top(t_{ij})\boldsymbol{\zeta})$).

Uma forma usual para $\rho[N_i(t_{ij}-)|\gamma]$ é $\gamma^{N_i(t_{ij}-)}$, em que $N_i(t_{ij}-)$ é o número de falhas que ocorreram antes de t_{ij} . Neste caso, se $\gamma > 1$, existe uma indicação de que a ocorrência de uma falha aumenta o risco futuro de falha; se $\gamma < 1$, cada ocorrência de uma falha diminui o risco de uma falha futura e, se $\gamma = 1$, a ocorrência da falha não afeta o risco de uma falha futura.

Segundo [Peña e Hollander \(2004\)](#), este modelo não é identificável com respeito à escala. Para eliminar este problema de identificabilidade, a distribuição de fragilidade precisa ter uma forma restrita, como a distribuição gama com o parâmetro de forma e de escala sendo iguais para que seu valor esperado seja igual a 1.

O modelo dado em (5.1) apresenta diversos casos particulares. Por exemplo, se não for de interesse verificar o efeito do acúmulo dos reparos, então $\rho[N_i(t_{ij}-)|\gamma] = 1$ e o modelo (5.1) resulta em um modelo de fragilidade multiplicativo, dado por

$$\lambda(t_{ij}|\mathbf{x}_i(t_{ij}), z_i) = z_i \lambda_0[\varepsilon_i(t_{ij})] \psi[\mathbf{x}_i^\top(t_{ij})\boldsymbol{\zeta}], \quad (5.2)$$

que considera o termo da fragilidade, covariáveis observadas e a idade virtual $\varepsilon_i(t)$. Este modelo não considera o efeito de acúmulo de reparos, portanto é indicado para, por exemplo, dados que

em cada sistema apresente um número muito pequeno de falhas. Quando não há a presença de covariáveis observadas e o reparo realizado é mínimo, temos exatamente o modelo proposto no Capítulo 4.

5.1.1 Modelo de classe geral sem fragilidade

A partir do modelo de classe geral (5.1), quando $Z = 1$, ou seja, quando não há a presença da variável de fragilidade na CGMF, temos um modelo com efeito de acumulação de reparos, efeito da idade virtual e efeito de covariáveis. A função de intensidade do modelo de classe geral sem fragilidade é dada por

$$\lambda(t_{ij}|\mathbf{X}_i) = \lambda_0[\varepsilon_i(t)]\rho[N_i(t_{ij}-)|\gamma]\exp(\boldsymbol{\zeta}'\mathbf{X}_i). \quad (5.3)$$

Este modelo pode ser aplicado em casos que acredita-se que não existam efeitos não-observáveis que atuam sobre os sistemas. Neste caso, os efeitos que estão sendo levados em consideração são: covariáveis observadas, o acúmulo dos reparos e a idade virtual. Por não ser um modelo de fragilidade, sua estimação fica simplificada e, portanto, pode ser facilmente aplicado em diversas situações práticas. E, embora não tenha o efeito de covariáveis não-observáveis, ainda temos um modelo bastante geral e que pode trazer muita informação sobre os sistemas.

O modelo (5.3), quando $\rho[N_i(t_{ij}-)|\gamma] = 1$, é dado por

$$\lambda(t_{ij}|\mathbf{X}_i) = \lambda_0[\varepsilon_i(t)]\exp(\boldsymbol{\zeta}'\mathbf{X}_i), \quad (5.4)$$

tornando-se um modelo de sistemas reparáveis com efeito de idade virtual e efeito de covariáveis observadas. Este modelo é quando não consideramos a variável de fragilidade nem o termo com o efeito de acúmulo de reparos. Poderia ser aplicado, por exemplo, quando há poucas falhas em cada sistema e quando não há motivos para acreditar que covariáveis não-observáveis causem efeito sobre os sistemas.

A seguir, apresentamos o método de estimação para a Classe Geral. Consideraremos a situação em que os reparos realizados são RM e que não há a presença de covariáveis, situação que se adequa perfeitamente nos conjuntos de dados da motivação do trabalho, encontrados na Seção 1.2. Apresentamos também o método de estimação para o modelo de classe geral sem fragilidade e fazemos uma comparação dos dois modelos na aplicação aos dados.

5.2 Modelo de classe geral sob reparo mínimo e fragilidade gama

Nesta seção propomos um método de estimação para os parâmetros da classe geral de modelos de fragilidade no caso particular em que são realizados RM e não há a presença de covariáveis.

Considerando o modelo (5.1) e que estamos em um caso em que a idade virtual é igual à idade real do sistema, ou seja, $\varepsilon_i(t_{ij}) = t$, então o efeito do reparo é RM, a forma paramétrica a ser utilizada é a do PLP (2.19), que os dados não têm a presença de covariáveis observadas e usaremos a forma usual $\rho[N_i(t_{ij}-)|\gamma] = \gamma^{N_i(t_{ij}-)}$ para medir o efeito dos acúmulos. Por simplicidade de notação, denotemos $g_{ij} = N_i(t_{ij}-)$, então $\rho[g_{ij}|\gamma] = \gamma^{g_{ij}}$. Sob estas condições, a função de intensidade do modelo é dada por

$$\lambda(t_{ij}|z_i) = z_i \lambda_0(t_{ij}) \gamma^{g_{ij}} = z_i \frac{\beta}{\eta} \left(\frac{t_{ij}}{\eta} \right)^{\beta-1} \gamma^{g_{ij}}, \quad (5.5)$$

e assim, temos a CGMF sob RM, ou seja, é um modelo para múltiplos sistemas reparáveis com um termo de fragilidade e com efeito da acumulação da ocorrência dos reparos em que são realizados RM após a ocorrência de cada falha. A função de confiabilidade referente ao modelo (5.5), condicionada à variável de fragilidade $Z = z_i$, é dada por

$$R(\tau|z_i) = [R(\tau)]^{z_i} = e^{[-\Lambda_0(\tau)z_i]} = \left\{ \exp \left[- \left(\frac{\tau}{\eta} \right)^{\beta} \gamma^{g_{ij}} \right] \right\}^{z_i}, \quad (5.6)$$

em que z_i representa a fragilidade do i -ésimo sistema e τ é o tempo de truncamento do estudo.

Consideramos, devido ao seu fácil tratamento algébrico, que a variável de fragilidade tem distribuição gama. A parametrização utilizada será Gama($1/\alpha, 1/\alpha$), em que $f(z_i)$ é a fdp da distribuição gama (4.2). Com essa parametrização, o valor esperado é $\mathbb{E}[Z] = 1$, evitando problemas de identificabilidade. A variância é $\text{Var}[Z] = \alpha$, assim α é o parâmetro que quantifica a heterogeneidade não-observada e a dependência entre as unidades em estudo, ou seja, quanto maior for a variância de Z maior será o grau de heterogeneidade entre os sistemas.

Para a construção da função de verossimilhança do modelo (5.5), assim como na Seção 3.1.1, precisamos fazer as seguintes suposições:

- Considere k sistemas reparáveis, para $k = 1, 2, \dots$, em que os sistemas são independentes;
- O i -ésimo sistema é truncado por tempo, sendo observado até um tempo pré-determinado τ , em que $0 < t_{i,1} < t_{i,2} < \dots < t_{i,n_i} < \tau$;
- n_i falhas são observadas no i -ésimo sistema, $i = 1, 2, \dots, k$;
- $N = \sum_{i=1}^k n_i$ é o número total de falhas observadas nos sistemas;
- Seja $t_{i,j}$, $i = (1, 2, \dots, k)$ e $j = (1, 2, \dots, n_i)$, observações da variável aleatória T que representa os tempos de falha para o i -ésimo sistema, registrados como os tempos desde o início do experimento ($t_{i,1} < t_{i,2} < \dots < t_{i,n_i}$);
- Seja $\boldsymbol{\mu} = (\beta, \eta, \alpha, \gamma)$ o vetor de parâmetros a ser estimado;
- g_{ij} é o número de falhas que ocorreram logo antes de t_{ij} .

Seguindo os mesmos passos para a construção da função de verossimilhança para o modelo de fragilidade (4.4), podemos reescrever a função de verossimilhança do modelo (5.5), condicionada à variável de fragilidade, como segue:

$$L(\boldsymbol{\mu}, z_i | \mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \{ z_i \lambda_0(t_{i,j}) \gamma^{g_{ij}} \} e^{-\Lambda_0(\tau) \gamma^{g_{ij} z_i} f(z_i)} \right\}. \quad (5.7)$$

Para determinar a função de verossimilhança não-condicional à variável de fragilidade, devemos integrá-la em relação a z_i . Como Z tem distribuição Gama($1/\alpha, 1/\alpha$), temos que a transformada de Laplace é $\mathcal{Q}(s) = (1 + s\alpha)^{-\alpha^{-1}}$. Fazendo $s = \Lambda_0(t)$, a função de confiabilidade não-condicional à variável de fragilidade pode ser encontrada através de (2.36), dada por

$$R(\tau) = (1 + \Lambda_0(\tau)\alpha)^{-\alpha^{-1}} = \left(1 + \left(\frac{\tau}{\eta} \right)^\beta \gamma^{g_{ij}} \alpha \right)^{-\alpha^{-1}}, \quad (5.8)$$

e a partir de (2.37), encontramos também a função de intensidade acumulada não-condicional à variável de fragilidade, dada por

$$\Lambda(\tau) = -\log \left[(1 + \Lambda_0(\tau)\alpha)^{-\alpha^{-1}} \right] = -\log \left[\left(1 + \left(\frac{\tau}{\eta} \right)^\beta \gamma^{g_{ij}} \alpha \right)^{-\alpha^{-1}} \right]. \quad (5.9)$$

Podemos, assim, chegar na função de intensidade não-condicional à variável de fragilidade:

$$\lambda(t) = \Lambda'(\tau) = \lambda_0(t) (1 + \Lambda_0(\tau)\alpha)^{-1} = \frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \gamma^{g_{ij}} \left(1 + \left(\frac{\tau}{\eta} \right)^\beta \gamma^{g_{ij}} \alpha \right)^{-1}. \quad (5.10)$$

Então, obtemos facilmente a função de verossimilhança não-condicionada à variável de fragilidade, que é dada por

$$L(\boldsymbol{\mu} | \mathbf{t}) = \prod_{i=1}^k \prod_{j=1}^{n_i} [\lambda_0(t_{i,j})] [(1 + \Lambda_0(\tau)\alpha)^{-(n_i + \alpha^{-1})}] \quad (5.11)$$

$$= \prod_{i=1}^k \prod_{j=1}^{n_i} \left[\frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta} \right)^{\beta-1} \gamma^{g_{ij}} \right] \left[1 + \left(\frac{\tau}{\eta} \right)^\beta \gamma^{g_{ij}} \alpha \right]^{-(n_i + \alpha^{-1})}. \quad (5.12)$$

O logaritmo da função de verossimilhança (5.12) é dado por

$$\begin{aligned} \ell(\boldsymbol{\mu} | \mathbf{t}) &= N \log(\beta) - N\beta \log(\eta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) + \sum_{i=1}^k \sum_{j=1}^{n_i} g_{ij} \log(\gamma) \\ &\quad - \sum_{i=1}^k (n_i + \alpha^{-1}) \log \left[1 + \left(\frac{\tau}{\eta} \right)^\beta \gamma^{g_{ij}} \alpha \right]. \end{aligned} \quad (5.13)$$

Como não é possível obter uma solução analítica, os métodos numéricos são considerados para estimação de parâmetros.

Considerando a reparametrização do PLP apresentada na Seção 2.2.3.1, a classe geral de modelos de fragilidade em múltiplos sistemas reparáveis, considerando fragilidade com a mesma distribuição Gama($1/\alpha, 1/\alpha$), tem função de intensidade de falha e função de intensidade de falha acumulada, que são dadas, respectivamente, por

$$\lambda(t_{i,j}|z_i) = \frac{z_i \beta t_{i,j}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^\beta} \quad (5.14)$$

e

$$\Lambda(\tau|z_i) = z_i \delta \gamma^{g_{ij}}. \quad (5.15)$$

A partir da função de verossimilhança dada em 5.11, basta substituir $\lambda(t)$ e $\Lambda(t)$ pelas funções usando o PLP reparametrizado. Assim, obtemos a seguinte função de verossimilhança:

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left\{ \frac{\beta t_{i,j}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^\beta} \right\} (1 + \delta \gamma^{g_{ij}} \alpha)^{-(n_i + \alpha^{-1})} \right\}, \quad (5.16)$$

e o logaritmo da função de verossimilhança, dado por

$$\begin{aligned} \ell(\boldsymbol{\mu}|\mathbf{t}) &= N \log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) + N \log(\delta) + \log(\gamma) \sum_{i=1}^k \sum_{j=1}^{n_i} g_{ij} - \beta N \log(\tau) \\ &- (N + k \alpha^{-1}) \log(1 + \delta \gamma^{g_{ij}} \alpha). \end{aligned} \quad (5.17)$$

Com a função de log-verossimilhança dada em 5.17, podemos obter analiticamente apenas o estimador do parâmetro β , que é exatamente o EMV de β no modelo de RM, ou seja,

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{i,j})}. \quad (5.18)$$

Para os outros parâmetros, novamente não é possível obter uma solução analítica, sendo necessário o uso dos métodos numéricos para estimação de parâmetros.

5.2.1 Modelo de classe geral sob reparo mínimo sem fragilidade

O modelo de classe geral sem fragilidade é um caso particular do modelo (5.5) quando não é levado em conta o efeito de covariáveis não-observáveis, ou seja, quando $Z = 1$; dessa forma, este é um modelo de classe geral sem fragilidade. Ele pode ser utilizado em situações que não tenham covariáveis observáveis, que possamos acreditar que não exista o efeito de covariáveis, e assim os sistemas não sejam considerados heterogêneos, e que o tipo de reparo realizado em todas as falhas de todos os sistemas não altere a idade virtual dos sistemas, isto é, os reparos são mínimos.

Novamente, supondo que temos múltiplos sistemas reparáveis, com RM realizados, sob forma paramétrica do PLP e, agora $Z = 1$, a partir de (5.5), podemos encontrar a função de intensidade do modelo de classe geral sem fragilidade, que é dada por

$$\lambda(t_{ij}) = \lambda_0(t_{ij})\gamma^{g_{ij}} = \frac{\beta}{\eta} \left(\frac{t_{ij}}{\eta}\right)^{\beta-1} \gamma^{g_{ij}}, \quad (5.19)$$

e a função de intensidade de falha acumulada é dada por

$$\Lambda(\tau) = \left(\frac{\tau}{\eta}\right)^{\beta} \gamma^{g_{ij}}. \quad (5.20)$$

Como não há a presença da variável de fragilidade, a função de verossimilhança é encontrada diretamente. Então, considerando k sistemas reparáveis, truncados no tempo τ e que o i -ésimo sistema tenha n_i falhas, ou seja, $0 < t_{i1} < t_{i2} \dots < t_{in_i} < \tau$ e o vetor de parâmetros $\boldsymbol{\mu} = (\beta, \eta, \gamma)$, a função de verossimilhança é dada por

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left[\frac{\beta}{\eta} \left(\frac{t_{i,j}}{\eta}\right)^{\beta-1} \gamma^{g_{ij}} \right] \exp\left(-\left(\frac{\tau}{\eta}\right)^{\beta} \gamma^{g_{ij}}\right) \right\}, \quad (5.21)$$

e a função log-verossimilhança é dada por

$$\begin{aligned} \ell(\boldsymbol{\mu}|\mathbf{t}) &= N \log(\beta) - \beta N \log(\eta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) \\ &+ \sum_{i=1}^k \sum_{j=1}^{n_i} g_{ij} \log \gamma - \left(\frac{\tau}{\eta}\right)^{\beta} \sum_{i=1}^k \sum_{j=1}^{n_i} \gamma^{g_{ij}}. \end{aligned} \quad (5.22)$$

Não é possível obter uma solução analítica, então métodos numéricos são considerados para estimação de parâmetros.

Considerando novamente a reparametrização do PLP apresentada na Seção 2.2.3.1, o modelo de classe geral sem fragilidade em múltiplos sistemas reparáveis, tem função de intensidade de falha e função de intensidade de falha acumulada, que são dadas, respectivamente, por

$$\lambda(t_{i,j}) = \frac{\beta t_{i,j}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^{\beta}} \quad (5.23)$$

e

$$\Lambda(\tau) = \delta \gamma^{g_{ij}}. \quad (5.24)$$

Assim, considerando o vetor de parâmetros $\boldsymbol{\mu} = (\beta, \delta, \gamma)$, a função de verossimilhança é dada por

$$L(\boldsymbol{\mu}|\mathbf{t}) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \left[\frac{\beta t_{i,j}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^{\beta}} \right] \exp(-\delta \gamma^{g_{ij}}) \right\}, \quad (5.25)$$

e a função log-verossimilhança é dada por

$$\begin{aligned} \ell(\boldsymbol{\mu}|\mathbf{t}) &= N\log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{i,j}) + N\log(\delta) + \sum_{i=1}^k \sum_{j=1}^{n_i} g_{ij} \log \gamma \\ &- \beta N\log(\tau) - \sum_{i=1}^k \delta \gamma^{g_{ij}}. \end{aligned} \quad (5.26)$$

A partir da função de log-verossimilhança dada em 5.26, podemos obter analiticamente apenas o estimador do parâmetro β , que é exatamente o EMV de β no modelo de RM, ou seja,

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{i,j})}. \quad (5.27)$$

Para os outros parâmetros, novamente não é possível obter uma solução analítica para a estimação de parâmetros, sendo necessário o uso dos métodos numéricos.

5.3 Estudo de simulação

Nesta seção apresentamos os procedimentos utilizados para o estudo de simulação para CGMF em sistemas reparáveis sob RM e fragilidade gama, com função de intensidade dada em 5.5, a fim de verificar as propriedades frequentistas desse modelo. Mais especificamente, avaliaremos o impacto do tamanho da amostra sobre as propriedades dos estimadores através de métricas como vício, EQM e PC.

Neste estudo, os diferentes cenários escolhidos são para avaliar os seguintes aspectos: tamanhos de amostra (entenda tamanho da amostra como número de sistemas) pequenos e grandes, para sistemas pouco ou muito heterogêneos e, principalmente, como os estimadores se comportam quando $\gamma > 1$, indicando que o acúmulo de falhas aumenta a função de intensidade de falha e quando $\gamma < 1$, indicando que o acúmulo de falhas diminui a intensidade de falha. Consideramos amostras de diferentes tamanhos: $k = 5$, $k = 10$ e $k = 30$ sistemas e assumimos que os sistemas são observados em um período de tempo fixado, ou seja, são truncados em um tempo τ , em que consideramos $\tau = 50$. Fixamos o parâmetro de escala do PLP em $\delta = 5$ e o parâmetro de forma em $\beta = 1,5$. Fixamos o parâmetro α , que representa a variância da variável de fragilidade, em três casos diferentes: $\alpha = 0,5$, $\alpha = 1$ e $\alpha = 3$, representando baixa, média e alta variância, respectivamente. Lembrando que quanto maior for a variância α , mais heterogêneos são os sistemas, ou seja, teremos três cenários diferentes: um com baixa heterogeneidade entre os sistemas; um com heterogeneidade média e outro com alta heterogeneidade. Por fim, consideramos dois diferentes cenários: $\gamma = 1,5$, situação em que o acúmulo de falhas aumenta a função de intensidade e $\gamma = 0,5$, situação em que o acúmulo de falhas diminui a função de intensidade. Em cada cenário, foram geradas $d = 50.000$ réplicas.

Considerando que as falhas seguem um PPNH e as fragilidades $z_i \sim \text{Gamma}(1/\alpha, 1/\alpha)$, seguimos o algoritmo abaixo de simulação de dados:

1. Fixamos os valores dos parâmetros;
2. Para cada sistema, geramos $z_i \sim \text{Gamma}(1/\alpha, 1/\alpha)$;
3. Para cada sistema, geramos o número de falhas $n_i \sim \text{Poisson}(z_i \delta \gamma^{g_{ij}})$, $i = 1, \dots, k$;
4. No i -ésimo sistema, os tempos de falha $t_{i1}, t_{i2}, \dots, t_{in_i}$ foram gerados através da inversa da função de intensidade, ou seja, $t_{ij} = \tau U_{ij}^{1/\beta}$, em que U_{ij} são números aleatórios da distribuição Uniforme(0,1);
5. Repetimos os itens anteriores d vezes.

Realizamos $d = 50.000$ réplicas, e a comparação entre os diferentes cenários é realizada usando o vício e o EQM, dados em (4.19). Com base nas propriedades frequentistas, espera-se que quanto maior for o tamanho da amostra, mais próximos de zero devem ser o vício e o EQM.

As PC dos parâmetros também são calculadas assumindo um nível de confiança de 95%, e devem incluir o valor verdadeiro com a proporção de 0,95 das d réplicas. Portanto, a PC é uma medida para avaliar a qualidade dos estimadores intervalares. Consideramos as variâncias assintóticas obtidas a partir da matriz de informação de Fisher observada para construir os intervalos. O *software* R foi usado para calcular os resultados.

A Tabela 11 apresenta o vício, o EQM e a PC para o cenário: $\beta = 1,5$, $\delta = 5$, $\gamma = 0,5$, diferentes valores de α fixados e diferentes tamanhos amostrais. Este cenário é quando os sistemas ficam com a função de intensidade diminuída com o acúmulo dos reparos. Encontramos a representação gráfica desses resultados na Figura 22.

Tabela 11 – Simulação com o cenário: $\beta = 1,5$, $\delta = 5$ e $\gamma = 0,5$.

Parâmetro α	k sistemas	Parâmetro	Vício	EQM	PC (95%)
0,5	5	β	0,012	0,024	0,942
		δ	0,039	0,417	0,939
		α	0,109	0,879	0,938
		γ	0,011	0,007	0,943
	10	β	0,009	0,020	0,944
		δ	0,035	0,341	0,942
		α	0,041	0,719	0,940
		γ	0,003	0,003	0,947
	30	β	0,003	0,008	0,951
		δ	0,009	0,071	0,948
		α	0,017	0,148	0,946
		γ	0,002	0,002	0,949
1	5	β	0,015	0,033	0,941
		δ	0,043	0,501	0,938
		α	0,108	0,914	0,938
		γ	0,008	0,009	0,940
	10	β	-0,011	0,029	0,942
		δ	0,037	0,407	0,941
		α	0,051	0,762	0,942
		γ	0,005	0,004	0,943
	30	β	0,005	0,011	0,948
		δ	-0,018	0,079	0,945
		α	0,025	0,211	0,943
		γ	0,004	0,002	0,950
3	5	β	0,021	0,057	0,940
		δ	0,053	0,534	0,933
		α	0,112	0,996	0,935
		γ	0,009	0,010	0,942
	10	β	0,013	0,042	0,941
		δ	0,041	0,438	0,939
		α	0,062	0,774	0,941
		γ	0,007	0,005	0,943
	30	β	0,008	0,021	0,947
		δ	-0,011	0,089	0,943
		α	0,029	0,397	0,942
		γ	0,004	0,003	0,946

Notamos que, de uma maneira geral, quanto maior o valor da variância α , maiores são os valores do vício e do EQM para os três parâmetros e, portanto, menos precisas são as suas estimativas. Entre os quatro parâmetros, podemos notar que o parâmetro α seguido do parâmetro δ apresenta os maiores vícios, EQM e são os parâmetros que mais demoram para alcançar a probabilidade de 0,95 de cobertura. Enquanto que o parâmetro que atinge mais rápido a cobertura nominal é o parâmetro γ , seguido do parâmetro β , que também são os parâmetros em que os estimadores apresentaram menores valores de vício e EQM. Também notamos que,

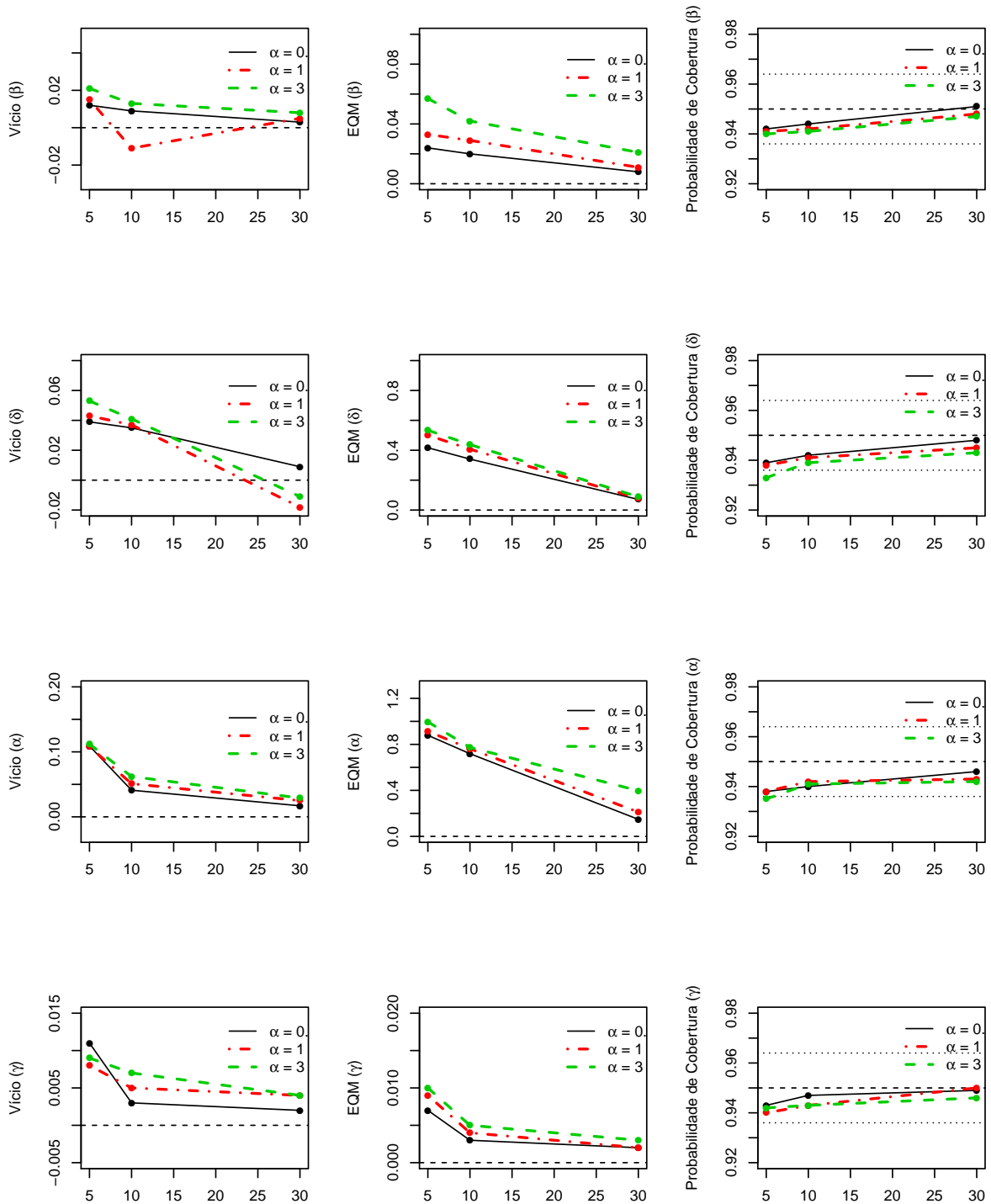


Figura 22 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais e para os valores de $\alpha = 0, 5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$ e $\gamma = 0, 5$.

para todos os parâmetros, o vício e o EQM vão se aproximando de zero conforme o tamanho da amostra aumenta e a PC vai se aproximando de 0,95 conforme o tamanho da amostra aumenta.

Na Tabela 12 também encontramos o vício, o EQM e a PC para as diferentes confi-

gurações do parâmetro α fixado e para os diferentes tamanhos amostrais, mas agora para o cenário: com $\beta = 1,5$, $\delta = 5$ e $\gamma = 1,5$, situação em que o acúmulo de falhas aumenta a função de intensidade de falha. A representação gráfica desses resultados encontra-se na Figura 23.

Tabela 12 – Simulação com o cenário: $\beta = 1,5$, $\delta = 5$ e $\gamma = 1,5$.

Parâmetro α	k sistemas	Parâmetro	Vício	EQM	PC
0,5	5	β	0,015	0,027	0,941
		δ	0,041	0,438	0,939
		α	0,101	0,864	0,938
		γ	0,015	0,011	0,941
	10	β	0,011	0,022	0,943
		δ	0,034	0,335	0,944
		α	0,051	0,801	0,939
		γ	0,005	0,007	0,945
	30	β	0,003	0,009	0,949
		δ	0,013	0,032	0,947
		α	0,011	0,052	0,942
		γ	0,004	0,002	0,950
1	5	β	0,017	0,031	0,940
		δ	0,045	0,497	0,939
		α	0,111	0,902	0,940
		γ	0,009	0,009	0,940
	10	β	0,013	0,027	0,943
		δ	0,041	0,391	0,940
		α	0,072	0,762	0,941
		γ	0,007	0,008	0,944
	30	β	0,006	0,011	0,947
		δ	0,022	0,087	0,943
		α	0,031	0,130	0,941
		γ	0,005	0,005	0,949
3	5	β	0,025	0,065	0,940
		δ	0,056	0,612	0,938
		α	0,118	0,998	0,937
		γ	0,010	0,011	0,940
	10	β	0,017	0,051	0,941
		δ	0,048	0,494	0,940
		α	0,071	0,918	0,939
		γ	0,008	0,009	0,941
	30	β	0,010	0,016	0,946
		δ	0,020	0,083	0,942
		α	0,033	0,348	0,942
		γ	0,006	0,006	0,943

Nesse cenário, com o acúmulo dos reparos, a função de intensidade é aumentada. Nessa configuração, chegamos às mesmas conclusões encontradas no cenário anterior, ou seja, os estimadores mais precisos são os dos parâmetros γ e β , e os estimadores menos precisos são os dos parâmetros α e δ ; conforme o valor de α aumenta, as estimativas em geral ficam menos

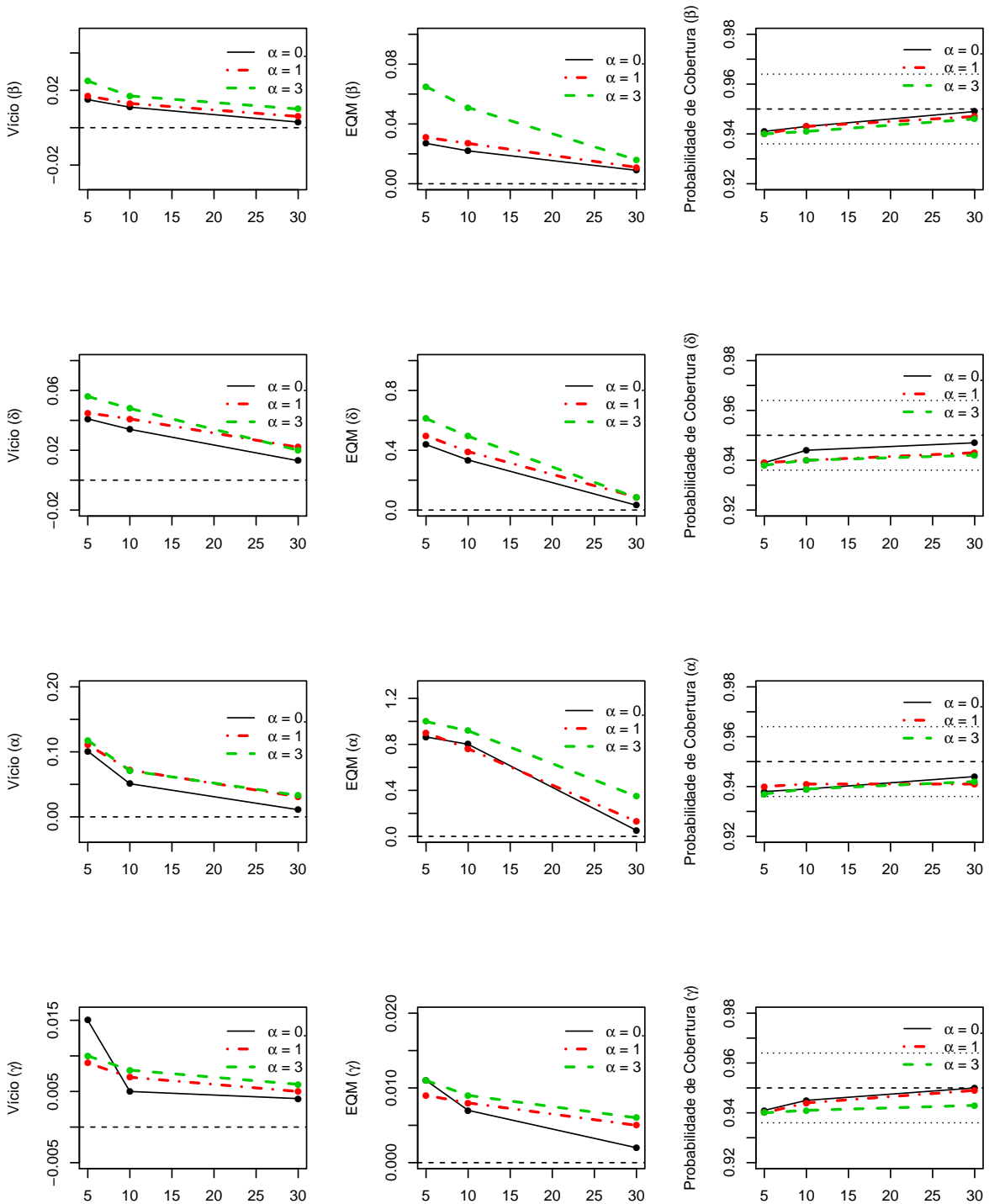


Figura 23 – Vício, EQM e PC das estimativas com diferentes tamanhos amostrais e para os valores de $\alpha = 0, 5, 1$ e 3 considerando $\beta = 3$ e $\delta = 5$ e $\gamma = 1, 5$.

precisas e, conforme o tamanho da amostra aumenta, o vício e o EQM vão se aproximando de zero e a PC vai se aproximando de 0,95. Podemos perceber também que, em geral, quando $\gamma = 1, 5$ as estimativas do parâmetro γ apresentam vício e EQM um pouco maiores.

No geral, os estimadores retornaram estimativas próximas às verdadeiras para ambos os parâmetros e retornaram intervalos de credibilidade com PC próxima à 0,95 em todos os parâmetros, conforme o tamanho da amostra aumenta. Chegamos à conclusão que os estimadores apresentaram as propriedades frequentistas desejáveis.

5.4 Aplicação - Dados de colheitadeiras

Os “dados de colheitadeiras” já foram analisados na Seção 3.4, em que verificamos que o modelo PLP é adequado em cada uma das colheitadeiras de cana-de-açúcar. Observamos também através do modelo de RM, um indicativo de que as colheitadeiras estão melhorando ao longo do tempo e espera-se que no período de 195 dias ocorram 12,5 falhas. Na Seção 4.3, a partir da estimativa da variância da variável de fragilidade, observamos que as colheitadeiras são pouco heterogêneas. Nesta seção, os modelos propostos (5.5) e (5.19) serão aplicados aos dados das colheitadeiras a fim de identificar se o acúmulo de reparos aumenta ou diminui a intensidade de falha das colheitadeiras.

Tabela 13 – Estimativas dos modelos para os dados das colheitadeiras.

Modelo	Parâmetros	EMV	EP	IC (95%)
Modelo de classe geral com fragilidade	β	0,861	0,140	(0,584 ; 1,137)
	δ	12,695	2,532	(7,732 ; 17,658)
	α	0,074	0,017	(0,041 ; 0,109)
	γ	1,019	0,072	(0,877 ; 1,161)
Modelo de classe geral sem fragilidade	β	0,880	0,142	(0,601 ; 1,160)
	δ	12,823	2,341	(8,235 ; 17,411)
	γ	1,011	0,120	(0,774 ; 1,247)

A Tabela 13 apresenta as estimativas para os parâmetros dos modelos, confirmando os resultados anteriormente encontrados. Através da estimativa de β , vemos que as colheitadeiras estão melhorando ao longo do tempo. Vemos também que espera-se que ocorram cerca de 12,5 falhas em 195 dias, através da estimativa de δ . A partir da estimativa de α , segue-se que a variância da variável de fragilidade é 0,074, indicando que as colheitadeiras são pouco heterogêneas.

Também, de acordo com a Tabela 13, em ambos os modelos a estimativa de γ é maior do que um, isso é um indicativo de que, à medida que aumenta a quantidade de falhas, aumenta o risco de uma nova falha. Porém, o intervalo de confiança do parâmetro γ nos dois modelos contém o número um, indicando que esse aumento pode não ser significativo.

É de interesse também verificar qual modelo se ajusta melhor aos dados: a classe geral de modelos de fragilidade, o modelo de classe geral sem fragilidade ou o modelo de fragilidade (aplicado na Seção 4.3). Para tal, utilizamos os critérios de seleção de modelos AIC, BIC e

log-verossimilhança. De acordo com esses critérios de seleção de modelos apresentados na Tabela 14, o modelo que se ajusta melhor aos dados é o modelo de classe geral sem fragilidade.

Tabela 14 – Critérios AIC, BIC e log-verossimilhança dos modelos para os dados das colheitadeiras.

Modelo	AIC	BIC	Log-verossimilhança
Modelo de classe geral com fragilidade	319,4	332,2	-155,7
Modelo de classe geral sem fragilidade	289,2	298,8	-141,6
Modelo usual de fragilidade	304,0	313,6	-149,0

Dessa forma, como o modelo escolhido é o modelo de classe geral sem fragilidade, a função de intensidade estimada para as colheitadeiras é dada por

$$\begin{aligned}\hat{\lambda}(t_{i,j}) &= \frac{\hat{\beta}t_{i,j}^{\hat{\beta}-1} \hat{\delta} \hat{\gamma}^{g_{ij}}}{\tau^{\hat{\beta}}} = \frac{0,88.t_{i,j}^{0,88-1} .12,823.1,011^{g_{ij}}}{195^{0,88}} \\ &= \frac{11,28.t_{i,j}^{-0,12} .1,011^{g_{ij}}}{103,57}.\end{aligned}\quad (5.28)$$

A qualidade do ajuste dos modelos foi avaliada comparando-se as estimativas da função de confiabilidade com o Kaplan-Meier (KM) empírico para eventos recorrentes de Wang e Chang (ver com mais detalhes no Anexo A). A Figura 24 mostra o KM de Wang Chang estimado versus a função de confiabilidade dos modelos ajustados.

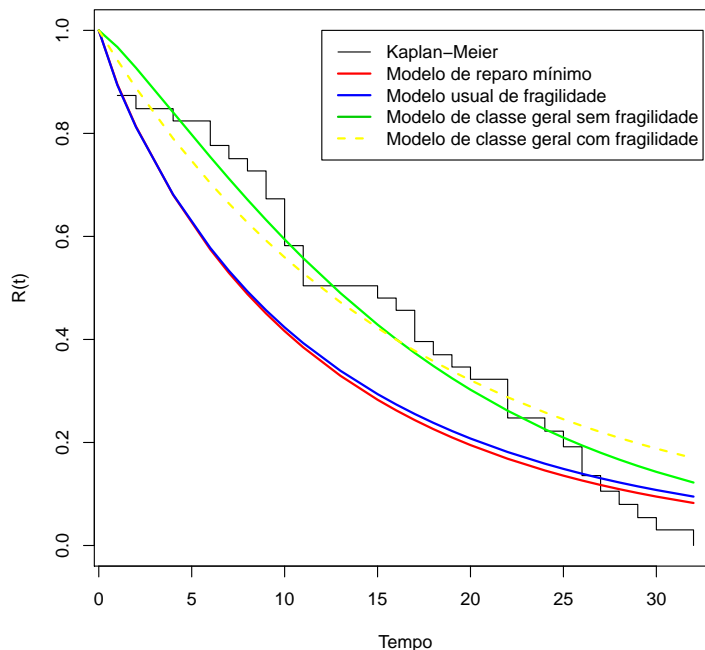


Figura 24 – Comparação do KM para eventos recorrentes com a função de confiabilidade estimada dos modelos para os dados de colheitadeiras de cana-de-açúcar.

Podemos notar que, a partir da Figura 24, a função de confiabilidade estimada do modelo de classe geral sem fragilidade segue melhor o decaimento da curva de KM, sugerindo que este

modelo é o que se ajusta melhor aos dados, confirmando a análise dos critérios de seleção de modelo.

5.5 Aplicação - Dados de caminhões basculantes

Nesta aplicação, queremos verificar se o histórico de reparos aumenta ou diminui o risco de um novo reparo e quantificar a heterogeneidade entre os caminhões basculantes do conjunto de dados de caminhões basculantes, já analisado por [Feitosa \(2015\)](#) e [D'Andrea et al. \(2017\)](#), que consideram apenas a heterogeneidade não-observada.

Já aplicamos esses dados na Seção 3.5, em que vimos que o modelo PLP se mostrou adequado para cada um dos cinco caminhões, através do gráfico *Duane plot* encontrado na Figura 17. Através das estimativas do modelo de RM, pudemos observar que há um indicativo de que os caminhões basculantes estão se deteriorando ao longo do tempo, ou seja, a função de intensidade é uma função crescente e espera-se que ocorram 25,8 falhas em cada caminhão basculante no período de 107 dias. Vimos também, na Seção 4.4, através do modelo de fragilidade, que os caminhões basculantes são pouco heterogêneos.

Utilizamos o *software* R para obter as estimativas dos parâmetros da classe geral de modelos de fragilidade sob RM e seu caso particular. Os resultados para ambos os modelos seguem na Tabela 15. A partir da estimativa de β , vemos novamente que o sistema está se deteriorando ao longo do tempo. Através da estimativa de δ , vemos que espera-se que ocorram aproximadamente 25 falhas em cada caminhão basculante no período de 107 dias. Com a estimativa do parâmetro α , vemos novamente que a variância da variável de fragilidade é pequena e, portanto, os caminhões são pouco heterogêneos. Também, de acordo com a Tabela 15, em ambos os modelos a estimativa de γ é maior do que um, isso é um indicativo de que, à medida que aumenta a quantidade de falhas, aumenta o risco de uma nova falha nos caminhões. Porém, o intervalo de confiança do parâmetro γ nos dois modelos contém o número um, indicando que esse aumento pode não ser significativo.

Tabela 15 – Estimativas dos modelos para os dados dos caminhões basculantes.

Modelo	Parâmetros	EMV	EP	IC (95%)
Modelo de classe geral com fragilidade	β	1,095	0,097	(0,906 ; 1,284)
	δ	25,108	3,077	(19,077 ; 31,139)
	α	0,058	0,007	(0,045 ; 0,071)
	γ	1,003	0,028	(0,947 ; 1,057)
Modelo de classe geral sem fragilidade	β	1,095	0,097	(0,906 ; 1,284)
	δ	24,602	2,622	(19,463 ; 29,742)
	γ	1,004	0,022	(0,960 ; 1,049)

A Tabela 16 apresenta os valores dos critérios AIC, BIC e log-verossimilhança de cada modelo, a fim de verificar qual modelo se ajusta melhor aos dados. De acordo com esses critérios,

temos indicação que o modelo que se ajusta melhor aos dados é o modelo de classe geral sem fragilidade.

Tabela 16 – Critérios AIC, BIC e log-verossimilhança dos modelos para os dados dos caminhões basculantes.

Modelo	AIC	BIC	Log-verossimilhança
Modelo de classe geral com fragilidade	767,6	779,0	-379,8
Modelo de classe geral sem fragilidade	629,7	638,2	-311,8
Modelo usual de fragilidade	677,5	686,1	-335,7

Dessa forma, como o modelo escolhido é o modelo de classe geral sem fragilidade, a função de intensidade estimada para as colheitadeiras é dada por

$$\begin{aligned}\hat{\lambda}(t_{i,j}) &= \frac{\hat{\beta}t_{i,j}^{\hat{\beta}-1} \hat{\delta} \hat{\gamma}^{g_{ij}}}{\tau \hat{\beta}} = \frac{1,095 \cdot t_{i,j}^{1,095-1} \cdot 24,602 \cdot 1,004^{g_{ij}}}{107^{1,095}} \\ &= \frac{26,9 \cdot t_{i,j}^{0,095} \cdot 1,011^{g_{ij}}}{166,8}.\end{aligned}\quad (5.29)$$

A Figura 25 mostra o KM de Wang Chang estimado versus a função de confiabilidade dos modelos ajustados. Podemos observar que a função de confiabilidade estimada do modelo de classe geral sem fragilidade segue melhor o decaimento da curva de KM, sugerindo que este modelo é o que se ajusta melhor aos dados, o que confirma o resultado encontrado com a análise dos critérios de seleção de modelo.

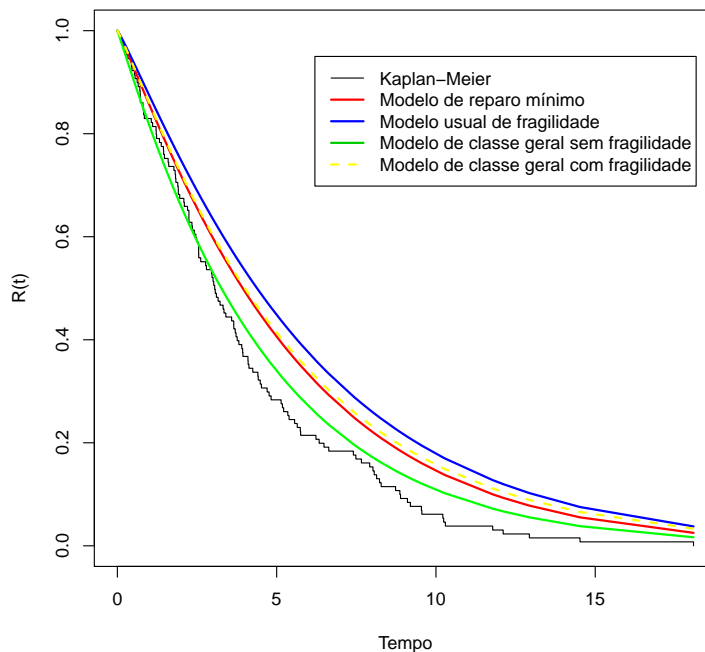


Figura 25 – Comparação do KM para eventos recorrentes com a função de confiabilidade estimada dos modelos para os dados de caminhões basculantes.

5.6 Considerações finais

Neste capítulo, apresentamos a classe geral de modelos de fragilidade para múltiplos sistemas reparáveis. Esta classe leva em consideração um efeito aleatório não-observável que causa heterogeneidade entre os sistemas, o efeito do acúmulo dos reparos, o tipo de reparo realizado e o efeito de covariáveis observadas. Propomos um método de estimação dos parâmetros para a situação em que são feitos RM, não há a presença de covariáveis, supondo que a fragilidade segue uma distribuição gama e usando o modelo PLP. Também, propomos um método de estimação dos parâmetros para o caso particular que não leva em conta o efeito da heterogeneidade não-observável.

Realizamos um estudo de simulação em que verificamos que as propriedades frequentistas são atendidas para a classe geral de modelos de fragilidade. Nessa simulação pudemos notar que maiores valores da variância da variável de fragilidade tornam as estimativas de todos os parâmetros um pouco menos precisas. Fizemos duas aplicações, com os conjuntos de dados das colheitadeiras de cana-de-açúcar e dos caminhões basculantes, em que foi possível mostrar a aplicabilidade dos modelos propostos. Nas duas aplicações, o modelo que melhor se ajustou aos dados foi o modelo de classe geral sem fragilidade.

MODELOS DE FRAGILIDADE NÃO-PARAMÉTRICA PARA MÚLTIPLOS SISTEMAS REPARÁVEIS

A variável de fragilidade é uma variável não-observável, portanto a escolha de sua distribuição é uma parte problemática nessa modelagem, comparável à escolha de distribuições à priori em métodos Bayesianos. Ademais, ao ajustar o modelo de fragilidade, é necessário fazer a integração em relação à variável de fragilidade e, em muitos casos, essa integral é feita numericamente, podendo causar problemas na estimação, principalmente em modelos muito sofisticados ou em grandes conjuntos de dados. A fim de contornar esses problemas, uma opção é o uso da abordagem não-paramétrica para a fragilidade, já que nela não se faz necessário supor alguma distribuição de probabilidade para a variável de fragilidade.

Outra vantagem do uso da fragilidade não-paramétrica é a capacidade de estimar as fragilidades individuais, muitas vezes não sendo feito em modelos paramétricos. Estimar as fragilidades não é útil somente na tarefa de conhecer quais sistemas são mais frágeis, mas também é possível, ao comparar as estimativas dos diferentes sistemas, ter indicações sobre covariáveis importantes que influenciam o processo de falha, como a idade do sistema no início da operação ou sua localização no momento do funcionamento.

Na literatura são bastante escassos modelos de fragilidade com abordagem não-paramétrica e, os que são encontrados, são da área de bioestatística. Podemos citar, por exemplo, [Manda \(2011\)](#), que apresenta uma abordagem Bayesiana para a estimação dos parâmetros. Recentemente, [Slimacek e Lindqvist \(2016\)](#) propuseram o modelo de fragilidade para o PPNH com uma abordagem não-paramétrica para a variável de fragilidade. Nesse artigo, a estimação da fragilidade individual de cada sistema é feita diretamente, sem supor alguma distribuição de probabilidade. Os autores consideraram o PPNH com a forma paramétrica do PLP, mas utilizaram uma parametrização que, embora seja algebricamente simples, não tem uma interpretação

prática para os parâmetros.

Neste capítulo iremos considerar a abordagem não-paramétrica para modelos de fragilidade para múltiplos sistemas reparáveis, proposta por [Slimacek e Lindqvist \(2016\)](#). Mas, diferentemente dos autores, vamos considerar o modelo reparametrizado do PLP por sua interpretação prática dos parâmetros, como discutido na Seção 2.2.3.1. A principal contribuição deste capítulo é que consideramos também a CGMF apresentada na Seção 5.2 sob a abordagem não-paramétrica das fragilidades individuais.

6.1 Múltiplos sistemas reparáveis com fragilidade não-paramétrica

Em sistemas reparáveis, é muito importante considerar a escolha de um modelo apropriado para o tempo de falha. Assim, caso acredita-se que haja uma fonte externa causando heterogeneidade entre os sistemas, essa heterogeneidade deve ser incluída no modelo; isso pode ser feito pelo modelo de fragilidade, ou seja, a incorporação de uma variável aleatória não-observável multiplicando a função de intensidade de falhas, assim como foi feito na Seção 4.1 utilizando o modelo de fragilidade gama.

Porém, diferentemente do Capítulo 4, a proposta é incluir a heterogeneidade não-observada no modelo considerando a fragilidade de forma não-paramétrica, como proposto por [Slimacek e Lindqvist \(2016\)](#), mas com a forma reparametrizada do PLP para que seja possível ter uma melhor interpretação. A vantagem de se usar a fragilidade não-paramétrica vem do fato de não precisar escolher um modelo paramétrico para a variável de fragilidade, diminuindo a chance de ter modelos mal ajustados. A segunda vantagem desse modelo é o fato de se conseguir estimar de forma direta as fragilidades individuais de cada sistema, conseguindo identificar quais sistemas são mais frágeis. Demonstraremos o uso deste modelo com aplicação a conjuntos de dados reais.

A fim de determinarmos o modelo de fragilidade não-paramétrica para dados de múltiplos sistemas reparáveis, considere $N(t)$ o número de falhas desde o início do monitoramento do sistema até um determinado momento t , considerando k PPNH sob a forma paramétrica do PLP reparametrizado dado em (2.23) com efeito aleatório não-observado, e considere a função de intensidade do modelo de fragilidade (2.30). Então, a função de intensidade de falhas do i -ésimo sistema, condicionada à variável de fragilidade $Z = z_i$, é dada por

$$\lambda(t_{ij}|z_i) = \frac{z_i \beta t_{ij}^{\beta-1} \delta}{\tau^\beta}, \quad (6.1)$$

em que δ é o número esperado de falhas antes do tempo de truncamento τ , β é uma medida de melhoria ou deterioração do sistema e z_i é a variável de fragilidade do i -ésimo sistema, para

$i = 1, \dots, k$. A função de intensidade acumulada desse modelo é dada por

$$\Lambda(\tau|z_i) = z_i \delta. \quad (6.2)$$

A interpretação da variável z_i é da mesma forma vista no Capítulo 4, ou seja, se $z_i > 1$, o i -ésimo sistema é mais frágil, pois aumenta a função de intensidade de falha e, se $z_i < 1$, o i -ésimo sistema é menos frágil, pois diminui a função de intensidade de falha.

Considere que n_i falhas são observadas no i -ésimo sistema ($N = \sum_{i=1}^k n_i$) e os tempos de falha dos sistemas são denotados por t_{ij} , em que $i = 1, \dots, k$ e $j = 1, \dots, n_i$. O i -ésimo sistema é observado até o tempo τ , isto é, os sistemas são truncados todos no mesmo tempo. Seja $\boldsymbol{\mu} = (\beta, \delta, z_i)$ o vetor de parâmetros a ser estimado, então a função de verossimilhança do i -ésimo sistema, condicional em z_i , é dada por

$$L(\boldsymbol{\mu}|z_i, \mathbf{t}) = \left(\prod_{j=1}^{n_i} \lambda(t_{ij}|z_i) \right) \exp(-\Lambda(\tau|z_i)) = \left(\prod_{j=1}^{n_i} \frac{z_i \beta t_{ij}^{\beta-1} \delta}{\tau^\beta} \right) \exp(-z_i \delta). \quad (6.3)$$

A função de log-verossimilhança para os k sistemas é

$$\begin{aligned} \ell(\boldsymbol{\mu}|z_i, \mathbf{t}) &= \sum_{i=1}^k n_i \log(z_i) + N \log(\beta) + N \log(\delta) - N \beta \log(\tau) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) \\ &\quad - \delta \sum_{i=1}^k z_i. \end{aligned} \quad (6.4)$$

Na abordagem de fragilidade não-paramétrica, assim como descrito em [Slimacek e Lindqvist \(2016\)](#), as fragilidades individuais não-observadas são consideradas diretamente parâmetros desconhecidos a serem estimados. Assim, para obter os estimadores dos parâmetros do modelo, encontramos primeiro as equações score, que são dadas por

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{N}{\beta} + \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) - N \log(\tau) = 0, \\ \frac{\partial \ell}{\partial \delta} &= \frac{N}{\delta} - \sum_{i=1}^k z_i = 0, \\ \frac{\partial \ell}{\partial z_i} &= \frac{n_i}{z_i} - \delta = 0, \quad i = 1, \dots, k. \end{aligned}$$

Sendo assim, pelo método de máxima verossimilhança, podemos obter diretamente o estimador de β , que é dado por

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{ij})}, \quad (6.5)$$

que é exatamente o EMV de β dado em (3.12) para o modelo de RM (3.8). Podemos notar que ao encontrar o estimador de δ , ele depende de z_i e ao encontrar o estimador de z_i , ele depende de δ . Precisamos, então, de algum artifício para encontrar os estimadores desses parâmetros.

Sabemos que a esperança condicional do número de falhas N dentro de um determinado tempo τ , com uma dada fragilidade Z em um sistema seguindo o PLP, é

$$\mathbb{E}[N|Z, \tau] = Z\delta, \quad (6.6)$$

e, através da suposição de $\mathbb{E}[Z] = 1$ que torna o modelo de fragilidade identificável, temos que

$$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|Z, \tau]] = \delta. \quad (6.7)$$

Assim, utilizando a ideia de [Slimacek e Lindqvist \(2016\)](#), a partir de 6.7 introduzimos um estimador $\bar{\delta}$ (baseado no método dos momentos), dado por

$$\bar{\delta} = \frac{1}{k} \sum_{i=1}^i n_i = \frac{N}{k}, \quad (6.8)$$

que coincide exatamente com o EMV de δ no modelo de RM. Utilizando a equação escore de z_i e substituindo δ pelo estimador $\bar{\delta}$, as fragilidades individuais z_i podem ser estimadas como

$$\hat{z}_i = \frac{n_i}{\bar{\delta}} = \frac{n_i}{\frac{N}{k}} = \frac{n_i k}{N}. \quad (6.9)$$

E assim, com essa abordagem conseguimos encontrar os estimadores dos parâmetros do modelo e das fragilidades de forma analítica, sendo essa mais uma vantagem dessa modelagem, o que permite a aplicação deste modelo de uma forma bastante simples na prática.

Caso seja de interesse estimar os parâmetros para o modelo de fragilidade com a seguinte parametrização do PLP:

$$\lambda(t_{ij}|z_i) = z_i a b t_{ij}^{b-1}, \quad (6.10)$$

os estimadores dos parâmetros podem ser obtidos de forma bastante simples, são encontrados em [Slimacek e Lindqvist \(2016\)](#).

6.2 Classe geral de modelos de fragilidade sob reparo mínimo e fragilidade não-paramétrica

A Seção 5.2 considerou o caso particular da CGMF em que são feitos RM, não há a presença de covariáveis observadas e a distribuição da variável de fragilidade considerada foi a distribuição gama. Nesta seção propomos a estimação dos parâmetros do modelo considerando a fragilidade de forma não-paramétrica. Ou seja, estudaremos uma extensão da modelagem apresentada na Seção 6.1, considerando também o efeito do acúmulo dos reparos.

Seja $N(t)$ o número de falhas desde o começo do estudo até um certo tempo t , considerando k independentes PPNH sob a forma paramétrica PLP (2.23), com a presença do efeito

de acumulação de ocorrências de eventos e da heterogeneidade não-observada; a função de intensidade de falha e a função de intensidade de falha acumulada do i -ésimo sistema são dadas, respectivamente, por

$$\lambda(t_{ij}|z_i) = \frac{z_i \beta t_{ij}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^\beta} \quad \text{e} \quad \Lambda(\tau|z_i) = z_i \delta \gamma^{g_{ij}}, \quad (6.11)$$

em que g_{ij} é o número de falhas logo antes do tempo t_{ij} , γ é o parâmetro que mede o efeito do acúmulo das ocorrências de reparo e z_i é a heterogeneidade não-observada do i -ésimo sistema, em que $i = 1, \dots, k$.

Considere novamente n_i o número de falhas do i -ésimo sistema e t_{ij} os tempos de falha ($i = 1, \dots, k$ e $j = 1, \dots, n_i$). O i -ésimo sistema é observado até o tempo τ , isto é, os sistemas são truncados por tempo. Seja $\boldsymbol{\mu} = (\beta, \delta, \gamma, z_i)$ o vetor de parâmetros a ser estimado, assim, a função de verossimilhança do i -ésimo sistema, condicional em z_i , é

$$L(\boldsymbol{\mu}|z_i, \mathbf{t}) = \left(\prod_{j=1}^{n_i} \frac{z_i \beta t_{ij}^{\beta-1} \delta \gamma^{g_{ij}}}{\tau^\beta} \right) \exp(-z_i \delta \gamma^{g_{ij}}). \quad (6.12)$$

E, então, podemos encontrar a função de log-verossimilhança para os k sistemas, que é dada por

$$\begin{aligned} \ell(\boldsymbol{\mu}|z_i, \mathbf{t}) &= \sum_{i=1}^k n_i \log(z_i) + N \log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) + N \log(\delta) + \log(\gamma) \sum_{i=1}^k n_i g_{ij} \\ &- N\beta \log(\tau) - \delta \sum_{i=1}^k z_i \gamma^{g_{ij}}. \end{aligned} \quad (6.13)$$

Considerando que a fragilidade não-observada dos sistemas de forma não-paramétrica, isto é, supondo que as fragilidades individuais são diretamente os parâmetros desconhecidos, as equações score são

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{N}{\beta} + \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) - N \log(\tau) = 0, \\ \frac{\partial \ell}{\partial \delta} &= \frac{N}{\delta} - \sum_{i=1}^k z_i \gamma^{g_{ij}} = 0, \\ \frac{\partial \ell}{\partial z_i} &= \frac{n_i}{z_i} - \delta \gamma^{g_{ij}} = 0, \quad i = 1, \dots, k, \\ \frac{\partial \ell}{\partial \gamma} &= \frac{\sum_{i=1}^k n_i g_{ij}}{\gamma} - \delta \sum_{i=1}^k z_i g_{ij} \gamma^{g_{ij}-1} = 0. \end{aligned}$$

Resolvendo a equação score de β , temos que o EMV de β é dado por

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k n_i \log(\tau) - \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij})}. \quad (6.14)$$

A partir da função escore de z_i , temos que as fragilidades individuais podem ser expressas por

$$\hat{z}_i = \frac{n_i}{\hat{\delta} \hat{\gamma}^{g_{ij}}}, \quad i = 1, \dots, k, \quad (6.15)$$

em que z_i depende dos estimadores dos parâmetros δ e γ .

Resolvendo a equação escore de δ , encontramos o EMV de δ dado por

$$\hat{\delta} = \frac{N}{\sum_{i=1}^k \hat{z}_i \hat{\gamma}^{g_{ij}}}, \quad (6.16)$$

mas em ambos os casos, os estimadores dependem dos outros parâmetros.

Ao utilizar \hat{z}_i ou $\hat{\delta}$ para encontrar o estimador de γ , temos um problema de identificabilidade. Então, vamos usar a mesma ideia de [Slimacek e Lindqvist \(2016\)](#), isto é, transformar o modelo original (6.11) para o modelo equivalente com novos parâmetros chamados de δ_i , dados por

$$\delta_i = z_i \delta, \quad i = 1, \dots, k, \quad (6.17)$$

e assim a função de intensidade e a função de intensidade acumulada desse modelo equivalente são, respectivamente, dadas por

$$\lambda(t_{ij}|z_i) = \frac{\delta_i \beta t_{ij}^{\beta-1} \gamma^{g_{ij}}}{\tau^\beta} \quad \text{e} \quad \Lambda(\tau|z_i) = \delta_i \gamma^{g_{ij}}. \quad (6.18)$$

A função de verossimilhança para o i -ésimo sistema do modelo equivalente é dada por

$$L(\boldsymbol{\mu}|z_i, \mathbf{t}) = \left(\prod_{j=1}^{n_i} \frac{\delta_i \beta t_{ij}^{\beta-1} \gamma^{g_{ij}}}{\tau^\beta} \right) \exp(-\delta_i \gamma^{g_{ij}}), \quad (6.19)$$

e a função de log-verossimilhança para os k sistemas é dada por

$$\begin{aligned} \ell(\boldsymbol{\mu}|z_i, \mathbf{t}) &= \sum_{i=1}^k n_i \log(\delta_i) + N \log(\beta) + (\beta - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) + \log(\gamma) \sum_{i=1}^k n_i g_{ij} \\ &- N \beta \log(\tau) - \sum_{i=1}^k \delta_i \gamma^{g_{ij}}. \end{aligned} \quad (6.20)$$

As equações escore para esse modelo são dadas por

$$\frac{\partial \ell}{\partial \beta} = \frac{N}{\beta} + \sum_{i=1}^k \sum_{j=1}^{n_i} \log(t_{ij}) - N \log(\tau) = 0,$$

$$\frac{\partial \ell}{\partial \delta_i} = \frac{n_i}{\delta_i} - \gamma^{g_{ij}} = 0, \quad i = 1, \dots, k,$$

$$\frac{\partial \ell}{\partial \gamma} = \frac{\sum_{i=1}^k n_i g_{ij}}{\gamma} - \sum_{i=1}^k \delta_i g_{ij} \gamma^{g_{ij}-1} = 0.$$

Utilizando a equação escore de β , temos novamente o estimador de β , dado por

$$\hat{\beta} = \frac{N}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\tau/t_{ij})}. \quad (6.21)$$

Da equação escore de δ_i , temos

$$\frac{n_i}{\delta_i} = \gamma^{g_{ij}}, \quad (6.22)$$

e então,

$$\hat{\delta}_i = \frac{n_i}{\hat{\gamma}^{g_{ij}}}. \quad (6.23)$$

Sabemos que a esperança condicional do número de falhas N dentro de um determinado tempo τ , com uma dada fragilidade Z e com um determinado número de falhas acumuladas g em um sistema seguindo o PLP, é

$$\mathbb{E}[N|Z, \tau, g] = Z\delta\gamma^g, \quad (6.24)$$

e, através da suposição de $\mathbb{E}[Z] = 1$, que torna o modelo de fragilidade identificável, temos que

$$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|Z, \tau, g]] = \delta\mathbb{E}[\gamma^g],$$

$$\mathbb{E}\left[\frac{N}{\gamma^g}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{N}{\gamma^g} \middle| Z, \tau, g\right]\right] = \mathbb{E}\left[\frac{1}{\gamma^g}\mathbb{E}[N|Z, \tau, g]\right] = \delta,$$

e assim,

$$\delta = \frac{\mathbb{E}[N]}{\mathbb{E}[\gamma^g]} = \mathbb{E}\left[\frac{N}{\gamma^g}\right]. \quad (6.25)$$

A partir da esperança de δ , nós então introduzimos um estimador $\bar{\delta}$ baseado no método dos momentos, dado por

$$\bar{\delta} = \frac{1}{k} \sum_{i=1}^k \bar{\delta}_i = \frac{1}{k} \sum_{i=1}^k \frac{n_i}{\gamma^{g_{ij}}}, \quad (6.26)$$

e então as fragilidades individuais z_i podem ser estimadas como

$$\hat{z}_i = \frac{\bar{\delta}_i}{\bar{\delta}} = \frac{\frac{n_i}{\gamma^{g_{ij}}}}{\frac{1}{k} \sum_{i=1}^k \frac{n_i}{\gamma^{g_{ij}}}}. \quad (6.27)$$

Os estimadores $\hat{\delta}$ e \hat{z}_i dependem do parâmetro γ , mas $\hat{\gamma}$ não tem forma fechada, então métodos numéricos são considerados.

6.3 Aplicação - Dados de colheitadeiras

Nesta seção, revisitaremos o conjunto de dados “Dados de colheitadeiras”. O objetivo aqui é novamente obter as estimativas dos parâmetros do PLP, assim como na Seção 3.4, mas também obter as estimativas das fragilidades individuais de cada colheitadeira para conseguir identificar qual colheitadeira é a mais frágil. Estamos interessados também, assim como no Capítulo 5, em verificar se, com o acúmulo de falhas, aumenta ou diminui a intensidade de falha.

Nas Seções 3.4, 4.3 e 5.4, observamos que o modelo PLP é adequado em cada uma das colheitadeiras de cana-de-açúcar. Notamos também que temos um indicativo de que as colheitadeiras estão melhorando ao longo do tempo; que espera-se que no período de 195 dias ocorram 12,5 falhas; que as colheitadeiras são pouco heterogêneas e que à medida que aumenta a quantidade de falhas, aumenta o risco de uma nova falha.

Para essa aplicação, primeiro ajustamos o modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica, dado em 6.1. Na Tabela 17 encontram-se as estimativas dos parâmetros do modelo.

Tabela 17 – Estimativas dos parâmetros do modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica para os dados de colheitadeiras

Parâmetros	Estimativa	EP	IC (95%)
$\hat{\beta}$	0,899	0,141	(0,623 ; 1,175)
$\hat{\delta}$	12,667	1,489	(9,749 ; 15,585)
z_1	0,868	0,089	(0,694 ; 1,042)
z_2	1,105	0,142	(0,827 ; 1,383)
z_3	1,026	0,101	(0,828 ; 1,224)

A partir da Tabela 17, vemos que $\hat{\beta} = 0,899$, então novamente temos um indicativo de que a confiabilidade dos sistemas está melhorando ao longo do tempo, pois $\hat{\beta} < 1$, e vemos que o número esperado de falhas é de aproximadamente 12,7 falhas até o tempo 195 dias, pois $\hat{\delta} = 12,667$.

A Tabela 17 também fornece as estimativas das fragilidades individuais de cada colheitadeira. A partir das estimativas das fragilidades individuais z_i , podemos notar que a colheitadeira mais frágil é a colheitadeira 2 ($\hat{z}_2 = 1,105$) e a colheitadeira menos frágil é a colheitadeira 1 ($\hat{z}_1 = 0,868$).

A partir da função de intensidade de falha dada em (6.1), a função estimada para os dados das colheitadeiras é

$$\hat{\lambda}(t_{ij}) = \frac{\hat{z}_i \hat{\beta} t_{ij}^{\hat{\beta}-1} \hat{\delta}}{\tau \hat{\beta}} = \frac{\hat{z}_i 0,899 t_{ij}^{0,899-1} 12,667}{195^{0,899}} = \hat{z}_i 0,099 t_{ij}^{-0,101}, \quad (6.28)$$

em que \hat{z}_i é a fragilidade estimada para cada colheitadeira. A representação gráfica dessa função é apresentada na Figura 26. Confirmando os resultados analisados da Tabela 17, a intensidade de

falha das colheitadeiras são funções decrescentes, ou seja, a intensidade com que ocorrem as falhas diminui ao longo do tempo. Vemos também que a função de intensidade de falha é maior para a colheitadeira 2, confirmando que ela é a mais frágil e a menor função de intensidade de falha é para a colheitadeira 1, demonstrando que ela é a menos frágil.

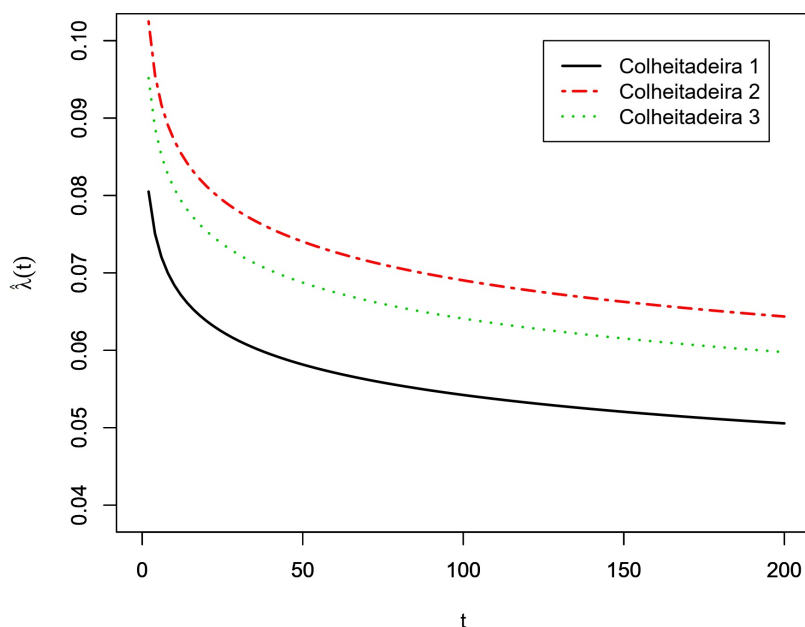


Figura 26 – Função de intensidade do modelo de fragilidade não-paramétrico ajustado aos dados das colheitadeiras.

Agora, além de verificar qual colheitadeira é mais frágil, queremos verificar também se o acúmulo de falhas aumenta ou diminui a intensidade de falhas. Para isso, ajustamos aos dados a CGMF sob RM e fragilidade não-paramétrica, dado em 6.11. As estimativas dos parâmetros se encontram na Tabela 18.

Tabela 18 – Estimativas dos parâmetros da CGMF sob RM e fragilidade não-paramétrica para os dados de colheitadeiras

Parâmetros	Estimativa	EP	IC (95%)
$\hat{\beta}$	0,899	0,106	(0,691 ; 1,107)
$\hat{\delta}$	12,695	2,532	(7,732 ; 17,658)
$\hat{\gamma}$	0,783	0,079	(0,628 ; 0,938)
\hat{z}_1	0,543	0,042	(0,461 ; 0,625)
\hat{z}_2	1,434	0,134	(1,171 ; 1,697)
\hat{z}_3	1,002	0,071	(0,863 ; 1,141)

Através das estimativas encontradas na Tabela 18, vemos que $\hat{\beta} = 0,899$, o que novamente sugere que os sistemas não estão se deteriorando, pois $\hat{\beta} < 1$. Com a estimativa de $\hat{\delta}$, novamente vemos que espera-se que ocorram aproximadamente 12,5 falhas em 195 dias. A partir das estimativas das fragilidades individuais de cada colheitadeira, podemos perceber que a colheitadeira mais frágil é a 2 e a menos frágil é a 1.

Com a Tabela 18, vemos que $\hat{\gamma} = 0,783$, sugerindo que o efeito da acumulação de falhas diminui a intensidade de falhas, pois $\hat{\gamma} < 1$.

6.4 Aplicação - Dados de caminhões basculantes

Aqui revisitaremos o conjunto de dados “Dados caminhões basculantes”. Nas Seções 3.5, 4.4 e 5.5, pudemos notar que o modelo PLP se mostrou adequado para cada um dos cinco caminhões; que há um indicativo de que os caminhões basculantes estão se deteriorando ao longo do tempo, ou seja, a função de intensidade é uma função crescente; que espera-se que ocorram 25,8 falhas em cada caminhão basculante no período de 107 dias; que os caminhões basculantes são pouco heterogêneos e que, à medida que aumenta a quantidade de falhas, aumenta o risco de uma nova falha nos caminhões.

Nessa seção, aplicaremos o modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica e a CGMF sob RM e fragilidade não-paramétrica. Na Tabela 19 encontram-se as estimativas dos parâmetros do modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica.

Tabela 19 – Estimativas dos parâmetros do modelo de múltiplos sistemas reparáveis com fragilidade não-paramétrica para os dados de caminhões basculantes.

Parâmetros	Estimativa	EP	IC (95%)
β	1,095	0,094	(0,911 ; 1,279)
δ	25,800	2,657	(20,592 ; 31,008)
z_1	0,891	0,116	(0,664 ; 1,118)
z_2	1,240	0,201	(0,846 ; 1,634)
z_3	0,891	0,116	(0,664 ; 1,118)
z_4	1,085	0,153	(0,785 ; 1,385)
z_5	0,891	0,116	(0,664 ; 1,118)

Assim como os resultados encontrados nas aplicações anteriores, vemos pela Tabela 19 que os caminhões estão se deteriorando com o tempo, pois $\hat{\beta} = 1,095$ e, como $\hat{\delta} = 25,8$, espera-se que cada caminhão falhe 25,8 vezes em 107 dias. Através das estimativas das fragilidades individuais z_i , podemos observar que o caminhão mais frágil é o caminhão 2 e os caminhões menos frágeis são os caminhões 1, 3 e 5, porém, o IC das estimativas contém o 1.

A função de intensidade de falha estimada para os dados dos caminhões, a partir da função (6.1), é dada por

$$\hat{\lambda}(t_{ij}) = \frac{\hat{z}_i \hat{\beta} t_{ij}^{\hat{\beta}-1} \hat{\delta}}{\tau \hat{\beta}} = \frac{\hat{z}_i 1,095 t_{ij}^{1,095-1} 25,8}{107^{1,095}} = \hat{z}_i 0,169 t_{ij}^{0,095}, \quad (6.29)$$

em que \hat{z}_i é a fragilidade estimada para cada caminhão. A representação gráfica dessa função é apresentada na Figura 27. Confirmando os resultados analisados da Tabela 19, a intensidade de

falha dos caminhões são funções crescentes, ou seja, a intensidade com que ocorrem as falhas aumenta ao longo do tempo. Também, podemos observar que a função de intensidade de falha é maior para o caminhão 2, confirmando que ele é o mais frágil e a menor função de intensidade de falha é para os caminhões 1, 3 e 5, demonstrando que eles são os menos frágeis, porém, o IC das estimativas contém o 1.

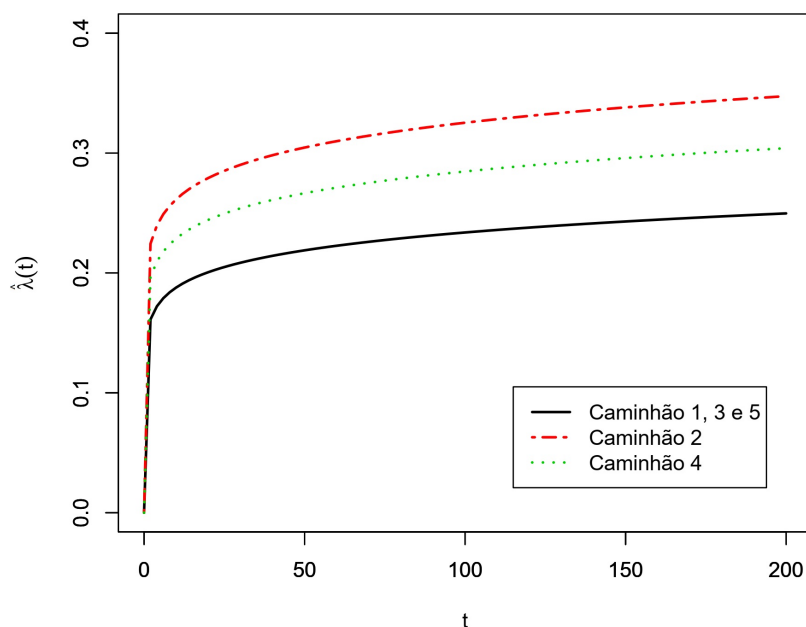


Figura 27 – Função de intensidade do modelo de fragilidade não-paramétrica ajustado aos dados dos caminhões.

Agora, a fim de verificar se o acúmulo de falhas aumenta ou diminui a intensidade de falha e, conjuntamente verificar qual caminhão é mais frágil, ajustamos aos dados a CGMF sob RM e fragilidade não-paramétrica. As estimativas dos parâmetros se encontram na Tabela 20.

Tabela 20 – Estimativas dos parâmetros da CGMF sob RM e fragilidade não-paramétrica para os dados de caminhões basculantes.

Parâmetros	Estimativa	EP	IC (95%)
β	1,095	0,096	(0,906 ; 1,283)
δ	25,108	3,077	(19,077 ; 31,139)
γ	1,002	0,004	(0,994 ; 1,009)
z_1	0,891	0,136	(0,624 ; 1,157)
z_2	2,265	0,683	(0,926 ; 3,603)
z_3	0,891	0,136	(0,624 ; 1,157)
z_4	1,695	0,256	(1,193; 2,196)
z_5	0,891	0,136	(0,624 ; 1,157)

Assim como na aplicação do modelo anterior, $\hat{\beta} = 1,095$ sugere que os sistemas estão se deteriorando, pois $\hat{\beta} > 1$ e espera-se que ocorram 25,11 falhas em 107 dias em cada caminhão. A partir das estimativas das fragilidades individuais de cada caminhão basculante, notamos

novamente que o caminhão mais frágil é o caminhão 2 e os menos frágeis são os caminhões 1, 3 e 5.

Também, a partir da Tabela 20, vemos que $\hat{\gamma} = 1,002$, sugerindo que o efeito da acumulação de falhas aumenta um pouco a intensidade de falhas, pois $\hat{\gamma} > 1$. Assim, caminhões basculantes que apresentam mais falhas devem ter maiores funções de intensidade de falha. É possível perceber que os parâmetros de fragilidade \hat{z}_i e as acumulações de falha g_{ij} estão intimamente ligados, pois, em geral, os caminhões que têm maior quantidade de falhas acumuladas também têm maior \hat{z}_i , o que está de acordo com o parâmetro estimado $\hat{\gamma}$.

6.5 Considerações finais

Neste capítulo propomos o modelo de fragilidade e a CGMF em sistemas reparáveis considerando a fragilidade de forma não-paramétrica. Nesta abordagem, as variáveis de fragilidade individuais de cada sistema são consideradas diretamente os parâmetros a serem estimados. Consideramos o modelo PLP reparametrizado, por sua interpretação simples e por seu tratamento algébrico facilitado.

A vantagem dessa abordagem em relação à abordagem paramétrica, em que supomos a fragilidade com distribuição gama, é que aqui não se faz necessária a suposição de distribuição para a fragilidade, evitando assim problemas de mau ajuste. Uma outra vantagem é a possibilidade de estimar facilmente a fragilidade individual, então conseguimos ter um indicativo de quais sistemas são mais ou menos frágeis. Na modelagem com fragilidade gama, apenas é possível estimar a variância da fragilidade e, caso a variância seja alta, indica que existem efeitos aleatórios causando heterogeneidade entre os sistemas. Para o modelo de fragilidade, obtivemos ainda a vantagem de obter analiticamente os estimadores dos parâmetros.

Propomos o método de estimação para os parâmetros, em que no modelo da classe geral, as estimativas precisam ser encontradas utilizando algum método numérico e no modelo com fragilidade foi possível encontrar os estimadores de forma analítica, sendo uma vantagem a mais dessa abordagem.

Através da aplicação com os dados das colheitadeiras de cana-de-açúcar e dos caminhões basculantes foi possível demonstrar a aplicabilidade dessa abordagem. Conseguimos identificar quais colheitadeiras e quais caminhões são mais frágeis, tornando ainda mais rica a aplicação e possibilitando que se possa pensar futuramente em manutenções preventivas principalmente nos sistemas mais frágeis.

CONCLUSÕES E PROPOSTAS FUTURAS

Nesta tese, estudamos certos aspectos da modelagem de dados relacionados aos tempos de falha de múltiplos sistemas reparáveis sob a suposição de reparo mínimo. O modelo de reparo mínimo é plausível para sistemas que consistem em vários componentes e em situações em que se deseja apenas corrigir o efeito causador da falha, deixando o sistema rapidamente pronto para o uso novamente. É uma situação bastante comum na prática por sua facilidade de execução e por ser economicamente mais viável em diversas situações. Para modelar de forma paramétrica o RM, utilizamos o processo de lei de potência. O trabalho foi motivado por um conjunto de dados reais relacionados ao tempo até sucessivas falhas no sistema de corte de colhedei­ras de cana-de-açúcar e por um conjunto de dados de caminhões basculantes que trabalharam sob condições extremas em uma mineradora brasileira.

O trabalho foi iniciado com uma revisão bibliográfica e uma revisão dos conceitos de confiabilidade, de processo de contagem, de sistemas reparáveis, do modelo de fragilidade e da estimação Bayesiana, o que ajuda no entendimento do trabalho e espera-se que seja útil em futuras pesquisas de diversas áreas. Apresentamos a estimação por máxima verossimilhança do modelo de reparo mínimo para múltiplos sistemas reparáveis já conhecida na literatura e propomos um método de estimação Bayesiana objetiva, utilizando a priori de referência. Através de um estudo de simulação, comparamos as propriedades utilizando tal priori com a estimação clássica e com a também já conhecida priori de Jeffreys, em que verificamos que em geral a estimação por máxima verossimilhança não tem as melhores propriedades. Verificamos que usando o estimador Bayesiano MAP com a priori de Jeffreys nos retornou melhores estimativas com relação ao vício e, com a priori de referência, nos retornou melhores estimativas com relação à probabilidade de cobertura do intervalo de credibilidade. Fizemos a aplicação aos dois conjuntos de dados, em que foi possível verificar que as colheitadeiras estão melhorando a confiabilidade conforme o tempo passa e os caminhões estão se deteriorando conforme o tempo passa. Em cada aplicação também verificamos o número esperado de falhas dentro do período de estudo.

O modelo de fragilidade sob RM foi apresentado e o método de estimação foi proposto medindo a heterogeneidade não-observada entre os sistemas, através da inserção de uma variável aleatória latente na função de intensidade de falha de forma multiplicativa, em que supomos que a variável de fragilidade tem distribuição gama. Encontramos a função de verossimilhança através de integração de forma direta e através da transformada de Laplace, em que a segunda maneira nos permitiu encontrar diretamente a função de confiabilidade e a função de intensidade. Fizemos um estudo de simulação, em que verificamos suas propriedades frequentistas e refizemos a aplicação aos conjuntos de dados, em que verificamos que as colheitadeiras de cana-de-açúcar e os caminhões basculantes são pouco heterogêneos.

Propomos uma classe geral de modelos de fragilidade para múltiplos sistemas reparáveis, incluindo o termos de fragilidade, o efeito do acúmulo dos reparos e covariáveis observadas, e apresentamos um método de estimação adequado para uma situação em específico: quando são realizados reparos mínimos, não são observadas covariáveis, que é exatamente a situação dos conjuntos de dados analisados, e supomos que a variável de fragilidade tem distribuição gama. Essa modelagem é muito interessante, pois possibilita verificar se em sistemas que são realizados mais reparos, eles tendem a ter mais ou menos falhas, o que em modelos tradicionais de sistemas reparáveis não é possível identificar. Conseguimos também, através do parâmetro da distribuição gama, verificar se existe algum fator externo que torna os sistemas heterogêneos. Além disso, fizemos um estudo de simulação, observando suas propriedades frequentistas e foram realizadas aplicações com dados reais, em que foi possível mostrar a aplicabilidade do modelo.

Consideramos também para a distribuição da fragilidade uma abordagem não-paramétrica, em que as fragilidades individuais são diretamente os parâmetros a serem estimados. Essa abordagem é bastante interessante pois não fazemos a suposição de alguma distribuição para a variável de fragilidade, evitando um possível problema de mal ajuste e, pois, possibilita estimar diretamente quais sistemas são mais frágeis e quais sistemas são menos frágeis. Essa identificação de sistemas mais frágeis é muito interessante na prática pois é possível prestar mais atenção nesses sistemas e talvez até adotar políticas de manutenção preventiva mais rigorosas nesses sistemas. Aplicamos essa abordagem no modelo de fragilidade em sistemas reparáveis e na classe geral de modelos de fragilidade. Na aplicação aos dados, conseguimos estimar quais colheitadeiras e quais caminhões dos conjuntos de dados são mais frágeis.

Como propostas futuras, podemos considerar a abordagem Bayesiana objetiva para estimar os parâmetros do modelo de fragilidade em sistemas reparáveis e da classe geral de modelos de fragilidade. Propomos também considerar a abordagem Bayesiana objetiva para estimar os parâmetros dos modelos em que utilizamos a abordagem não-paramétrica para a variável de fragilidade. Além disso, podemos considerar outra forma paramétrica além do processo de lei de potência. E, para se obter uma modelagem mais flexível, propomos considerar outros tipos de reparo, como o reparo perfeito e o reparo imperfeito. Por fim, podemos elaborar uma política de manutenção preventiva a fim de que, a partir da modelagem apresentada neste

trabalho, seja possível obter o tempo ótimo de manutenção preventiva.

REFERÊNCIAS

- AALEN, O. Nonparametric inference for a family of counting processes. **The Annals of Statistics**, JSTOR, p. 701–726, 1978. Citado nas páginas 20 e 30.
- AALEN, O.; BORGAN, O.; GJESSING, H. **Survival and event history analysis: a process point of view**. [S.l.]: Springer Science & Business Media, 2008. Citado na página 74.
- ANDERSEN, P. K.; BORGAN, O.; GILL, R. D.; KEIDING, N. **Statistical models based on counting processes**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 44.
- ASCHER, H.; FEINGOLD, H. **Repairable systems reliability: modeling, inference, misconceptions and their causes**. [S.l.]: M. Dekker New York, 1984. Citado na página 19.
- BAIN, L. J.; WRIGHT, F. The negative binomial process with applications to reliability. **Journal of Quality Technology**, v. 14, p. 60–66, 1982. Citado na página 21.
- BARLOW, R. E.; PROSCHAN, F. **Mathematical Theory of Reliability**. [S.l.]: Society for Industrial and Applied Mathematics, 1987. Citado na página 20.
- BERGER, J. O.; BERNARDO, J. M. Estimating a product of means: Bayesian analysis with reference priors. **Journal of the American Statistical Association**, Taylor & Francis, v. 84, n. 405, p. 200–207, 1989. Citado na página 49.
- BERGER, J. O.; BERNARDO, J. M. *et al.* On the development of reference priors. **Bayesian statistics**, v. 4, n. 4, p. 35–60, 1992. Citado na página 49.
- BERGER, J. O.; BERNARDO, J. M.; SUN, D. *et al.* Overall objective priors. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 10, n. 1, p. 189–221, 2015. Citado na página 50.
- BERMAN, M.; TURNER, T. R. Approximating point process likelihoods with glim. **Applied Statistics**, JSTOR, p. 31–38, 1992. Citado na página 20.
- BERNARDO, J.; SMITH, A. Bayesian theory. **Chichester: Wiley**, 1994. Citado na página 47.
- BERNARDO, J. M. Reference posterior distributions for bayesian inference. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 113–147, 1979. Citado na página 49.
- BERNARDO, J. M.; RAMÓN, J. M. An introduction to bayesian reference analysis: inference on the ratio of multinomial parameters. **Journal of the Royal Statistical Society: Series D (The Statistician)**, Wiley Online Library, v. 47, n. 1, p. 101–135, 1998. Citado na página 49.
- BLOCK, H. W.; BORGES, W. S.; SAVITS, T. H. Age-dependent minimal repair. **Journal of applied probability**, Cambridge University Press, v. 22, n. 2, p. 370–385, 1985. Citado na página 42.
- BROWN, M.; PROSCHAN, F. Imperfect repair. **Journal of Applied probability**, JSTOR, p. 851–859, 1983. Citado na página 42.

- 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**, Biometrika Trust, v. 65, n. 1, p. 141–151, 1978. Citado nas páginas 44 e 46.
- COETZEE, J. L. The role of nhpp models in the practical analysis of maintenance failure data. **Reliability Engineering & System Safety**, Elsevier, v. 56, n. 2, p. 161–168, 1997. Citado na página 35.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: **ABE-Projeto Fisher**. [S.l.]: Edgard Blücher, 2006. Citado na página 28.
- COOK, R. J.; LAWLESS, J. **The statistical analysis of recurrent events**. [S.l.]: Springer Science & Business Media, 2007. Citado nas páginas 29, 41 e 42.
- COX, D. The statistical analysis of dependencies in point processes. **Stochastic point processes**, New York: Wiley, p. 55–66, 1972. Citado na página 22.
- COX, D. R. Regression models and life tables (with discussion). **Journal of the Royal Statistical Society, Series B**, v. 34, n. 2, 1972. Citado na página 44.
- COX, D. R.; ISHAM, V. **Point processes**. [S.l.]: CRC Press, 1980. v. 12. Citado nas páginas 20 e 21.
- COX, D. R.; LEWIS, P. A. The statistical analysis of series of events. **Monographs on Applied Probability and Statistics, London: Chapman and Hall, 1966**, v. 1, 1996. Citado nas páginas 20 e 21.
- CROW, L. H. **Reliability analysis for complex, repairable systems**. In **Reliability and Biometry**, eds. F. Proschan and R. J. Serfling, Philadelphia. [S.l.], 1974. Citado nas páginas 35, 36 e 37.
- D'ANDREA, A. M. E.; FEITOSA, C. C.; TOMAZELLA, V. L. D.; VIEIRA, A. M. C. Frailty modeling for repairable systems with minimum repair: An application to dump truck data of a brazilian mining company. **Journal of Mathematics and Statistical Science**, v. 3, n. 1, p. 179–198, 2017. Citado nas páginas 21 e 103.
- DATTA, G. S.; MUKERJEE, R. **Probability matching priors: higher order asymptotics**. [S.l.]: Springer Science & Business Media, 2012. v. 178. Citado na página 51.
- DOYEN, L. On the brown–proschan model when repair effects are unknown. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 27, n. 6, p. 600–618, 2011. Citado na página 42.
- DOYEN, L.; GAUDOIN, O. Classes of imperfect repair models based on reduction of failure intensity or virtual age. **Reliability Engineering & System Safety**, Elsevier, v. 84, n. 1, p. 45–56, 2004. Citado nas páginas 40 e 42.
- DUANE, J. Learning curve approach to reliability monitoring. **IEEE transactions on Aerospace**, IEEE, v. 2, n. 2, p. 563–566, 1964. Citado na página 35.
- ELBERS, C.; RIDDER, G. True and spurious duration dependence: The identifiability of the proportional hazard model. **The Review of Economic Studies**, Oxford University Press, v. 49, n. 3, p. 403–409, 1982. Citado na página 47.

FEITOSA, C. C. **Modelagem de dados de sistemas reparáveis com fragilidade**. Dissertação (Mestrado) — Universidade Federal de São Carlos, 2015. Citado nas páginas 21 e 103.

FERNANDES, R. da S. Modelos com reparo imperfeito no r. Universidade Federal de Minas Gerais, 2015. Citado na página 39.

GIORGIO, M.; GUIDA, M.; PULCINI, G. Repairable system analysis in presence of covariates and random effects. **Reliability Engineering & System Safety**, Elsevier, v. 131, p. 271–281, 2014. Citado na página 74.

GONZÁLEZ, J. R.; PEÑA, E. A.; SLATE, E. H. Modelling intervention effects after cancer relapses. **Statistics in medicine**, Wiley Online Library, v. 24, n. 24, p. 3959–3975, 2005. Citado na página 87.

GOUVÊA, G. D. R.; TOMAZELLA, V. L. D.; SCALON, J. D. Estimação bayesiana das fragilidades individuais de pacientes com insuficiência renal crônica da cidade de lavras-mg. **Rev. Bras. Biom**, v. 29, n. 3, p. 534–548, 2011. Citado na página 88.

GUIDA, M.; CALABRIA, R.; PULCINI, G. Bayes inference for a non-homogeneous poisson process with power intensity law (reliability). **IEEE Transactions on Reliability**, IEEE, v. 38, n. 5, p. 603–609, 1989. Citado na página 20.

GUO, H. R.; LIAO, H.; ZHAO, W.; METTAS, A. A new stochastic model for systems under general repairs. **IEEE Transactions on Reliability**, IEEE, v. 56, n. 1, p. 40–49, 2007. Citado na página 22.

HAMADA, M. S.; WILSON, A.; REESE, C. S.; MARTZ, H. **Bayesian reliability**. [S.l.]: Springer Science & Business Media, 2008. Citado nas páginas 27, 32, 34 e 36.

HOUGAARD, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. **Biometrika**, Biometrika Trust, v. 71, n. 1, p. 75–83, 1984. Citado na página 46.

_____. Frailty models for survival data. **Lifetime data analysis**, Springer, v. 1, n. 3, p. 255–273, 1995. Citado na página 44.

_____. **Analysis of multivariate survival data**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 21.

HUANG, Y.-S. A decision model for deteriorating repairable systems. **Iie Transactions**, Springer, v. 33, n. 6, p. 479–485, 2001. Citado na página 20.

HUSSAIN, V. M. S.; NAIKAN, V. Point process based maintenance modeling for repairable systems: A review. 2010. Citado na página 21.

JEFFREYS, H. An invariant form for the prior probability in estimation problems. **Proc. R. Soc. Lond. A**, The Royal Society, v. 186, n. 1007, p. 453–461, 1946. Citado na página 49.

JEFFREYS, S. H. **Theory of Probability: 3d Ed**. [S.l.]: Clarendon Press, 1967. Citado na página 48.

KIJIMA, M. Some results for repairable systems with general repair. **Journal of Applied probability**, JSTOR, p. 89–102, 1989. Citado nas páginas 20, 39, 40, 42 e 88.

- KIJIMA, M.; MORIMURA, H.; SUZUKI, Y. Periodical replacement problem without assuming minimal repair. **European Journal of Operational Research**, Elsevier, v. 37, n. 2, p. 194–203, 1988. Citado na página 42.
- KIM, T.; PARK, C.; AHN, S. Determining the optimal maintenance action for a deteriorating repairable system. **Probabilistic Engineering Mechanics**, Elsevier, v. 23, n. 1, p. 95–101, 2008. Citado na página 20.
- KRIVTSOV, V. V. Practical extensions to nhpp application in repairable system reliability analysis. **Reliability Engineering & System Safety**, Elsevier, v. 92, n. 5, p. 560–562, 2007. Citado na página 21.
- KUMAR, U.; KLEFSJÖ, B. Reliability analysis of hydraulic systems of lhd machines using the power law process model. **Reliability Engineering & System Safety**, Elsevier, v. 35, n. 3, p. 217–224, 1992. Citado na página 35.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 1982. Citado na página 28.
- _____. Regression methods for poisson process data. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 82, n. 399, p. 808–815, 1987. Citado nas páginas 45 e 74.
- LEÃO, J.; LEIVA, V.; SAULO, H.; TOMAZELLA, V. Birnbaum–saunders frailty regression models: Diagnostics and application to medical data. **Biometrical Journal**, Wiley Online Library, v. 59, n. 2, p. 291–314, 2017. Citado na página 46.
- LEHMANN, A. Degradation based reliability models with imperfect repair. Citeseer, 2002. Citado na página 21.
- LEUNG, F. K.; FONG, C. A repair-replacement study for gearboxes using geometric processes. **International Journal of Quality & Reliability Management**, MCB UP Ltd, v. 17, n. 3, p. 285–304, 2000. Citado na página 21.
- LINDQVIST, B. The trend-renewal process, a useful model for repairable systems. In: **Society in Reliability Engineers, Scandinavian Chapter, Annual Conference, Malmö, Sweden**. [S.l.: s.n.], 1993. Citado na página 21.
- LINDQVIST, B. H. On the statistical modeling and analysis of repairable systems. **Statistical science**, JSTOR, p. 532–551, 2006. Citado nas páginas 37 e 74.
- LINDQVIST, B. H.; ELVEBAKK, G.; HEGGLAND, K. The trend-renewal process for statistical analysis of repairable systems. **Technometrics**, Taylor & Francis, v. 45, n. 1, p. 31–44, 2003. Citado na página 21.
- LINDQVIST, B. H.; SKOGSRUD, G. Modeling of dependent competing risks by first passage times of wiener processes. **IIE Transactions**, Taylor & Francis, v. 41, n. 1, p. 72–80, 2008. Citado na página 21.
- MANDA, S. O. A nonparametric frailty model for clustered survival data. **Communications in Statistics—Theory and Methods**, Taylor & Francis, v. 40, n. 5, p. 863–875, 2011. Citado na página 107.

- MOALA, F. A.; GARCIA, L. M. A bayesian analysis for the parameters of the exponential-logarithmic distribution. **Quality Engineering**, n. 25:3, p. 282–291, 2013. Citado na página 20.
- MURALIDHARAN, K. A review of repairable systems and point process models. In: **ProbStat Forum**. [S.l.: s.n.], 2008. v. 1, p. 26–49. Citado nas páginas 20 e 39.
- NAKAGAWA, T. **Maintenance theory of reliability**. [S.l.]: Springer Science & Business Media, 2006. Citado nas páginas 41 e 42.
- 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 nas páginas 44 e 46.
- O'HAGAN, A. **Kendall's Advanced Theory of Statistics, Vol. 2b: Bayesian Statistics**. [S.l.]: John Wiley & Sons, 1994. Citado na página 48.
- OLIVEIRA, M. D. D.; COLOSIMO, E. A.; GILARDONI, G. L. Power law selection model for repairable systems. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 42, n. 4, p. 570–578, 2014. Citado nas páginas 35 e 36.
- OLIVEIRA, M. D. de; COLOSIMO, E. A.; GILARDONI, G. L. Bayesian inference for power law processes with applications in repairable systems. **Journal of Statistical Planning and Inference**, Elsevier, v. 142, n. 5, p. 1151–1160, 2012. Citado nas páginas 20, 37 e 57.
- PARK, D. H.; JUNG, G. M.; YUM, J. K. Cost minimization for periodic maintenance policy of a system subject to slow degradation. **Reliability Engineering & System Safety**, Elsevier, v. 68, n. 2, p. 105–112, 2000. Citado na página 20.
- PEÑA, E. A.; HOLLANDER, M. Models for recurrent events in reliability and survival analysis. In: **Mathematical reliability: An expository perspective**. [S.l.]: Springer, 2004. p. 105–123. Citado nas páginas 22, 87, 88 e 89.
- PEÑA, E. A.; SLATE, E. H.; GONZÁLEZ, J. R. Semiparametric inference for a general class of models for recurrent events. **Journal of Statistical Planning and Inference**, Elsevier, v. 137, n. 6, p. 1727–1747, 2007. Citado na página 88.
- PERCY, D. F.; ALKALI, B. M. Generalized proportional intensities models for repairable systems. **IMA Journal of Management Mathematics**, Oxford University Press, v. 17, n. 2, p. 171–185, 2006. Citado na página 22.
- PHELPS, R. Replacement policies under minimal repair. **Journal of the Operational Research Society**, Springer, v. 32, n. 7, p. 549–554, 1981. Citado na página 20.
- PROSCHAN, F. Theoretical explanation of observed decreasing failure rate. **Technometrics**, Taylor & Francis Group, v. 5, n. 3, p. 375–383, 1963. Citado na página 21.
- PULCINI, G. Modeling the failure data of a repairable equipment with bathtub type failure intensity. **Reliability Engineering & System Safety**, Elsevier, v. 71, n. 2, p. 209–218, 2001. Citado na página 21.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. Citado na página 62.

- RIGDON, S. E.; BASU, A. P. **Statistical methods for the reliability of repairable systems**. [S.l.]: Wiley New York, 2000. Citado nas páginas 19, 20, 37, 38, 41, 42, 55, 61, 66 e 78.
- SCHERVISH, M. J. **Theory of statistics**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 48.
- SEN, A. Bayesian estimation and prediction of the intensity of the power law process. **Journal of Statistical computation and simulation**, Taylor & Francis, v. 72, n. 8, p. 613–631, 2002. Citado na página 20.
- SINHA, D.; DEY, D. K. Semiparametric bayesian analysis of survival data. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 92, n. 439, p. 1195–1212, 1997. Citado na página 44.
- SLIMACEK, V.; LINDQVIST, B. H. Nonhomogeneous poisson process with nonparametric frailty. **Reliability Engineering & System Safety**, Elsevier, v. 149, p. 14–23, 2016. Citado nas páginas 47, 107, 108, 109, 110 e 112.
- STOCKER IV, R. S.; PEÑA, E. A. A general class of parametric models for recurrent event data. **Technometrics**, Taylor & Francis, v. 49, n. 2, p. 210–221, 2007. Citado na página 88.
- TIAO, G. C.; BOX, G. E. Some comments on “bayes” estimators. **The American Statistician**, Taylor & Francis Group, v. 27, n. 1, p. 12–14, 1973. Citado na página 49.
- TIBSHIRANI, R. Noninformative priors for one parameter of many. **Biometrika**, Oxford University Press, v. 76, n. 3, p. 604–608, 1989. Citado na página 51.
- TOLEDO, M. L. G. d. **Determination of the optimal periodic maintenance policy under imperfect repair assumption**. Tese (Doutorado) — Escola de Engenharia - Departamento de Engenharia da Produção, 2014. Citado na página 24.
- TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação, 2003. Citado nas páginas 20, 29 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 21 e 44.
- VERSSANI, B. A. W. **Modelo de regressão para sistemas reparáveis: um estudo da confiabilidade de colhedoras de cana-de-açúcar**. Dissertação (Mestrado) — Universidade de São Paulo, 2018. Citado na página 23.
- WANG, H. A survey of maintenance policies of deteriorating systems. **European journal of operational research**, Elsevier, v. 139, n. 3, p. 469–489, 2002. Citado na página 20.
- WANG, M.-C.; CHANG, S.-H. Nonparametric estimation of a recurrent survival function. **Journal of the American Statistical Association**, Taylor & Francis, v. 94, n. 445, p. 146–153, 1999. Citado na página 131.
- WHITAKER, L. R.; SAMANIEGO, F. J. Estimating the reliability of systems subject to imperfect repair. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 84, n. 405, p. 301–309, 1989. Citado na página 42.

WIENKE, A. **Frailty models in survival analysis**. [S.l.]: CRC Press, 2010. Citado nas páginas 21, 44 e 45.

YU, J.-W.; TIAN, G.-L.; TANG, M.-L. Predictive analyses for nonhomogeneous poisson processes with power law using bayesian approach. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4254–4268, 2007. Citado na página 20.

ZHAO, M.; XIE, M. On maximum likelihood estimation for a general non-homogeneous poisson process. **Scandinavian journal of statistics**, JSTOR, p. 597–607, 1996. Citado na página 20.

ESTIMADOR NÃO-PARAMÉTRICO DE WANG E CHANG

Existem estimativas diferentes da função de confiabilidade, mas neste trabalho foi usado o estimador de Wang e Chang (WANG; CHANG, 1999); é um estimador para dados recorrentes, tempos correlacionados ou independentes e de distribuição idêntica. Sua função é derivada de um estimador do tipo KM para eventos recorrentes com base nos intervalos entre os tempos, que é $t_{i,j-1} - t_{i,j}$, chamado *gap times*. Considere T_{ij} o tempo entre o $j - 1$ -ésimo e j -ésimo evento do i -ésimo do sistema, C_i o “tempo de censura”, ou seja, o tempo entre o último evento e o fim do monitoramento do sistema i e m_i é de modo que

$$\sum_{j=1}^{m_i-1} T_{ij} \leq C_i$$

e

$$\sum_{j=1}^{m_i} T_{ij} > C_i.$$

Assim, o termo dados “não-censurados” refere-se ao conjunto $(t_{i1}t_{i2}, \dots, t_{i,m_i})$ e o termo “censurado” refere-se ao intervalo $(t_{i1}t_{i2}, \dots, t_{i,m_i}^+)$ em que t_{i,m_i}^+ refere-se ao tempo entre o $m_i - 1$ -ésimo evento e o final do acompanhamento. Então, m_i é o número de eventos recorrentes para o sistema i .

Também considere $m_i^* = 1$ se $m_i = 1$ e $m_i^* = m_i - 1$ se $m_i \geq 2$, y_{ij} os tempos recorrentes observados definidos por $y_{ij} = t_{ij}$ se $j = 1, 2, \dots, m_i - 1$ e $y_{ij} = t_{i,m_i}^+$ se $j = m_i$ e $R^*(t)$ o número total de sistemas em risco em t , dado por

$$R^*(t) = \sum_{i=1}^n \left[\frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} \geq t) \right],$$

em que a_i é uma função para valores positivos dos sistemas com valores censurados em que $\mathbb{E}[a_i^2] < \infty$ e $I(y_{ij} \geq t)$ é 0 se $y_{ij} < t$ e 1 se $y_{ij} \geq t$.

Seja

$$d^*(t) = \sum_{i=1}^n \left[\frac{a_i I(m_i \geq 2)}{m_i^*} \sum_{j=1}^{m_i^*} I(y_{ij} = t) \right].$$

Então, se $y_1^*, y_2^*, \dots, y_1^K$ são os tempos “não-censurados”, distintos e ordenados, o estimador de KM de Wang Chang é dado por

$$\hat{R}(t) = \prod_{y_i^* \geq t} \left\{ 1 - \frac{d^*(y_i^*)}{R^*(y_i^*)} \right\}.$$

Essa fórmula expressa a função de confiabilidade do tempo entre dois eventos sucessivos, conhecida como função de confiabilidade recorrente. Supõe-se que os sistemas tenham uma função de confiabilidade marginal comum e, às vezes, nas unidades intercorrentes, eles estejam correlacionados. O gráfico de KM de Wang Chang pode ser feito utilizando a função *WC.fit* do pacote *newTestSurvRec* do software R.

