**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelo de regressão para sistemas reparáveis: um estudo da
confiabilidade de colhedoras de cana-de-açúcar**

Bruna Aparecida Wruck Verssani

Dissertação apresentada para obtenção do título de Mestra
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2018**

Bruna Aparecida Wruck Verssani
Bacharel em Estatística

**Modelo de regressão para sistemas reparáveis: um estudo da
confiabilidade de colhedoras de cana-de-açúcar**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Profa. Dra. **SÔNIA MARIA DE STEFANO PIEDADE**

Dissertação apresentada para obtenção do título de Mestra
em Ciências. Área de concentração: Estatística e Experimen-
tação Agronômica

Piracicaba
2018

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Verssani, Bruna Aparecida Wruck

Modelo de regressão para sistemas reparáveis: um estudo da confiabilidade de colhedoras de cana-de-açúcar / Bruna Aparecida Wruck Verssani. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2018 .
53 p.

Dissertação (Mestrado) – – USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Sistema reparável 2. Processo de Poisson não-homogêneo 3. Distribuição odd log-logística Weibull flexível generalizada 4. Regressão GOLLW 5. Regressão Weibull. . I. Título.

DEDICATÓRIA

Aos *meus pais* **Jair Verssani** e **Saete Wruck**, que me apresentaram o caminho da família e a importância da honestidade, persistência e determinação.

Ao *meu irmão* **João Victor**, pela amizade e incentivo.

Ao *meu companheiro* **Felipe Braga** pelo apoio incondicional nos momentos de incerteza.

Sem vocês nenhuma conquista valeria a pena.

AGRADECIMENTOS

A Deus, pela fé em algo superior, pela força para seguir em frente perante os obstáculos e sabedoria para decisões difíceis.

A minha orientadora Prof^o Dra Sônia Maria De Stefano Piedade, pela acolhida e orientação nos momentos que precisei, sem ela esse trabalho não seria possível.

A CAPES, pela bolsa concedida.

A minha família, meus avós Odenilze e Luizinho, meus tios Emerson e Aparecida, minhas madrinhas Rosemeire e Vera Lúcia pelas orações e por sempre acreditarem em mim e darem força.

A todos os meus professores de graduação, em especial aos meus orientadores Prof^o Dr Sérgio Minoru Oikawa e Prof^o Dr José Gilberto Spasiane Rinaldi por me inserir no mundo das pesquisas.

A todos os professores do Programa de Pós-Graduação em Estatística e Experimentação Agronômica, pelos ensinamentos e dedicação a cada um dos alunos, em especial aos professores Prof^o Dr Carlos Tadeu Dos Santos Dias, Prof^o Dra Renata Alcarde Sermarini e Prof^o Dr César Gonçalves De Lima.

Às Secretárias do Departamento Solânge de Assis Paes Sabadin e Luciane Brajao e aos técnicos de informática Eduardo Bonilha e Jorge Alexandre Wiendl, por todo suporte, ajuda e disponibilidade junto aos alunos.

Aos meus amigos Patrícia, Alex, Raquel, Cayan, Portya, Leomir, Gustavo, Djair e Isaneli, pelos momentos de estudo, descontração, apoio e amizade. Quando as coisas ficam difíceis sempre temos o FrutaGurt e as Paletas Mexicanas.

A todos que de forma direta ou indireta me ajudaram no andamento e conclusão deste trabalho.

SUMÁRIO

RESUMO	6
ABSTRACT	7
1 INTRODUÇÃO	9
Referências	11
2 DISTRIBUIÇÃO ODD LOG-LOGÍSTICA WEIBULL FLEXÍVEL GENERALIZADA	13
2.1 Introdução	13
2.2 Definição do Modelo	14
2.3 Simulação	18
2.4 Modelo de Regressão GOLLFW	20
2.5 Aplicação	22
2.6 Conclusão	23
Referências	25
3 ANÁLISE DE CONFIABILIDADE PARA UM ÚNICO SISTEMA REPARÁVEL	27
3.1 Introdução	27
3.2 Material e Método	29
3.3 Resultados e Discussão	33
3.4 Conclusão	35
Referências	37
4 MODELO DE REGRESSÃO PARA DOIS OU MAIS SISTEMAS REPARÁVEIS	39
4.1 Introdução	39
4.2 Material e Método	40
4.3 Resultados e Discussão	43
4.4 Conclusão	44
Referências	45
APÊNDICES	47
ANEXOS	51

RESUMO

Modelo de regressão para sistemas reparáveis: um estudo da confiabilidade de colhedoras de cana-de-açúcar

A análise de confiabilidade desempenha um papel fundamental para estudos de durabilidade e otimização de tempos de reparo em sistemas reparáveis. Equipamentos como colhedoras de cana-de-açúcar que após a falha e um reparo voltam a exercer sua função objetivo são classificados como sistemas reparáveis. O objetivo deste trabalho consistiu em propor alternativas de modelagem para sistemas complexos, que apresentam grande variabilidade no comportamento da função intensidade de falha. Foi proposta a nova distribuição odd log-logística Weibull flexível generalizada (GOLFW) e um modelo de regressão Weibull aplicado ao processo lei de potência usado para analisar sistemas reparáveis. Para a nova distribuição foi apresentada a família de distribuições odd log-logística generalizada, realizado um estudo de simulação para verificar algumas propriedades dos estimadores de máxima verossimilhança e incluídas covariáveis na análise dos tempos de falha através do modelo de regressão GOLFW. Para a análise de regressão considerando os sistemas reparáveis, foram apresentados os principais modelos de contagem para um único sistema reparável e realizado a análise deles de forma separada e, em seguida, foram considerados mais de dois sistemas e acrescentado um modelo de regressão Weibull ao processo lei de potência (PLP). A característica de bimodalidade da distribuição GOLFW garantiu a adequabilidade e um melhor ajuste aos dados. Já a inclusão de covariáveis através do modelo de regressão Weibull no PLP permitiu modelar sistemas que antes somente os processos de contagens tradicionais, processo lei de potência e processo de renovação, não se adequariam bem.

Palavras-chave: Sistema reparável; Processo de Poisson não-homogêneo; Distribuição odd log-logística Weibull flexível generalizada; Regressão GOLFW; Regressão Weibull.

ABSTRACT

Regression model for repairable systems: a study of the reliability of sugarcane harvesters

The confiability analysis carries out an important role for durability studies and optimization of repair time in repairable systems. Repairable systems are equipments that returns to execute its function after a fail, for example, sugarcane harvester. This work aimed to propose modeling alternatives for complex systems with great variability in the behaviour of fail intensity function. It was proposed a new distribution on generalized odd log-logistic flexible Weibull (GOLLFW) and an Weibull regression model applied to potential law used to analyze repairable systems. It was presented the distribution family generalized odd log-logistic, was carried out a simulation study to verify some properties of maximum likelihood estimators and was included covariables in the fail time by regression model GOLLFW. To the regression analysis considering repairable systems, it was presented the main counting models for a single repairable system and it was performed an analysis of each model singly, then, it was considered more than two systems and it was added a Weibull regression model to the potential law process (PLP). The bimodality characteristic of GOLLFW distribution guaranteed the suitability and a better adjust to tested datas. While, the inclusion of covariables by regression model GOLLFW in the PLP allowed to model systems which traditional counting process, PLP and renewal process, would not fit well.

Keywords: Repairable system; Nonhomogeneous Poisson process; Generalized odd log logistic flexible Weibull; GOLLW regression; Weibull regression.

1 INTRODUÇÃO

O tempo é um importante fator de estudo da Estatística, em particular nas áreas médicas e biológicas (análise de sobrevivência) e nas áreas de engenharia e industrial (confiabilidade) e, essas áreas são de grande interesse de estudos e pesquisa devido sua aplicabilidade na resolução de problemas (COLOSIMO; GIOLO, 2006). A confiabilidade desempenha um papel fundamental no desenvolvimento de produtos de qualidade e na competitividade de mercado, sendo que a durabilidade e a possibilidade de reparo são importantes características vistas pelos clientes na hora da escolha do produto. Calcular a probabilidade de um produto falhar durante o período de garantia ou garantia estendida e quanto tempo é a vida útil do produto até que o cliente precise comprar um novo, são questionamentos importantíssimos dentro de uma empresa, uma vez que seus gastos e lucros estão diretamente relacionados a isso. Os bens duráveis estão entre os produtos mais afetados pelo cenário econômico atual, realizar um conserto ou a troca de uma peça se tornou mais atrativo, fazendo com que o mercado começasse a exigir produtos com maior vida útil e que após uma falha seja possível a realização de um reparo para que o mesmo possa voltar a desempenhar sua função. Produtos com essas características são chamados de sistemas reparáveis (ASCHER; FEINGOLD, 1984).

As máquinas e equipamentos agrícolas, dentre elas as colhedoras de cana-de-açúcar, são exemplos de sistemas reparáveis amplamente usados pelo setor de açúcar e etanol que, dado o seu recente crescimento na utilização, merecem destaque nos estudos de confiabilidade. Até o início dos anos 2000, praticamente todos os canaviais eram colhidos manualmente e dessa forma o uso da queima era essencial, pois permitia um melhor desempenho do cortador. Porém, um conjunto de esforços de órgãos ambientalistas e do próprio setor ajudaram a regulamentar leis que limitaram o uso da queima e já na safra 2009/2010 a colheita mecanizada foi utilizada em 56% dos canaviais da região Centro-Sul (RUDDORF et al., 2010). Enfim, a utilização da colhedora mecânica se tornou indispensável com a regulamentação da Lei Estadual 47.700 que determinou a definitiva proibição em 2014 da queima da cana-de-açúcar (MANARIM, 2017). Esses equipamentos apresentam um alto grau de complexidade dos seus sistemas e são altamente exigidos durante o período de safra que, na região sudeste, compreende os meses de março a novembro. Um dos principais gastos do setor, que é responsável por 2% do Produto Interno Bruto (PIB) do Brasil (CONAB, 2017), são reparos e manutenções com colhedoras. Os valores estão em torno de 70 mil reais por máquina/ano, sem contar a mão de obra para os reparos e o que a máquina deixou de ganhar durante esse período. Considerando uma Usina com uma frota de 100 colhedoras, os gastos podem superar 7 milhões de reais, o que dá indicativos da complexidade e da importância desses sistemas.

O comportamento dos sistemas reparáveis são descritos por repetidas falhas durante o período de tempo em estudo e pode-se assumir um processo de contagem para a análise. Dentre os processos de contagem utilizados para sistemas reparáveis, o processo de Poisson é o mais importante, sendo o processo de Poisson homogêneo (PPH) e o processo de Poisson não homogêneo (PPNH) os principais (GILARDONI; COLOSIMO, 2007), assim como seus casos particulares: processo de renovação (PR) e processo lei de potência (PLP), respectivamente. A variação do processo de contagem utilizado depende do comportamento do sistema ao longo do tempo e, principalmente, do reparo realizado no sistema após uma falha. Barlow e Hunter (1960) descrevem dois tipos de reparos frequentemente usados na literatura, o reparo perfeito e o reparo mínimo e, Pham e Wang (1996) descrevem o reparo imperfeito que é uma condição intermediária aos outros dois tipos de reparo. No processo de renovação os tempos entre as falhas são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) comumente distribuído pela distribuição exponencial e, é usado quando o reparo é perfeito e o sistema volta a condição de novo. No caso de um reparo mínimo que tem como objetivo retornar o sistema a sua função e a uma condição imediatamente igual à anterior a falha o PPNH é o indicado, podendo ser modelado por uma distribuição Weibull e, em especial, o processo lei de potência (RIGDON; BASU, 2000). O reparo imperfeito é uma

condição intermediária aos outros dois tipos de reparo e utiliza como modelagem o processo de renovação generalizado.

Na prática, dificilmente existirão sistemas que sofrem reparos perfeitos e a condição do equipamento retorna à novo, da mesma maneira que é improvável considerar apenas um sistema para análise. A construção do modelo probabilístico para sistemas reparáveis depende da quantidade, do reparo e do comportamento dos sistemas envolvidos. Outro fator importante para a modelagem é o tipo de truncamento, ou seja, como foram coletados os dados. Existem dois tipos de truncamento, o truncamento por tempo, no qual os dados são coletados até um tempo pré-determinado e o número de falhas é aleatório, e o truncamento por falha no qual os dados são coletados até que um número pré-determinado de falhas ocorram e o tempo de coleta não é mais fixo. Decidir o tipo de truncamento usado para o experimento depende do comportamento dos sistemas de um modo geral e do meio onde ele está inserido, por exemplo, se falham muito em pouco tempo ou não e por quanto tempo eles trabalham. O tipo de reparo, de truncamento e a quantidade de sistemas em estudo dão as diretrizes para quais processos de Poisson devem ser testados e para a construção da função de verossimilhança e por consequência a estimação dos parâmetros.

A complexidade de alguns sistemas reparáveis e a quantidade de variáveis que os influenciam faz com que o processo lei de potência não seja suficiente para descrever o comportamento de algumas funções intensidade de falhas. Em uma Usina composta por uma frota de colhedoras de cana-de-açúcar, por exemplo, têm-se vários equipamentos expostos a diferentes condições de solo, motorista, vegetação, mecânico, dentre outros, e a construção da função intensidade de falha para cada sistema pode indicar grande variabilidade no comportamento dessas funções, ou seja, assumindo curvas crescentes, decrescentes, unimodal, bimodal, “forma de U” e constante, por isso, somente a distribuição Weibull usada no PLP não seria a mais indicada para essa modelagem. Como alternativas de modelagem surgiram duas propostas: a inclusão de covariáveis através de um modelo de regressão para explicar a variabilidade presente nos tempos de falha, principalmente os modelos de regressão paramétricos com algumas distribuições clássicas associada a variável resposta, Weibull e exponencial (LAWLESS, 2003), por exemplo, que acrescentam um modelo de regressão nos parâmetros da distribuição Weibull e exponencial usada no PLP e o PR, respectivamente. E, a generalização e modificação de uma distribuição contínua para obter uma nova distribuição de probabilidade com mais parâmetros e mais flexível nas formas da função de risco, baseada no modelo Weibull flexível (BEBBINGTON; LAI; ZITIKIS, 2007), ou seja, a distribuição odd log-logística Weibull flexível generalizada (GOLLFW). Essa nova distribuição utiliza o mesmo método de generalização de distribuições da odd log-logística Weibull generalizada (CORDEIRO et al, 2016) e tem como base a distribuição Weibull flexível.

O objetivo deste trabalho consistiu em apresentar a família odd log-logística generalizada, propor a distribuição odd log-logística Weibull flexível generalizada e mostrar seus casos particulares a odd log-logística Weibull flexível (OLLFW) e o modelo base Weibull flexível (WF) que tem como vantagem a modelagem de dados bimodais e a flexibilização da função intensidade de falha para comportamentos crescente, decrescente, forma de U, entre outros, e uma aplicação desses modelos em dados de confiabilidade. E, analisar um conjunto de dados dos tempos de falha das lâminas Chopper usadas no sistema de corte de colhedoras de cana-de-açúcar de um grupo de Usinas através dos processos de contagem conhecidos para sistemas reparáveis e propor um modelo de regressão paramétrico, ou seja, modelo de regressão Weibull e exponencial para sistemas reparáveis que consiga incorporar a influência das covariáveis presentes nos dados.

REFERÊNCIAS

- ASCHER, H.; FEINGOLD, H. Repairable systems reliability. Lecture notes in statistics, v. 7, 1984.
- BEBBINGTON, M.; LAI, C.-D.; ZITIKIS, R. A exible weibull extension. Reliability Engineering and System Safety, v. 92, p. 719-726, 2007.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. [S.l.]: Edgard Blucher, 2006.
- CONAB – Companhia Nacional de Abastecimento. Disponível em: <<http://www.conab.gov.br>> Acesso em: 13 novembro 2017.
- CORDEIRO, G. M.; ALIZADEH, M.; OZEL, G.; HOSSEINI, B.; ORTEGA, E. M. M.; ALTUN, E. The generalized odd log-logistic family of distributions: properties, regression models and applications. Journal of Statistical Computation and Simulation, 2016.
- GILARDONI, G.; COLOSIMO, E. Optimal maintenance time for repairable systems. Journal of Quality Technology, v. 39, n. 1, p. 48-53, 2007.
- LAWLESS, J. F. Parametric regression models. In: Statistical Models and Methods for Lifetime Data. John Wiley & Sons, Inc., p. 269-339, 2003.
- MANARIM, G. R. Interferentes analíticos provenientes da colheita, carregamento e transporte nas análises tecnológicas de caldo de cana-de-açúcar. Dissertação (Mestrado em Microbiologia Agrícola) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2017.
- RIGDON, S. E.; BASU, A. P. Statistical Methods for the Reliability of Repairable Systems. [S.l.]: Wiley Series in Probability and Statistics, 2000.
- RUDORFF, B. F. T.; AGUIAR, D. A.; SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using landsat data. Remote Sensing, v. 2, n. 4, p. 1057-1076, 2010.

2 DISTRIBUIÇÃO ODD LOG-LOGÍSTICA WEIBULL FLEXÍVEL GENERALIZADA

Resumo

Uma nova distribuição de probabilidade chamada de odd log-logística Weibull flexível (GOLLFW) foi desenvolvida a partir da família odd log-logística generalizada e usando como função base a distribuição Weibull Flexível. O foco dessa nova distribuição e dos seus casos particulares, odd log-logístico Weibull flexível e Weibull flexível, está na aplicação em dados de confiabilidade por ser capaz de acomodar diferentes formas da função de risco, como monótona, unimodal, em forma de banheira, crescente e decrescente. Um novo modelo de regressão paramétrica com base na distribuição GOLLFW foi apresentado como alternativa aos modelos de regressão localização-escala. Foi utilizado o método de máxima verossimilhança para estimação dos parâmetros do modelo em estudo e realizaram-se várias simulações para diferentes configurações de parâmetros, tamanhos de amostra e porcentagens de censura. Uma aplicação a dados de confiabilidade provenientes dos tempos de falha de colhedoras de cana-de-açúcar foi realizada para ilustrar a flexibilidade dos modelos propostos.

Palavras-chave: Distribuição Weibull flexível; Odd log-logística Weibull flexível generalizada; modelo de regressão; simulação com censura.

Abstract

A new probability distribution named generalized odd log-logistic flexible Weibull (GOLLFW) was developed from the generalized odd log-logistic. It was used the flexible Weibull distribution as base function. The focus of this new distribution and their particular cases, odd log-logistic flexible Weibull and flexible Weibull, is to apply confiability datas in the accommodation of different hazard rate form, as monotonous, unimodal, bathtub-shaped, crescent, descending and other possible. A new parametric regression model based on GOLLFW distribution was presented as alternative to the location-scale regression model. It was used the maximum likelihood method to estimate the parameters of the studied model and it was carried out many simulations to different parameter configurations, sample sizes and censoring percentages. An application to confiability datas from failure time of sugarcane harvesters was executed to illustrate the flexibility of the proposed model.

Keywords: Flexible Weibull distribution; Generalized odd log-logistic flexible; Regression model; Simulation with censored.

2.1 Introdução

Análise de sobrevivência ou confiabilidade é a área da estatística destinada ao estudo do tempo, em que a variável de interesse é definida como o tempo até a ocorrência do evento de interesse (falha) e, as principais distribuições de probabilidade usadas para modelar tal variável são exponencial, log-normal, log-logística, Gama e em especial a distribuição Weibull, já que sua função de risco é monótona, crescente ou decrescente (COLOSIMO; GIOLO, 2006). Entretanto, para modelagem de fenômenos que apresentam taxas de falha não-monótonas, como forma de banheira ou unimodal, comuns na confiabilidade e em dados biológicos esta distribuição não é adequada. Um equipamento eletrônico que diminui pra um nível mínimo suas taxas de falha e permanece assim por alguns anos antes de sofrer o efeito do desgaste com o tempo e aumentar é um exemplo da taxa de falha em forma de banheira. As taxas de falha unimodais podem ser observadas em dados com determinadas doenças que atingem um pico após um tempo e depois declinam.

Quando as distribuições de risco não são suficientemente flexíveis para se ajustar a diferentes funções intensidade de falha de um determinado conjunto de dados, novas distribuições mais flexíveis tornam-se necessárias. Nos últimos anos modelos com mais parâmetros em relação aos comumente usados em análise de sobrevivência, foram modificados tendo como base a distribuição Weibull, como exemplo, a Weibull exponenciada (MUDHOLKAR; SRIVASTAVA, 1993), Weibull flexível (BEBBINGTON; LAI; ZITIKIS, 2007), Weibull modificada generalizada (CARRASCO; ORTEGA; CORDEIRO, 2008) e odd log-logística Weibull generalizada (CORDEIRO et al., 2016).

O desenvolvimento de um novo modelo probabilístico com quatro parâmetros, denominado odd log-logístico Weibull flexível generalizado (GOLLFW), se justifica pela flexibilização da sua função de risco, pela importância dos casos particulares (odd log-logístico Weibull flexível e Weibull flexível) e pela aplicabilidade a diversos fenômenos biológicos e problemas de confiabilidade, como o estudo dos tempo de falha das colhedoras de cana-de-açúcar. Além disso, foi definido um modelo de regressão paramétrico para dados de sistemas reparáveis, chamado de modelo de regressão GOLLFW, com base na distribuição proposta.

2.2 Definição do Modelo

Considere uma função de densidade acumulada (fda) base $G(x; \boldsymbol{\tau})$ com vetor de parâmetros $\boldsymbol{\tau}$ da ordem $p \times 1$ de parâmetros desconhecidos. Glean e Lynch (2006) definiram uma nova fda pertencente a família log-logística generalizada com um parâmetro extra $\alpha > 0$ de forma e cuja expressão é dada por

$$H(x) = \frac{G(x; \boldsymbol{\tau})^\alpha}{G(x; \boldsymbol{\tau})^\alpha + \bar{G}(x; \boldsymbol{\tau})^\alpha}, \quad (2.1)$$

em que $\bar{G}(x; \boldsymbol{\tau}) = 1 - G(x; \boldsymbol{\tau})$.

Baseado no método do gerador (T-X) (ALZAATREH; LEE; FAMOYE, 2013), Cordeiro et al. (2016) propuseram uma nova classe de distribuições contínuas denominada odd log-logística generalizada - G (GOLL-G) com fda dada por

$$F(t; \alpha, \theta, \boldsymbol{\tau}) = \int_0^{\frac{G(t; \boldsymbol{\tau})^\theta}{1 - G(t; \boldsymbol{\tau})^\theta}} \frac{\alpha x^{\alpha-1}}{(1-x)^2} dx = \frac{G(t; \boldsymbol{\tau})^{\alpha\theta}}{G(t; \boldsymbol{\tau})^{\alpha\theta} + [1 - G(t; \boldsymbol{\tau})^\theta]^\alpha}, \quad (2.2)$$

em que $\alpha > 0$ e $\theta > 0$ são dois parâmetros extras de forma e, para $\alpha = \theta = 1$ temos a função base $G(x; \boldsymbol{\tau})$. A função densidade de probabilidade (fdp) correspondente a Equação (2.2) é dada por

$$f(t) = \frac{\alpha \theta g(t; \boldsymbol{\tau}) G(t; \boldsymbol{\tau})^{\alpha\theta-1} [1 - G(t; \boldsymbol{\tau})^\theta]^{\alpha-1}}{\left\{ G(t; \boldsymbol{\tau})^\alpha \theta + [1 - G(t; \boldsymbol{\tau})^\theta]^\alpha \right\}^2}, \quad (2.3)$$

em que $g(x; \boldsymbol{\tau})$ é a fdp da função base. A variável aleatória X tem função densidade apresentada na Equação (2.3) e denotado por $X \sim GOLL - G(\alpha, \theta, \boldsymbol{\tau})$. O desenvolvimento da distribuição GOLLFW utiliza o método de generalização, citado anteriormente, e por transitividade a distribuição Weibull, o que justifica seu potencial em ajustar diferentes curvas de risco. A distribuição GOLLFW tem como função base a distribuição Weibull flexível (FW) (BEBBINGTON; LAI; ZITIKIS, 2007). Assim, seja T uma variável aleatória com distribuição Weibull flexível e parâmetros $a, b > 0$. A fda e a fdp são dadas, respectivamente, por

$$G(t; a, b) = 1 - \exp \left\{ -e^{at-b/t} \right\}, \quad (2.4)$$

e

$$g(t; a, b) = (a + b/t^2) \exp(at - b/t) \exp \left\{ -e^{at-b/t} \right\}, \quad t > 0. \quad (2.5)$$

Os parâmetros a e b controlam a forma da distribuição e quando $b = 0$ e $a = \log(\lambda)$ para $\lambda > 0$ na Equação (2.4) tem-se a distribuição exponencial.

Substituindo a função base da Equação (2.4) na função geradora da Equação(2.2) temos a fda da distribuição GOLLFW

$$F(t) = \frac{\{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta}}{\{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta} + [1 - \{1 - \exp[-\kappa_{ab}(t)]\}^\theta]^\alpha}, \quad (2.6)$$

em que $\kappa_{ab}(t) = \exp(at - b/t)$. A fdp da distribuição GOLLFW é dada por

$$\begin{aligned} f(t) &= \alpha\theta \left(a + \frac{b}{t^2}\right) \exp\left[\left(at - \frac{b}{t}\right) - \kappa_{ab}(t)\right] \{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta-1} \times \\ &\quad [1 - \{1 - \exp[-\kappa_{ab}(t)]\}^\theta]^{\alpha-1} \times \\ &\quad \left\{ \{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta} + [1 - \{1 - \exp[-\kappa_{ab}(t)]\}^\theta]^\alpha \right\}^{-2}. \end{aligned} \quad (2.7)$$

Uma importante propriedade da distribuição GOLLFW são seus casos particulares. Por exemplo, se $\theta = 1$ se obtém a distribuição odd log-logística Weibull flexível e para $\alpha = \theta = 1$ temos a distribuição base Weibull flexível.

A função intensidade de falha da distribuição GOLLFW é dada pela seguinte expressão:

$$h(t) = \frac{\alpha\theta \left(a + \frac{b}{t^2}\right) \exp\left[\left(at - \frac{b}{t}\right) - \kappa_{ab}(t)\right] \{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta-1}}{\left[1 - \{1 - \exp[-\kappa_{ab}(t)]\}^\theta\right]^\alpha \left\{ \{1 - \exp[-\kappa_{ab}(t)]\}^{\alpha\theta} + [1 - \{1 - \exp[-\kappa_{ab}(t)]\}^\theta]^\alpha \right\}}. \quad (2.8)$$

Os gráficos contendo algumas formas da função densidade de probabilidade (Figura 2.1) e da função intensidade de falha (Figura 2.2) evidenciam a flexibilidade desse novo modelo. É possível observar o comportamento unimodal e de bimodalidade, dificilmente encontrado em distribuições comumente usadas em confiabilidade (Figura 2.1) e formas crescente, decrescente, forma de banheira e banheira invertida (Figura 2.2) dependendo dos valores que os parâmetros assumem.

Figura 2.1. Gráficos da função densidade da distribuição GOLLFW para diferentes valores de parâmetros. (a) Para diferentes valores de α , a e b com $\theta = 1$. (b) Para diferentes valores de a com $\alpha = 0, 20$, $b = 2$ e $\theta = 3$. (c) Para diferentes valores de a , b , α e θ .

Estimador de máxima verossimilhança

Em problemas de sobrevivência ou confiabilidade é comum a presença de dados censurados, ou seja, observações parciais da variável resposta. Para a simulação adota-se que os dados são censurados a direita. Seja T uma variável aleatória com distribuição (2.7) e com vetor de parâmetros $\tau = (a, b, \alpha, \theta)^T$.

Figura 2.2. Gráfico da função intensidade de falha da distribuição GOLLFW para diferentes valores dos parâmetros. (a) Para diferentes valores de a com $\alpha = 0, 10$, $b = 1, 20$ e $\theta = 1$. (b) Para diferentes valores de a , b , α e θ . (c) Para diferentes valores de a , b , α e θ .

Um mecanismo pra descrever a censura é aquele em que cada indivíduo i é assumido com um tempo de falha T_i e um tempo de censura C_i , onde T_i e C_i são variáveis aleatórias. Suponha que os dados consistam em n observações independentes e $t_i = \min(T_i, C_i)$ para $i = 1, \dots, n$. A estimação dos parâmetros, do ponto de vista paramétrico, é baseada no método de máxima verossimilhança e suas formas assintóticas. Considere $\tau = (a, b, \alpha, \theta)^T$ um vetor de parâmetros 4×1 , a função de log-verossimilhança para para τ é dada por

$$\begin{aligned}
 l(\tau) = & r \log(\alpha \theta) + \sum_{i \in F} \log \left(a + \frac{b}{t_i^2} \right) + \sum_{i \in F} \left[a t_i + \frac{b}{t_i} - \kappa_{ab}(t_i) \right] + (\alpha \theta - 1) \sum_{i \in F} \log \{\omega_i\} \\
 & e - 2 \sum_{i \in F} \log \left\{ \omega_i^{\alpha \theta} + [1 - \omega_i^{\theta}]^{\alpha} \right\} + (\alpha - 1) \sum_{i \in F} \log [1 - \omega_i^{\theta}] \\
 & + \alpha \sum_{i \in C} \log [1 - \omega_i^{\theta}] - \sum_{i \in C} \log \left\{ \omega_i^{\alpha \theta} + [1 - \omega_i^{\theta}]^{\alpha} \right\},
 \end{aligned}$$

em que $\kappa_{ab}(t_i) = \exp(at_i - b/t_i)$, $\omega_i = 1 - \exp[-\kappa_{ab}(t_i)]$, r é o número de falhas, F e C denotam se é tempo de falha ou censura, respectivamente.

As funções escore para os parâmetros a , b , α e θ são dadas por

$$\begin{aligned}
 U_a(\tau) = & \sum_{i \in F} \frac{1}{(a + b/t_i^2)} + \sum_{i \in F} t_i [1 - \kappa_{ab}(t_i)] + (\alpha \theta - 1) \sum_{i \in F} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i) t_i}{\omega_i} \\
 & - 2\alpha \theta \sum_{i \in F} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i) t_i}{\omega_i^{\alpha \theta} + [1 - \omega_i^{\theta}]^{\alpha}} \left\{ \omega_i^{\alpha \theta - 1} - [1 - \omega_i^{\theta}]^{\alpha - 1} \omega_i^{\theta - 1} \right\} \\
 & - (\alpha - 1) \theta \sum_{i \in F} \frac{\omega_i^{\theta} \exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i) t_i}{\omega_i [1 - \omega_i^{\theta}]} - \alpha \theta \sum_{i \in C} \frac{\omega_i^{\theta} \exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i) t_i}{\omega_i [1 - \omega_i^{\theta}]} \\
 & - \alpha \theta \sum_{i \in C} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i) t_i}{\omega_i^{\alpha \theta} + [1 - \omega_i^{\theta}]^{\alpha}} \left\{ \omega_i^{\alpha \theta - 1} - [1 - \omega_i^{\theta}]^{\alpha - 1} \omega_i^{\theta - 1} \right\},
 \end{aligned}$$

$$\begin{aligned}
U_b(\boldsymbol{\tau}) = & \sum_{i \in F} \frac{1}{(t_i^2)(a + b/t_i)} + \sum_{i \in F} \frac{1}{t_i} [1 + \kappa_{ab}(t_i)] + (\alpha\theta - 1) \sum_{i \in F} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i)}{t_i \omega_i} \\
& - 2\alpha\theta \sum_{i \in F} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i)}{t_i(\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha)} \left\{ \omega_i^{\alpha\theta-1} - [1 - \omega_i^\theta]^{\alpha-1} \omega_i^{\theta-1} \right\} \\
& - (\alpha - 1)\theta \sum_{i \in F} \frac{\omega_i^\theta \exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i)}{t_i \omega_i [1 - \omega_i^\theta]} - \alpha\theta \sum_{i \in C} \frac{\omega_i^\theta \exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i)}{t_i \omega_i [1 - \omega_i^\theta]} \\
& - \alpha\theta \sum_{i \in C} \frac{\exp[-\kappa_{ab}(t_i)] \kappa_{ab}(t_i)}{t_i(\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha)} \left\{ \omega_i^{\alpha\theta-1} - [1 - \omega_i^\theta]^{\alpha-1} \omega_i^{\theta-1} \right\},
\end{aligned}$$

$$\begin{aligned}
U_\alpha(\boldsymbol{\tau}) = & \frac{r}{\alpha} - 2 \sum_{i \in F} \frac{1}{\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha} \left\{ \theta \omega_i^{\alpha\theta} \log[\omega_i] + [1 - \omega_i^\theta]^\alpha \log[1 - \omega_i^\theta] \right\} \\
& + \theta \sum_{i \in F} \log[\omega_i] + \sum_{i \in F} \log[1 - \omega_i^\theta] + \sum_{i \in C} \log[1 - \omega_i^\theta] \\
& - \sum_{i \in C} \frac{1}{\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha} \left\{ \theta [\omega_i^\alpha]^{\theta-1} \omega_i^\alpha \log[\omega_i] + [1 - \omega_i^\theta]^\alpha \log[1 - \omega_i^\theta] \right\},
\end{aligned}$$

$$\begin{aligned}
U_\theta(\boldsymbol{\tau}) = & \frac{r}{\theta} - 2\alpha \sum_{i \in F} \frac{\omega_i^\theta \log[\omega_i]}{\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha} \left\{ [\omega_i^\theta]^{\alpha-1} - [1 - \omega_i^\theta]^{\alpha-1} \right\} + \alpha \sum_{i \in F} \log[\omega_i] \\
& - (\alpha - 1) \sum_{i \in F} \frac{\omega_i^\theta \log[\omega_i]}{1 - \omega_i^\theta} - \alpha \sum_{i \in C} \frac{1}{1 - \omega_i^\theta} \left\{ \omega_i^\theta \log[\omega_i] \right\} \\
& - \sum_{i \in C} \frac{\omega_i^\theta \log[\omega_i]}{\omega_i^{\alpha\theta} + [1 - \omega_i^\theta]^\alpha} \left\{ [\omega_i^\theta]^{\alpha-1} - [1 - \omega_i^\theta]^{\alpha-1} \right\}.
\end{aligned}$$

O estimador de máxima verossimilhança $\hat{\boldsymbol{\tau}}$ de $\boldsymbol{\tau}$ é obtido resolvendo as equações não lineares $U_a(\boldsymbol{\tau}) = 0$, $U_b(\boldsymbol{\tau}) = 0$, $U_\alpha(\boldsymbol{\tau}) = 0$ e $U_\theta(\boldsymbol{\tau}) = 0$. Entretanto, essas equações não podem ser resolvidas analiticamente, e a utilização de software estatístico se torna imprescindível para resolvê-las numericamente. O método iterativo usado no presente trabalho é o algoritmo Newton-Raphson e o software utilizado foi o R.

Para intervalos de confiança de (a, b, α, θ) e testes de hipótese para os parâmetros é necessário a obtenção da matriz de informação observada, porém, o cálculo dessa matriz também requer métodos numéricos. A matriz de informação observada $J(\boldsymbol{\tau})$ de dimensões 4×4 é:

$$J(\boldsymbol{\tau}) = - \begin{pmatrix} \mathbf{L}_{aa} & \mathbf{L}_{ab} & \mathbf{L}_{a\alpha} & \mathbf{L}_{a\theta} \\ \cdot & \mathbf{L}_{bb} & \mathbf{L}_{b\alpha} & \mathbf{L}_{b\theta} \\ \cdot & \cdot & \mathbf{L}_{\alpha\alpha} & \mathbf{L}_{\alpha\theta} \\ \cdot & \cdot & \cdot & \mathbf{L}_{\theta\theta} \end{pmatrix},$$

em que cada elemento dessa matriz é calculado numericamente.

Uma condição de regularidade para os parâmetros é que, a distribuição assintótica de $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ é $N_4(0, I(\boldsymbol{\tau})^{-1})$, em que $I(\boldsymbol{\tau})$ é a matriz de informação esperada. Esse comportamento assintótico é válido se $I(\boldsymbol{\tau})$ é substituída por $J(\boldsymbol{\tau})$, isto é, a matriz de informação observada é avaliada no $\hat{\boldsymbol{\tau}}$. A distribuição assintótica normal multivariada $N_4(0, J(\hat{\boldsymbol{\tau}})^{-1})$ pode ser usada para construir intervalos de confiança aproximados para os parâmetros individuais e regiões de confiança. A normalidade assintótica é usada para testar o ajuste da distribuição GOLLFW e compará-la com outros modelos encaixados, como o modelo OLLFW, usando um dos três métodos bem conhecidos: razão de verossimilhanças, Wald e estatística escore (Rao). Pode-se calcular as verossimilhanças restrita e não restrita para construir o teste de razão de verossimilhanças e testar alguns submodelos da distribuição GOLLFW. Por exemplo, pode-se testar se o ajuste da distribuição GOLLFW com relação ao ajuste dos submodelos OLLFW e FW

é estatisticamente superior. Sendo assim, a hipótese testada é $H_0 : \tau = \tau_0$ versus $H : \tau \neq \tau_0$. Para a comparação com o submodelo OLLFW citado, a hipótese testada é $H_0 : \theta = 1$ versus $H : \theta \neq 1$ e nesse caso, a estatística do teste de razão de verossimilhanças é dada por $w = 2\{l(\hat{a}, \hat{b}, \hat{\alpha}, \hat{\theta}) - l(\tilde{a}, \tilde{b}, \tilde{\alpha}, 1)\}$, em que $\hat{a}$, $\hat{b}$, $\hat{\alpha}$ e $\hat{\theta}$ são os estimadores de máxima verossimilhança sob H e $\tilde{a}$, $\tilde{b}$ e $\tilde{\alpha}$ são os estimadores de máxima verossimilhança sob H_0 .

2.3 Simulação

Algumas propriedades dos estimadores de máxima verossimilhança (EMVs) são avaliadas por meio de estudos de simulação e para isso a inversa da distribuição acumulada GOLLFW. Assim, se T é uma variável aleatória com função densidade apresentada na Equação (2.7) e denotada por $T \sim \text{GOLLFW}(a, b, \alpha, \theta)$, invertendo a Equação (2.6), tem-se $t = Q(u) = F^{-1}(u)$, ou seja,

$$t = Q(u) = Q_{FW} \left(\left[\frac{\left(\frac{u}{1-u} \right)^{\frac{1}{\alpha}}}{1 + \left(\frac{u}{1-u} \right)^{\frac{1}{\alpha}}} \right]^{\frac{1}{\theta}} \right), \quad (2.9)$$

em que $Q_{FW}(u) = G^{-1}(t)$ é a função quantil da distribuição FW.

A simulação da distribuição GOLLFW é feita da seguinte forma:

- i. Calcular a função inversa $F^{-1}(\cdot)$ da distribuição (2.5).
- ii. Gerar $u \sim \text{Uniforme}(0,1)$.
- iii. Aplicar u em $F^{-1}(u) = Q(u)$, da Equação (2.9).
- iv. Definir $t = Q(u)$, os valores gerados da distribuição GOLLFW, em que $Q(u)$ é a função quantile da distribuição apresentada na Equação (2.5).

Considerando diferentes tamanhos de amostras para cada replicação, são calculadas as EMVs $\hat{\alpha}$, $\hat{\theta}$, $\hat{a}$ e $\hat{b}$. Este processo é repetido 1000 vezes e calculadas as médias das estimativas (MEs), o viés e os erros quadráticos médios (EQMs). Nesse estudo, foram considerados $\alpha = 0,25$, $\theta = 5, 50$, $a = 3, 50$ e $b = 1, 00$ e as estimativas de α , θ , a e b são obtidas resolvendo as equações não lineares $U_{\alpha}(\tau) = 0$, $U_{\theta}(\tau) = 0$, $U_a(\tau) = 0$, $U_b(\tau) = 0$. Os resultados da simulação de Monte Carlo (Tabela 2.1) indicam que os EQMs dos EMVs de α , θ , a e b tendem a zero à medida que aumenta o tamanho amostral, como esperado pela teoria assintótica. A curva da densidade estimada baseada nas MEs de 1000 amostras dos parâmetros α , θ , a e b , respectivamente, e $n = 80, 250$ e 500 para esse estudo (Figura 2.3) converge para a curva da densidade esperada à medida que n aumenta.

Estudo de simulação para censura aleatória:

Da mesma maneira feita para dados completos considera-se o estudo de simulação na presença de dados censurados e, seja uma simulação para diferentes tamanhos de amostras e porcentagens de censura de 0%, 10% e 20%. Assume-se, também, que os valores dos parâmetros são $\alpha = 0,25$, $\theta = 5, 50$, $a = 3, 50$ e $b = 1, 00$. Os tempos de censura são gerados considerando o mecanismo de censura aleatória da seguinte forma:

- i. Calcular a função inversa $F^{-1}(\cdot)$ da distribuição (2.5).
- ii. Gerar $u \sim \text{Uniforme}(0,1)$.
- iii. Aplicar u em $F^{-1}(u) = Q(u)$, da Equação (2.9).

Tabela 2.1. MEs, viés e EQMs para as estimativas dos parâmetros da GOLLFW

n	Parâmetro	MEs	Viés	EQMs
80	α	0,2745	0,0245	0,0267
	θ	7,5244	2,0244	47,1398
	a	3,6849	0,1849	0,4170
	b	1,1634	0,1634	0,4325
250	α	0,2557	0,0057	0,0044
	θ	5,9165	0,4165	7,6824
	a	3,5663	0,0633	0,1256
	b	1,0536	0,0536	0,0644
500	α	0,2539	0,0039	0,0019
	θ	5,6201	0,1201	2,1671
	a	3,5223	0,0223	0,0579
	b	1,0224	0,0224	0,0284

Figura 2.3. Função densidade de probabilidade da GOLLFW para os verdadeiros parâmetros e para as MEs no estudo de simulação.

- iv. Então $t = Q(u)$ são os valores gerados da distribuição GOLLFW, em que $Q(u)$ é a função quantile da distribuição da Equação (2.5).
- v. Criar um vetor δ de dimensão n contendo 1's.
- vi. Definir uma amostra de tamanho np (em que p representa o percentual fixo de censuras) de índices $1, \dots, n$ e alocados em w .
- vii. Para $j = 1, \dots, \text{length}(w)$, definir $t[w[j]] = \text{uniforme}(0, t[w[j]])$ e $\delta[w[j]] = 0$. Assim, w contém a posição onde t recebeu as censuras com porcentagem fixa p .

Os resultados da simulação (Tabela 2.2) indicam que os EQMs das EMVs de α , θ , a e b tendem a zero à medida que aumenta o tamanho amostral e que, ao aumentar o tamanho da amostra e diminuir o percentual de censura, as estimativas estão mais próximas dos verdadeiros valores com EQMs menores, conforme esperado pela teoria assintótica.

Nas Figuras 2.4 e 2.5 são apresentadas as densidades estimadas baseadas em 1000 amostras de MEs dos parâmetros a , b , α e θ , respectivamente, e $n = 80, 250$ e 500 e diferentes percentuais de censura. Para esse estudo foram consideradas 10% e 20% de censura e nota-se que esses gráficos estão de acordo com a teoria assintótica para os EMVs, ou seja, indicam uma convergência da curva estimada para a esperada à medida que n aumenta, mesmo contendo dados censurados.

Tabela 2.2. MEs e (EQMs) para as estimativas dos parâmetros da GOLLFW.

n	Parâmetros	0%	10%	20%
80	α	0,2745 (0,0267)	0,2683 (0,0249)	0,2648 (0,0238)
	θ	7,5244 (47,1398)	8,1450 (51,1536)	8,9646 (81,8937)
	a	3,6849 (0,4170)	3,6712 (0,4714)	3,6296 (0,4864)
	b	1,1634 (0,4325)	1,1104 (0,4087)	1,0447 (0,2507)
250	α	0,2557 (0,0044)	0,2611 (0,0056)	0,2601 (0,0057)
	θ	5,9165 (7,6824)	6,4892 (19,0287)	6,7054 (12,9884)
	a	3,5663 (0,1256)	3,5015 (0,1444)	3,4786 (0,1490)
	b	1,0536 (0,0644)	0,9963 (0,0626)	0,9632 (0,0636)
500	α	0,2539 (0,0019)	0,2549 (0,0022)	0,2577 (0,0027)
	θ	5,6201 (2,1671)	5,9389 (3,1489)	5,9609 (2,8700)
	a	3,5223 (0,0579)	3,4855 (0,0542)	3,4754 (0,0672)
	b	1,0224 (0,0284)	0,9831 (0,0249)	0,9734 (0,0256)

Figura 2.4. Função densidade de probabilidade da GOLLFW para os verdadeiros parâmetros e para as MEs no estudo de simulação com 10% de censura. **Figura 2.5.** Função densidade de probabilidade da GOLLFW para os verdadeiros parâmetros e para as MEs no estudo de simulação com 20% de censura.

2.4 Modelo de Regressão GOLLFW

Modelos de regressão podem ser propostos de diferentes formas em análise de confiabilidade. Por exemplo, o modelo de regressão locação-escala apresentado por Lawless (2002) e o modelo semi-paramétrico de Cox (1972) são frequentemente utilizados para ensaios clínicos. Para a distribuição GOLLFW é proposto um modelo de regressão paramétrica com dados censurados e denotado por modelo

de regressão GOLLFW para a análise de tempos de confiabilidade, como alternativa a utilização do modelo de regressão localização-escala. Na prática, existem muitas covariáveis influenciando a variável resposta tempo de falha, o que torna importante incorporar essa informação na análise utilizando uma abordagem baseada em um modelo de regressão.

A análise de regressão para a distribuição do tempo de falha da variável aleatória T , requer a especificação da distribuição dessa variável e um vetor de covariáveis denotado por $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$. Nesse trabalho, os parâmetros a e b estão relacionados com as covariáveis por meio da função de ligação log-linear $a_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)$ e $b_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)$, $i = 1, \dots, n$ respectivamente, em que $\boldsymbol{\beta}_1 = (\beta_{10}, \dots, \beta_{1p})^T$ e $\boldsymbol{\beta}_2 = (\beta_{20}, \dots, \beta_{2p})^T$ denotam os vetores de coeficientes da regressão e $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$. Assim, esse modelo pode ser usado em diversas situações com os parâmetros a e b dependendo de $\mathbf{x}$, a função de sobrevivência $(T|\mathbf{x})$ é dada por

$$S(t|\mathbf{x}) = \frac{\left[1 - \{1 - \exp[-\kappa(t|\mathbf{x})]\}^\theta\right]^\alpha}{\{1 - \exp[-\kappa(t|\mathbf{x})]\}^{\alpha\theta} + \left[1 - \{1 - \exp[-\kappa(t|\mathbf{x})]\}^\theta\right]^\alpha}, \quad (2.10)$$

em que

$$\kappa(t|\mathbf{x}) = \exp \left[\exp(\mathbf{x}^T \boldsymbol{\beta}_1) t - \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_2)}{t} \right].$$

Considere a amostra $(t_1, \mathbf{x}_1), \dots, (t_n, \mathbf{x}_n)$ de n observações independentes, em que a variável resposta é definida por $y_i = \min\{t_i, c_i\}$. Assume-se que os tempos de censura e os tempos de falha observados são independentes e seja F e C o conjunto de indivíduos para o qual t_i é tempo de falha ou censura, respectivamente. As técnicas usuais para o cálculo da verossimilhança podem ser aplicadas aqui e a função log-verossimilhança para um vetor de parâmetros $\boldsymbol{\tau} = (\alpha, \theta, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ para o modelo da Equação (2.10) tem a seguinte forma

$$l(\boldsymbol{\theta}) = \sum_{i \in F} l_i(\boldsymbol{\tau}) + \sum_{i \in C} l_i^{(c)}(\boldsymbol{\tau}),$$

em que $l_i(\boldsymbol{\tau}) = \log[f(t_i|\mathbf{x}_i)]$, $l_i^{(c)}(\boldsymbol{\theta}) = \log[S(t_i|\mathbf{x}_i)]$, $f(t_i|\mathbf{x}_i)$ e $S(t_i|\mathbf{x}_i)$ são a função densidade e de sobrevivência de T_i , respectivamente. A função de log-verossimilhança total para $\boldsymbol{\theta}$ reduz à

$$\begin{aligned} l(\boldsymbol{\tau}) &= r \log(\alpha \theta) + \sum_{i \in F} \log \left[\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1) + \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)}{t_i^2} \right] + \sum_{i \in F} \left[\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1) + \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)}{t_i} \right] - \\ &\quad \sum_{i \in F} \kappa(t_i|\mathbf{x}_i) + (\alpha \theta - 1) \sum_{i \in F} \log\{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\} + \\ &\quad (\alpha - 1) \sum_{i \in F} \log[1 - \{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^\theta] - \\ &\quad 2 \sum_{i \in F} \log \left\{ \{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^{\alpha\theta} + \left[1 - \{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^\theta\right]^\alpha \right\} + \\ &\quad \sum_{i \in C} \log \left\{ \frac{\left[1 - \{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^\theta\right]^\alpha}{\{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^{\alpha\theta} + \left[1 - \{1 - \exp[-\kappa(t_i|\mathbf{x}_i)]\}^\theta\right]^\alpha} \right\}, \end{aligned} \quad (2.11)$$

em que $\kappa(t_i|\mathbf{x}_i)$ é definido na Equação (2.10), e r é o número de falhas.

As estimativas de máxima verossimilhança (EMV) $\hat{\boldsymbol{\theta}}$ do vetor de parâmetros são calculados maximizando a função de verossimilhança da Equação (2.11), sendo usado o software R para calcular essas estimativas $\hat{\boldsymbol{\theta}}$ com os valores iniciais para $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ tomados a partir do ajuste do modelo de regressão da WF com $\alpha = \theta = 1$.

2.5 Aplicação

As análises a seguir referem-se a dados de colhedoras de cana-de-açúcar coletados em uma usina produtora de açúcar e etanol no Estado de São Paulo durante toda a safra de 2014/2015. Foram analisados os tempos 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, com o tempo contado em dias. As covariáveis são tempo até a falha da peça t_i e agente causador da falha x_{i1} , por exemplo, desgaste natural ou quebra por outros motivos. Essa análise tem por objetivo descrever o comportamento da lâmina Chopper em duas colhedoras de cana-de-açúcar na mesma usina e expostas ao mesmo tempo de trabalho e verificar a qualidade do ajuste dos modelos propostos.

O modelo de regressão usado é

$$\kappa(t_i|\mathbf{x}_i) = \exp \left[\exp(\beta_{10} + \beta_{11}x_{i1}) t_i - \frac{\exp(\beta_{20} + \beta_{21}x_{i1})}{t_i} \right],$$

em que a variável t_i segue uma distribuição GOLLFW dada pela Equação (2.10), $i = 1, 2, \dots, n$.

Foram considerados todos os tempos de falhas das duas colhedoras e feito uma análise conjunta delas. Foi construído o gráfico de TTT-plot (Figura 2.6 (a)) e a densidade estimada com os modelos propostos GOLLFW, OLLFW e FW (Figura 2.6 (b)) e encontrado as estimativas dos parâmetros, os erros de estimação e algumas estatísticas dos critérios de seleção do modelo (Tabela 2.3).

Figura 2.6. Análise dos tempos de falha das colhedoras. (a) TTT-plot. (b) Função densidade empírica e estimada para os tempos de falha das colhedoras.

Tabela 2.3. EMVs para os parâmetros para os dados das colhedoras de cana-de-açúcar e critérios de informação.

Modelo	a	b	α	θ	AIC	CAIC	BIC
GOLLFW	0,1291 (8,96x10 ⁻⁶)	1,1724 (1,77x10 ⁻⁵)	0,1417 (0,0820)	18,9900 (0,00001)	205,9	207,4	211,8
OLLFW	0,1190 (0,0270)	12,1480 (5,4529)	0,2641 (0,1140)	1 (-)	204,2	208,6	208,6
FW	0,0486 (0,0076)	3,3912 (0,7367)	1 (-)	1 (-)	1794,8	1795,0	1802,6

A bimodalidade presente no modelo estimado, sugere o uso das distribuições GOLLFW e OLLFW, por exemplo, já os critérios de seleção de modelos critério de informação de Akaike - AIC

(AKAIKE, 1974), critério de informação de Akaike corrigido - CAIC (BOZDOGAN, 1987) e critério de informação Bayesiano - BIC (SCHWARZ et al, 1978) indicam que o modelo OLLFW tem o melhor ajuste (Tabela 2.3). Os valores das estimativas dos parâmetros dos modelos testados são apresentados na Tabela 2.3 e a qualidade dos ajustem pode ser verificado na curva da função densidade (Figura 2.6), na curva de sobrevivência (Figura 2.7 (a)) e na curva da intensidade de falha estimada e empírica (Figura 2.7 (b)).

Figura 2.7. Análise dos tempos de falha das colhedoras. (a) Função de sobrevivência estimada e empírica. (b) Função intensidade de falha estimada.

A Tabela 2.4 apresenta as EMVs para os parâmetros dos modelos de regressão GOLLWF, OLLWF e FW. O teste de razão de verossimilhanças para comparar os modelos de regressão GOLLWF, OLLWF e FW (Tabela 2.5) apresenta indicações favoráveis ao modelo OLLWF, para este o $\beta_{.1}$ não é significativo ($p > 0,05$) indicando que não existe uma grande diferença para os tempos de falhas da colhedora 1 e da colhedora 2.

A curva empírica e a curva estimada da função de sobrevivência do modelo de regressão OLLFW são apresentadas na Figura 2.8 (a) e conclui-se que o modelo de regressão OLLFW se ajusta bem a esses dados. Na Figura 2.8 (b) é apresentado a curva da função de intensidade de falha estimada e nota-se que ela possui uma forma não comum e que o modelo OLLFW é capaz de modelar bem.

2.6 Conclusão

A nova distribuição de probabilidade GOLLFW apresentou como principais ganhos sua flexibilidade e dos seus diferentes casos particulares OLLFW e FW e, principalmente, a modelagem de dados que apresentam bimodalidade. A inclusão de covariáveis gerando o modelo de regressão GOLLFW como alternativa aos modelos de regressão locação-escala foi um ganho em relação aos modelos tradicionais de confiabilidade. Para o problema analisado dos tempos de falha das colhedoras de cana-de-açúcar, a distribuição OLLFW foi a que melhor se ajustou aos dados, e a rejeição do parâmetro $\beta_{.1}$ mostrou que a variável contendo o agente causador da falha, ou seja, desgaste natural ou quebra por outros motivos não influencia na durabilidade dos produtos.

Tabela 2.4. EMVs, SEs, p -valor e estatísticas para os modelos de regressão ajustados para os dados de colhedoras de cana-de-açúcar.

GOLLWF						
Parâmetros	Estimativa	SE	p-valor	AIC	BIC	CAIC
α	0,1271	0,0210	-	207,9	217,1	213,7
θ	7,1794	0,1675	-			
β_{10}	-1,8246	0,0158	$<<0,0001$			
β_{11}	-0,2362	0,1065	0,0335			
β_{20}	1,6541	0,0127	$<<0,0001$			
β_{21}	-0,6531	0,1592	0,0002			
OLLWF						
Parâmetro	Estimativa	SE	p-valor	AIC	BIC	CAIC
α	0,2274	0,0973	-	207,2	214,9	211,6
θ	(1)	(-)	-			
β_{10}	-1,8515	0,2507	$<0,0001$			
β_{11}	-0,3237	0,1636	0,0562			
β_{20}	2,9824	0,4666	$<0,0001$			
β_{21}	-0,5918	0,4155	0,1637			
FW						
Parâmetro	Estimativa	SE	p-valor	AIC	BIC	CAIC
α	(1)	(-)	-	214,9	221,0	218,0
θ	(1)	(-)	-			
β_{10}	-2,9199	0,2266	$<<0,0001$			
β_{11}	-0,2578	0,3251	0,4334			
β_{20}	1,5863	0,2635	$<0,0001$			
β_{21}	-0,7031	0,4258	0,1081			

Tabela 2.5. Teste de razão de verossimilhanças para dados de colhedoras de cana-de-açúcar.

Modelos	Hipóteses	Estatística w	p -valor
GOLLFW vs FW	$H_0 : \alpha = \theta = 1$ vs $H_1 : H_0$ é falso	11,0	0,0040
GOLLFW vs OLLFW	$H_0 : \theta = 1$ vs $H_1 : H_0$ é falso	1,55	0,2125

(a)

(b)

Figura 2.8. Distribuição OLLFW (a) Função de sobrevivência empírica e estimada para os tempos de falha de colhedoras de cana-de-açúcar. (b) Função intensidade de falha empírica e estimada para os tempos de falha de colhedoras de cana-de-açúcar.

REFERÊNCIAS

- AKAIKE, H., A new look at the statistical model identification. IEEE transactions on automatic control, v. 19, p. 716-723, 1974.
- ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. METRON, v. 71, n. 1, p. 63-79, 2013.
- BEBBINGTON, M.; LAI, C.-D.; ZITIKIS, R. A exible weibull extension. Reliability Engineering and System Safety, v. 92, p. 719-726, 2007.
- BOZDOGAN, H., Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. Psychometrika, v. 52, p. 345-370, 1987
- CARRASCO, J.; ORTEGA, E. M. M.; CORDEIRO, G. M. A generalized modied weibull distribution for lifetime modeling. 2008.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. [S.l.]: Edgard Blucher, 2006.
- CORDEIRO, G. M.; ALIZADEH, M.; OZEL, G.; HOSSEINI, B.; ORTEGA, E. M. M.; ALTUN, E. The generalized odd log-logistic family of distributions: properties, regression models and applications. Journal of Statistical Computation and Simulation, 2016.
- COX, D., Regression models and life-tables. Journal of the Royal Statistical Society - Series B, v. 34, p. 187-220, 1972.
- GLEATON, J.; LYNCH, J. Properties of generalized log-logistic families of lifetime distributions. Journal of Probability and Statistical Science, v. 4, n.1, p. 51-64, 2006.
- LAWLESS, J. F. Parametric regression models. In: Statistical Models and Methods for Lifetime Data. John Wiley & Sons, p. 269-339, 2003.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability, IEEE, v. 42, n. 2, p. 299-302, 1993.
- SCHWARZ, G. et al., Estimatiog the dimension os a model. The annals of statistics, v. 6, p. 461-464, 1978.

3 ANÁLISE DE CONFIABILIDADE PARA UM ÚNICO SISTEMA REPARÁVEL

Resumo

A análise de um sistema reparável exige entender as características que o geram, como o tipo de truncamento e o tipo de reparo, sendo o primeiro a forma como os dados são coletados e o segundo, fundamental para a construção do modelo que descreve o sistema, para o processo de estimação e as inferências. Por se tratar de um processo de contagem, os processos de Poisson homogêneo (PPH) e o processo de Poisson não-homogêneo (PPNH), assim como seus casos particulares, processo de renovação e processo lei de potência, respectivamente, modelam a maior parte dos problemas. Uma aplicação com dados de tempo de falha de quatro colhedoras de cana-de-açúcar mostraram o ajuste do PPH e do PPNH. A análise gráfica dos dados das colhedoras apresentou indicativos de diferentes comportamentos da função intensidade de falha ao longo do tempo e o teste de razão de verossimilhança comparou os dois processos de contagem em estudo para cada uma delas. Após o ajuste e verificação da adequação do modelo, foi estimado o número esperado de falhas para os sistemas modeladas por um processo lei de potência.

Palavras-chave: sistema reparável; processo de Poisson; processo de renovação; processo lei de potência; truncamento por tempo.

Abstract

The repairable system analysis demands to understand the characteristics that cause it, for example, truncation and repair type. Truncation is the way data is collected and repair type is fundamental to model construction, estimation process and inferences. Because it is a counting process, the homogeneous Poisson process (HPP) and the non homogeneous Poisson process (NHPP), as well their particular cases, renewal process and power law process, respectively, modelate the most part of problems. An application with failure time data of four sugarcane harvesters showed the HPP and NHPP adjusted well. The graphic analysis of harvesters indicated different behaviours of fail intensity function over time and the likelihood ratio test compared both process for each them. After the adjustment and model verification, it was estimated the expected fail number for the systems modeled by power law process.

Keywords: repairable system; Poisson process; renewal process; power law process; failure time.

3.1 Introdução

Um sistema pode ser a combinação de partes reunidas para obter um resultado. Ascher e Feingold (1984) definem que um sistema é um conjunto de duas ou mais unidades que desempenham uma ou mais funções e podem ser classificados como reparáveis e não reparáveis. Sistema não reparável após um problema não pode ser restaurado à sua função original, por exemplo, uma lâmpada queimada. Sistema reparável após uma falha admite que a condição de funcionamento pode ser restabelecida por algum processo de reparo e sem necessariamente substituição dos sistema inteiro, tendo como exemplos, máquinas e equipamentos eletrônicos (RIGDON; BASU, 2000). A natureza do reparo realizado no sistema após a falha é de vital importância na construção do modelo que o descreve, uma vez que o tipo de reparo realizado no sistema influencia no tempo de durabilidade restante do equipamento. Sendo assim, pode-se classificar o reparo em três tipos:

- Reparo Perfeito: é dito “*tão bom quanto novo*” e o sistema volta a condição de novo (BARLOW; HUNTER, 1960);
- Reparo Mínimo: é dito “*tão ruim quanto velho*” e o sistema volta a condição imediatamente anterior a falha (BARLOW; HUNTER, 1960);
- Reparo Imperfeito: é uma condição intermediária aos outros dois tipos de reparos (PHAM; WANG, 1996).

Os sistemas reparáveis após apresentarem falhas têm sua condição de operação restabelecida por uma ação de reparo quantas vezes for necessário, e assim, não é mais válida a suposição de independência entre os tempos de falha, o que inviabiliza a utilização de técnicas tradicionais de confiabilidade para sistema não reparável (COLOSIMO; GIOLO, 2006). Pode-se definir o tempo médio até a primeira falha do sistema como $TMPF$, o tempo médio de reparo como TMR , e também, o tempo médio entre as falhas $TMEF$, que é medido do momento que ocorre o reparo do sistema até a próxima falha, portanto, não se leva em consideração o tempo de reparo do sistema. A realização de um reparo perfeito informa que o tempo médio até a primeira falha é igual ao tempo médio entre as falhas ($TMPF = TMEF$), ou seja, garante que os comportamentos das falhas são independentes e podem ser descritos por uma distribuição de probabilidade exponencial. Porém, na realização de um reparo mínimo o tempo médio até a primeira falha é diferente do tempo médio entre as falhas ($TMPF \neq TMEF$) o que acarreta na dependência entre os tempos e inviabiliza a utilização da distribuição exponencial para descrevê-los. A modelagem utilizada para o reparo mínimo é feita por meio de um processo de Poisson não-homogêneo, quando o sistema está se deteriorando ou melhorando ao longo do tempo que é o comportamento frequentemente encontrado em sistemas reais.

Um importante ponto de atenção na teoria de sistemas reparáveis é a diferença na interpretação de censura e truncamento, já que esses conceitos geralmente são confundidos devido suas semelhanças. Censura é a observação parcial da variável de interesse e ocorre quando se tem uma informação mas ela é incompleta, sendo utilizada principalmente em sistemas não reparáveis. No caso de sistema reparável o conceito utilizado é de truncamento, ou seja, como a coleta de dados é realizada, sendo que o tipo de truncamento do sistema influencia na construção da função de verossimilhança para o modelo. Rigdon e Basu (2000) classificam truncamento em dois tipos:

- Truncamento por falha: é quando o encerramento do estudo do sistema ocorre após um número pré-determinado de falhas;
- Truncamento por tempo: é quando o estudo do sistema acaba depois de decorrido um período de tempo pré-determinado.

A escolha da técnica paramétrica para a modelagem dos tempos de falha depende entre outras coisas do reparo realizado, sendo assim, Ascher e Feingold (1984) apresentam o processo de renovação e o processo de Poisson não-homogêneo para reparo perfeito e reparo mínimo, respectivamente, que foram discutidos posteriormente por Rigdon e Basu (2000). Já Kijima e Sumita (1986) consideram a realização de um reparo imperfeito e apresentam o processo de renovação generalizado, entretanto essa modelagem não será discutida no presente trabalho.

Um estudo preliminar dos tempos de falhas da lâmina Chopper presente no sistema de corte de colhedoras de cana-de-açúcar através de técnicas apresentadas para análise de um sistema reparável, considerando um truncamento por tempo e um reparo mínimo do sistema mostraram que o PPH e o PPNH, assim como seus casos particulares, modelam bem esse tipo de situação e dão indicativos para a construção do modelo quando são considerados mais de um sistema reparável.

3.2 Material e Método

As análises de falhas de um sistema reparável sempre envolvem um processo de contagem, ou seja, um modelo estocástico que descreve a ocorrência de eventos no tempo (RIGDON; BASU, 2000). Suponha $N(t)$ uma variável aleatória que denota o número de falhas no intervalo $[0, t]$ e seja T uma variável aleatória que representa o tempo de falha, quando N tem como argumento um intervalo $N(a, b]$, por exemplo, ele descreve o número de falhas que ocorreram neste intervalo e pode-se reescrevê-lo como $N(a, b] = N(b) - N(a)$. Um processo de contagem deve satisfazer as seguintes propriedades:

1. $N(t) \geq 0$ e $N(0) = 0$;
2. $N(t)$ é um número inteiro;
3. $t \mapsto N(t)$ é contínuo á direita;
4. $\Delta N(t) = \lim_{h \rightarrow 0} [N(t) - N(t - h)] = 0$.

A função média de um processo de contagem é definida como a esperança do número de falhas, ou seja,

$$\Lambda(t) = E(N(t)) = \int_0^t \lambda(u) du, \quad (3.1)$$

em que $\lambda(u)$ é a função intensidade de falha, que será definida posteriormente, e $\Lambda(t)$ é uma função crescente que representa o número esperado de falhas até o tempo t . Quando $\Lambda(t)$ é uma função diferenciável define-se a taxa de falha acumulada (*Rate of occurrence of failures - ROCOF*) como

$$\mu(t) = \frac{d}{dt} \Lambda(t). \quad (3.2)$$

A função ROCOF é interpretada como a taxa instantânea de mudança no número esperado de falhas. Outra importante definição para um processo pontual é a função intensidade (RIGDON; BASU, 2000) que é dada por

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(N(t, t + \Delta t] \geq 1)}{\Delta t}, \quad (3.3)$$

e definida como a probabilidade de falhar em um pequeno intervalo.

É intuitivo associar a função de risco $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$ com a função intensidade, entretanto, a diferença entre elas está na interpretação, uma vez que a função de risco é o limite de uma probabilidade condicional (COLOSIMO; GIOLO, 2006) e a função intensidade não. Segundo Rigdon e Basu (2000) a função de risco é a probabilidade condicional de uma e, apenas uma falha ocorrer em um pequeno intervalo de tempo, ou seja, está condicionada à sobrevivência até o início de intervalo e a função intensidade não é condicionada a nada. Nota-se que a função intensidade está associada à confiabilidade de um sistema reparável sendo plausível considerar que a probabilidade de falhas simultâneas é zero, então, temos que $\mu(t) = \lambda(t)$. É importante ressaltar que a função intensidade não é condicionada ao tempo que ficou sem falhar, porém, para alguns modelos é apropriado considerar uma probabilidade condicional associada ao histórico de falhas do processo.

Dentre os processos de contagem utilizados para análise de sistemas reparáveis o processo de Poisson é amplamente citado e tem como base a distribuição de Poisson, que descreve a probabilidade de um certo número de eventos de interesse ocorrer em um determinado período de tempo e dependendo do comportamento da função intensidade de falha, o processo de Poisson pode ser classificado como processo de Poisson homogêneo e processo de Poisson não-homogêneo, descrevendo diferentes tipos de sistemas reparáveis.

Processo de Poisson homogêneo

O processo de Poisson homogêneo (PPH) é um processo de Poisson com função intensidade de falha constante, sendo o modelo mais simples usado em sistemas reparáveis pois supõe um reparo perfeito e a renovação do sistema após a falha, ou seja, o sistema volta ao estado inicial “*tão bom quanto novo*”. O PPH está diretamente relacionado com a distribuição exponencial e nesse caso é conhecido como processo de renovação (PR), Rigdon e Basu (2000) mostram (Anexo A) que um processo com função intensidade λ é um PPH se e somente se, os tempos de falha são variáveis aleatórias *iid* com distribuição exponencial de média $1/\lambda$. A função densidade de probabilidade - fdp conjunta para as variáveis aleatórias $0 < T_1 < T_2 < \dots < T_n$ é dada por

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = \lambda^n \exp(-\lambda t_n). \quad (3.4)$$

A função de verossimilhança para o PPH é dada por

$$L(\lambda) = \lambda^n \exp(-\lambda t_n), \quad 0 < t_1 < t_2 < \dots < t_n. \quad (3.5)$$

Processo de Poisson não-homogêneo

Um processo de Poisson não-homogêneo (PPNH) é um processo de Poisson com função intensidade de falha não constante e ocorre quando o tempo que o sistema já operou influencia na probabilidade de ocorrência da próxima falha, ou seja, o reparo realizado foi mínimo. Normalmente, usa-se a distribuição Weibull para modelar os tempos de falha e quando a função intensidade de falha é $\lambda(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1}$ e a função de média é $\Lambda(t) = \left(\frac{t}{\theta}\right)^\beta$, em que $\beta > 0$ e $\theta > 0$, o processo é chamado lei de potência (CROW, 1974). Esse processo é usual na literatura de sistemas reparáveis devido a flexibilidade da função intensidade e ao importante papel que os parâmetros θ e β assumem. O parâmetro θ é interpretado como o tempo durante o qual se espera exatamente que uma falha ocorra, $E[N(\theta)] = 1$, e β representa o comportamento do número esperado de falhas em relação ao tempo, ou seja, se $\beta > 1$ a função intensidade é crescente e o sistema está deteriorando, se $\beta < 1$ a função intensidade é decrescente e o sistema está melhorando e se $\beta = 1$ a função intensidade é constante e temos o PPH. Essa característica permite testar qual dos dois modelos, PPH ou PPNH, melhor se ajusta aos dados em estudo.

Para modelar sistemas reparáveis com função intensidade de falha não constante, a construção da fdp conjunta depende também do tipo de truncamento. A seguir tem-se a construção para os dois tipos de truncamento.

Truncamento por falha: Quando a coleta de dados termina depois de um número de falhas pré-determinado e o tempo total do estudo não é fixado. Rigdon e Basu (2000) descrevem (Anexo B) que a fdp conjunta para as variáveis aleatórias com os tempos de falha $T_1, T_2, \dots, T_n$ de um PPNH com função intensidade de falha $\lambda(t)$ é

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp \left(- \int_0^{t_n} \lambda(x) dx \right). \quad (3.6)$$

A função de verossimilhança é dada por

$$L(\lambda) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp \left(- \int_0^{t_n} \lambda(x) dx \right), \quad \lambda > 0. \quad (3.7)$$

Considerando um processo lei de potência com a função intensidade de falha da distribuição Weibull, $\lambda(t) = (\beta/\theta) (t/\theta)^{\beta-1}$, temos que

$$\begin{aligned}
L(\beta, \theta) &= \left(\prod_{i=1}^n (\beta/\theta) (t_i/\theta)^{\beta-1} \right) \exp \left(- \int_0^{t_n} (\beta/\theta) (x/\theta)^{\beta-1} dx \right) \\
&= \beta^n / \theta^{n\beta} \left(\prod_{i=1}^n t_i \right)^{\beta-1} \exp \left[- (t_n/\theta)^\beta \right], \quad \theta > 0, \quad \beta > 0.
\end{aligned} \tag{3.8}$$

Truncamento por tempo: Se fixamos o tempo de duração da coleta de dados, o número de falhas $N(t)$ é uma variável aleatória e isso deve ser considerado para a construção da fdp conjunta dos tempos de falha (RIGDON; BASU, 2000). Sendo assim, considere um PPNH com função intensidade $\lambda(t)$ observado até o tempo t , com tempos de falha $T_1, T_2, \dots, T_{N(t)}$ em que $N(t)$ é uma variável aleatória que descreve o número de falhas no intervalo $(0, t]$. Se condicionado a $N(t) = n$ as variáveis aleatórias $T_1 < T_2 < \dots < T_n$ são distribuídas como n estatísticas de ordem com distribuição acumulada (Anexo C) dada por

$$G(y) = \begin{cases} 0, & y \leq 0 \\ \Lambda(y)/\Lambda(t), & 0 < y \leq t \\ 1, & y > t \end{cases}, \tag{3.9}$$

em que $\Lambda(t)$ é o número esperado de falhas até o tempo t e é dado pela Equação 3.1, $\Lambda(y)$ é o número de falhas esperado até o tempo y .

A fdp conjunta de $N, T_1, T_2, \dots, T_n$ pode ser escrita como

$$f(n, t_1, \dots, t_n) = \begin{cases} f_N(n) f(t_1, \dots, t_n | n), & n \geq 1 \\ f_N(0), & n = 0. \end{cases} \tag{3.10}$$

A variável aleatória N segue uma distribuição de Poisson (JAMES, 1996) com média $(t/\theta)^\beta$ e $n = 0, 1, 2, \dots$, ou seja, $f_N(n) = \frac{[(t/\theta)^\beta]^n \exp[-(t/\theta)^\beta]}{n!}$, e a distribuição condicional de $T_1 < T_2 < \dots < T_N$ dado $N = n$ é $f(t_1, \dots, t_n | n) = n! \prod_{i=1}^n \frac{\lambda(t_i)}{\Lambda(t)} = n! \prod_{i=1}^n G'(t_i)$.

A fdp conjunta de N e $T_1 < T_2 < \dots < T_N$ é

$$f(n, t_1, \dots, t_n) = \frac{[(t/\theta)^\beta]^n \exp[-(t/\theta)^\beta]}{n!} n! \prod_{i=1}^n G'(t_i), \quad t_1 < \dots < t_n. \tag{3.11}$$

Considerando um processo lei de potência temos que $f(t_1, \dots, t_n | n) = n! \prod_{i=1}^n \frac{\beta}{t} \left(\frac{t_i}{t} \right)^{\beta-1}$, uma vez que $G(y) = \frac{(y/\theta)^\beta}{(t/\theta)^\beta} = \left(\frac{y}{t} \right)^\beta$ e $G'(y) = \frac{\beta}{t} \left(\frac{y}{t} \right)^{\beta-1}$ para $0 \leq y \leq t$. Assim, a fdp conjunta de N e $T_1 < \dots < T_N$ para esse caso particular é:

$$\begin{aligned}
f(n, t_1, \dots, t_n) &= \frac{[(t/\theta)^\beta]^n \exp[-(t/\theta)^\beta]}{n!} n! \prod_{i=1}^n \frac{\beta}{t} \left(\frac{t_i}{t} \right)^{\beta-1} = \frac{t^{n\beta}}{\theta^{n\beta}} \frac{\beta^n}{t^n} \left[\prod_{i=1}^n \left(\frac{t_i}{t} \right) \right]^{\beta-1} \exp \left[- (t/\theta)^\beta \right] \\
&= \frac{\beta^n}{\theta^{n\beta}} \left(\prod_{i=1}^n t_i \right)^{\beta-1} \exp \left[- (t/\theta)^\beta \right], \quad n \geq 0, \quad 0 < t_1 < \dots < t_n < t
\end{aligned} \tag{3.12}$$

Observe que para um PPNH, a função de verossimilhança considerando o estudo com truncamento por falha dado por (3.8) e a função de verossimilhança considerando o truncamento por tempo dado pela Equação 3.12 são iguais e isso acontece para o caso especial do processo lei de potência.

Decidir qual modelo deve ser usado nem sempre é uma tarefa fácil. Muitas vezes não é possível classificar o tipo de reparo apenas com a descrição do mesmo e nesses casos os métodos gráficos são importantes ferramentas para análise de sistemas reparáveis, seja na análise descritiva ou no indicativo

da modelagem que deve ser usada posteriormente. Com essas ferramentas é possível verificar o comportamento do sistema estudado ao longo das várias falhas e entender se o sistema está melhorando, deteriorando ou ainda, se é razoável pensar em um PPH ou um PPNH.

Considerando 12 tempos de falhas de cada um dos três sistemas fictícios (RIGDON; BASU, 2000) relataram que com o simples gráfico dos tempos de falha t em função do número de falhas acumuladas $N(t)$ é possível verificar o comportamento do sistema ao longo do tempo (Figura 3.1).

Figura 3.1. Gráfico do número de falhas acumuladas $N(t)$ ao longo do tempo t . (a) Sistema melhorando. (b) Sistema constante. (c) Sistema deteriorando.

Ao longo do tempo os sistemas podem assumir três diferentes comportamentos: estar melhorando, ficar constante ou se deteriorando (Figura 3.1). A relação linear existente na Figura 3.1 (b) indica estabilidade do sistema, o tempo entre as falhas são independentes e não estão aumentando ou diminuindo, indicando o uso de um PR ou PPH. Observa-se, na Figura 3.1 (a) e (c) a existência de curvatura, possibilitando considerar que os tempos independentes não é mais plausível. Na Figura 3.1 (a) o tempo entre as falhas está aumentando ao longo do tempo e o sistema está melhorando, na Figura 3.1 (c) os tempos entre as falhas estão diminuindo indicando uma piora na condição do sistema que está deteriorando. Quando o gráfico apresenta funções com concavidade para cima ou para baixo o PPNH e o processo lei de potência são indicados na modelagem.

Gráfico Duane

Uma técnica gráfica que pode ser usada para verificar a adequabilidade do processo lei de potência é o gráfico de Duane descrito por Duane (1964), que é um gráfico de dispersão de $N(t_i)/t_i$ versus t_i para $i = 1, 2, \dots, n$, em que $N(t_i)/t_i$ é a taxa de falha acumulada. Duane (1964) sugeriu utilizar o logaritmo desses valores e os pontos plotados no gráfico seriam $(\log t_i, \log (N(t_i)/t_i))$, assim, é adequado pensar que a curva seja aproximadamente linear se o sistema reparável puder ser modelado por um processo lei de potência.

Considerando os mesmos valores utilizados na Figura 3.1, pode-se construir e verificar o comportamento dos sistemas reparáveis, através do gráfico de Duane (Figura 3.2). Observa-se que a inclinação decrescente dos pontos indica uma melhora no sistema ao longo do tempo (Figura 3.2 (a)), e que o sistema está se deteriorando quando a curva é crescente ao longo do tempo (Figura 3.2 (c)). Com relação á Figura 3.2 (b) não é possível afirmar que exista uma relação linear entre o logaritmo do tempo e o número de falhas, entretanto, se desconsiderarmos os dois pontos iniciais (possíveis outliers) o sistema permanece estável ao longo do tempo observado e, portanto, pode ser um indicativo de função intensidade constante o que nos leva a pensar em um PPH.

Figura 3.2. Gráfico de Duane. (a) Sistema melhorando. (b) Sistema constante. (c) Sistema deteriorando.

Gráfico Tempo Total em Teste - TTT-plot

O gráfico do tempo total em teste, conhecido por TTT-plot, foi desenvolvido por Barlow e Campo (1975) para selecionar a distribuição adequada para modelar os tempos de falha de um sistema não reparável ou dados de sobrevivência; porém, esse gráfico pode ser utilizado para sistemas reparáveis. Klefsjo e Kumar (1992) verificaram que, se considerarmos os dados dos tempos de falha $T_1 < T_2 < \dots < T_n$ de um sistema reparável, possivelmente modelado por um processo lei de potência, e realizado uma transformação nos dados, então

$$w_j = -\log\left(\frac{T_{n-j}}{T_n}\right), \quad j = 1, 2, \dots, n-1 \quad (3.13)$$

são estatísticas de ordem que seguem uma distribuição exponencial de média $1/\beta$. Assim, o TTT-plot utiliza $w_1 < w_2 < \dots < w_{n-1}$ no lugar de $t_{(1)} < t_{(2)} < \dots < t_{(n)}$ para a construção do gráfico que verifica o ajuste de um processo lei de potência. Se a curva da função estimada se aproximar da reta diagonal do gráfico, o processo lei de potência é o indicado.

Considerando os dados do primeiro sistema apresentado por Rigdon e Basu (2000) e usados para exemplificar as técnicas gráficas anteriores (Figura 3.1 e Figura 3.2), nota-se a aproximação da curva na reta diagonal (Figura 3.3), indicando que o processo lei de potência pode ser adequado para analisar esse sistema.

É importante ressaltar que as técnicas gráficas descritas servem apenas como indicativos do comportamento dos sistemas reparáveis ao longo do tempo, sendo que o ajuste dos modelos e os testes que verificam a qualidade desses ajustes são primordiais para o estudo.

3.3 Resultados e Discussão

Os dados analisados são referentes aos tempos de falhas de uma peça denominada Lâmina Chopper presente em quatro colhedoras de cana-de-açúcar pertencentes a um grupo de usinas com quatro unidades produtoras de açúcar e etanol durante a safra de 2014/2015. Essa lâmina é responsável por cortar a cana em pequenos pedaços para serem transportados pelas esteiras até os caminhões. Os tempos de falhas das lâminas são contados em dias e, foi selecionada uma colhedora de cada unidade para analisar seu comportamento separadamente.

No gráfico de Duane (Figura 3.4) observa-se uma inclinação decrescente dos pontos em algumas máquinas, que é um indicativo de função intensidade de falha não constante, levando a um PPNH. A

Figura 3.3. Gráfico de TTT-plot para um sistema reparável em que o processo lei de potência é indicado.

partir da construção do gráfico TTT-plot, observa-se na Máquina 1 (Figura 3.5) uma aproximação da curva com a reta diagonal maior do que nas outras colhedoras, dando indicativos de que o PPH modela melhor esses dados. Já para a Máquina 3 (Figura 3.5) essa situação é inversa e a distância da curva à reta diagonal indica que o PPNH é melhor. Para as Máquinas 2 e 4 (Figura 3.5), não é possível afirmar ao certo qual processo de Poisson é melhor, entretanto, nas duas máquinas temos que a curva obtida foi côncava ou convexa, ou seja, que podem ser modelados por um PPNH ou seu caso particular, o processo lei de potência.

Figura 3.4. Gráfico de Duane das quatro Máquinas em estudo.

Figura 3.5. TTT-plot para as quatro colhedoras de cana-de-açúcar em estudo.

Com a análise gráfica do comportamento dos sistemas não é possível determinar ao certo qual o melhor modelo, apenas tem indicativos, e a utilização de métodos computacionais se torna necessária para o processo de inferência, ou seja, para determinar os estimadores de α e β , o erro padrão, seus respectivos intervalos de confiança e os critérios de seleção de modelos, o critério de Akaike - AIC (AKAIKE, 1974) e critério de informação bayesiano - BIC (SCHWARZ et al, 1978).

O teste de razão de verossimilhança para comparar os modelos PPH e PPNH (Tabela 3.2) foi realizado para cada máquina em estudo e apresenta indicações favoráveis ao PPH para as Máquinas 1 e 4 e favorável ao PPNH para as Máquinas 2 e 3, sendo os valores das estimativas dos parâmetros desses modelos apresentada na Tabela 3.1.

A partir dos valores estimados dos parâmetros do PPNH para as Máquinas 2 e 3 foi possível construir o gráfico contendo o número de falha esperado ($E(N(t))$) (Figura 3.6). Observa-se que a Máquina 2 tem um número de falhas esperado maior que a Máquina 3. Por exemplo, para $t = 10$ dias tem-se que o número de falhas esperado para a Máquina 2 é 5, já para a Máquina 3 é de 2 falhas.

3.4 Conclusão

A análise gráfica feita para os dados de falha da lâmina Chopper e os testes para comparação dos modelos foi de total importância para a modelagem desses sistemas separados e, nesse caso, o tipo de truncamento realizado na coleta de dados não influenciou na escolha do modelo, já que no caso do processo lei de potência a função de verossimilhança é a mesma para os dois tipos de truncamento. As curvas de confiabilidade para as Máquinas 2 e 3 indicaram uma função intensidade de falha crescente, ou seja, ocorre deterioração do sistema ao longo da safra; já as Máquinas 1 e 4 podem ser modeladas por um PPH, o que indica uma taxa de falha constante ao longo da safra. Com essas informações, será possível estimar o tempo de operação e o tempo ideal para reparos, otimizando ao máximo esse serviço.

Tabela 3.1. EMVs, SEs, intervalos de confiança e estatísticas para os modelos ajustados para os dados de colhedoras de cana-de-açúcar.

Weibull						
Máquina	Parâmetro	Estimativa	SE	IC (95%)	AIC	BIC
1	β	0,950	0,224	[0,511; 1,388]	41,898	37,898
	θ	0,906	0,688	[-0,442; 2,253]		
2	β	2,681	0,716	[1,276; 4,085]	27,382	23,382
	θ	6,352	1,787	[2,849; 9,854]		
3	β	3,568	0,921	[1,762; 5,374]	15,362	11,362
	θ	6,554	1,369	[3,871; 9,238]		
4	β	0,662	0,171	[0,327; 0,998]	42,566	38,566
	θ	0,369	0,415	[-0,445; 1,182]		
Exponencial						
Máquina	Parâmetro	Estimativa	SE	IC(95%)	AIC	BIC
1	β	(1)	-	-	39,946	37,946
	θ	1,056	0,249	[0,567; 1,543]		
2	β	(1)	-	-	35,436	33,436
	θ	1,214	0,325	[0,578; 1,850]		
3	β	(1)	-	-	29,930	27,930
	θ	0,933	0,241	[0,461; 1,410]		
4	β	(1)	-	-	42,490	38,490
	θ	1,467	0,379	[0,724; 2,209]		

Tabela 3.2. Teste de razão de verossimilhanças para dados de cada uma das colhedoras de cana-de-açúcar.

Máquina	Modelos	Hipóteses	Estatística w	p -valor
1	PPNH vs PPH	$H_0 : \beta = 1$ vs $H_1 : H_0$ é falso	0,049	0,976
2	PPNH vs PPH	$H_0 : \beta = 1$ vs $H_1 : H_0$ é falso	10,055	0,007
3	PPNH vs PPH	$H_0 : \beta = 1$ vs $H_1 : H_0$ é falso	16,569	0,001
4	PPNH vs PPH	$H_0 : \beta = 1$ vs $H_1 : H_0$ é falso	2,934	0,231

Figura 3.6. Gráficos do número esperado de falhas para as Máquinas modeladas por um processo lei de potência.

Vale ressaltar que o estudo de um único sistema separado pode ser apenas o início das análises, uma vez que cada unidade da usina possui várias colhedoras de cana-de-açúcar.

REFERÊNCIAS

- AKAIKE, H., A new look at the statistical model identification. IEEE transactions on automatic control, v. 19, p. 716-723, 1974.
- ASCHER, H.; FEINGOLD, H. Repairable systems reliability. Lecture notes in statistics, v. 7, 1984.
- BARLOW, R.; HUNTER, L. Optimum preventive maintenance policies. Operations Research, v. 8, n. 1, p. 90-100, 1960.
- BARLOW, R. E.; CAMPO, R. A. Total Time on Test Processes and Applications to Failure Data Analysis. [S.l.], 1975.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. [S.l.]: Edgard Blucher, 2006.
- CROW, L. H. Reliability analysis for complex, repairable systems. [S.l.], 1974.
- DUANE, J. Learning curve approach to reliability monitoring. IEEE transactions on Aerospace, IEEE, v. 2, n. 2, p. 563-566, 1964.
- JAMES, B. R. B. R. Probabilidade: um curso em nível intermediário. [S.l.: s.n.], 1996.
- KIJIMA, M.; SUMITA, U. A useful generalization of renewal theory: counting processes governed by non-negative markovian increments. Journal of Applied Probability, Cambridge University Press, Cambridge, UK, v. 23, n. 1, p. 71-88, 1986.
- KLEFSJO, B.; KUMAR, U. Goodness-of-t tests for the power-law process based on the ttt-plot. IEEE transactions on reliability, IEEE, v. 41, n. 4, p. 593-598, 1992.
- PHAM, H.; WANG, H. Imperfect maintenance. European journal of operational research, North-Holland, v. 94, n. 3, p. 425-438, 1996.
- RIGDON, S. E.; BASU, A. P. Statistical Methods for the Reliability of Repairable Systems. [S.l.]: Wiley Series in Probability and Statistics, 2000.
- SCHWARZ, G. et al., Estimating the dimension of a model. The annals of statistics, v. 6, p. 461-464, 1978.

4 MODELO DE REGRESSÃO PARA DOIS OU MAIS SISTEMAS REPARÁVEIS

Resumo

Na análise de sistemas reparáveis características importantes para a modelagem são levadas em consideração, como truncamento, tipo de reparo e igualdade dos sistemas. Os principais processos de contagem utilizados para esse tipo de modelagem são: processo de Poisson homogêneo, no caso dos tempos de falha constante e processo lei de potência para o caso em que o sistema está melhorando ou deteriorando ao longo do tempo. A presença de covariável no problema em estudo exigiu a utilização de um modelo de regressão Weibull, ou seja, foi acrescentado um modelo de regressão ao parâmetro de forma da distribuição Weibull usada no processo lei de potência. O modelo de regressão Weibull aplicado ao processo lei de potência foi o modelo que melhor se ajustou aos dados dos tempos falhas do sistema de corte das colhedoras de cana-de-açúcar analisado, e a significância dos parâmetros da regressão indicou que a covariável associada aos tempos de falha influenciou na durabilidade da peça em estudo.

Palavras-chave: sistema reparável; processo lei de potência; modelo de regressão Weibull.

Abstract

In the analysis of more than one repairable system, important characteristics for modeling are taken into account, as truncation, repair type and systems equality. The mainly counting process used for this type of modeling are homogeneous Poisson process, in cases of constant failure time, and power law process for the case that the system is improving or deteriorating over time. The covariable presence in the analyzed problem demanded a Weibull regression model and it was added a regression model to the form parameter of Weibull distribution used in the power law process. The Weibull regression model applied to the power law process was the fittest model to the datas of failure time of sugarcane harvest cut systems and the parameters significance indicated that the covariable associated to the failure time influenced in the durability of the piece.

Keywords: repairable system; power law process; Weibull regression model.

4.1 Introdução

Sistema reparável é definido como após a ocorrência de uma falha e realização de um reparo o sistema volta a exercer sua função de origem (ASCHER; FEINGOLD, 1984). Na prática, quando analisamos máquinas ou equipamentos eletrônicos temos interesse em entender o comportamento dos diversos sistemas conjuntamente. Para a modelagem de sistemas reparáveis o processo de contagem utilizado é o processo de Poisson (RIGDON; BASU, 2000), sendo os principais: o processo de Poisson homogêneo e o processo lei de potência. O primeiro considera que os tempos entre as falhas são os mesmos e a função intensidade de falhas é constante; já o segundo, que os tempos entre as falhas variam, podendo aumentar ou diminuir ao longo do tempo. Este processo modela sistemas que estão melhorando ou deteriorando ao longo do tempo e a distribuição associada é a Weibull (CROW, 1974). Rigdon e Basu (2000) e Barlow e Hunter (1960) mostraram que a construção da função de verossimilhança para estimação dos parâmetros do modelo depende, além do processo de contagem adequado, do pressuposto de igualdade dos sistemas e para cada caso uma técnica diferente deve ser assumida para a modelagem e, portanto, testes para verificar a igualdade dos sistemas devem ser realizados.

Além dos tempos de falha, outras medidas e características podem estar associadas aos sistemas, como a localização geográfica de cada um, o fornecedor das peças ou o motorista em caso de

máquinas. Quando mais de uma variável influencia na durabilidade dos sistemas, técnicas que incluam essas covariáveis devem ser usadas. Os modelos de regressão mais utilizados em análise de sobrevivência são: o semi-paramétrico de Cox (COX, 1972) que assume o pressuposto de taxa de falhas proporcionais e o modelo de regressão Weibull (COLOSIMO; GIOLO, 2006) que acrescenta um modelo de regressão nos parâmetros da distribuição usada, neste caso a Weibul.

Um banco de dados contendo os tempos de falhas e covariáveis que influenciam na durabilidade de máquinas analisado por meio das técnicas de sistemas reparáveis. Os equipamentos estudados são colhedoras de cana-de-açúcar pertencentes a um grupo de Usinas produtoras de Açúcar e Etanol e o objetivo é encontrar um modelo que se adeque ao dados, verificar se as covariáveis são significativas e descrevem o comportamento dessas colhedoras ao longo da safra.

4.2 Material e Método

Processo de Poisson homogêneo

Considere k sistemas reparáveis que assumem o pressuposto de serem modelados por um PPH. Se os sistemas são idênticos, cada sistema é um PPH com função intensidade $1/\theta$ e θ é o mesmo para todos. Para a estimação pontual de θ considere que t_{ij} é a j -ésima falha do i -ésimo sistema e que n_i falhas são observadas, seja T_i uma variável aleatória com os tempos de falha do i -ésimo sistema, assuma também que a sistemas sejam truncados por falha ($T_i = t_{i,n_i}$) e $k - a$ sistemas são truncados por tempo ($T_i > t_{i,n_i}$). Sabendo que é um PPH e que os tempos entre as falhas são *iid* com distribuição exponencial, temos que a função de verossimilhança para o caso dos a sistemas truncados por falha é dada por:

$$\begin{aligned} L_1(\theta) &= \prod_{i=1}^a \prod_{j=1}^{n_i} \frac{1}{\theta} \exp[-(t_{i,j} - t_{i,j-1})/\theta] \\ &= \theta^{-\sum_{i=1}^a n_i} \exp\left[-\sum_{i=1}^a t_{i,n_i}/\theta\right] \\ &= \theta^{-\sum_{i=1}^a n_i} \exp\left[-\sum_{i=1}^a T_i/\theta\right]. \end{aligned}$$

Para os $k - a$ sistemas truncados por tempo, o número de falhas N_i no intervalo $[0, T_i]$ segue uma distribuição de Poisson com média T_i/θ e a função de verossimilhança é dada por

$$\begin{aligned} L_2(\theta) &= \prod_{i=a+1}^k P(N_i = n_i) \\ &= \prod_{i=a+1}^k \frac{(T_i/\theta)^{n_i} \exp(-T_i/\theta)}{n_i!} \\ &= \left(\prod_{i=a+1}^k \frac{T_i^{n_i}}{n_i!} \right) \theta^{-\sum_{i=1}^k n_i} \exp\left(-\sum_{i=a+1}^k T_i/\theta\right). \end{aligned}$$

A função de verossimilhança conjunta dos k sistemas é

$$\begin{aligned} L(\theta) &= L_1(\theta)L_2(\theta) \\ &\propto \left\{ \theta^{-\sum_{i=1}^a n_i} \exp\left(-\sum_{i=1}^a T_i/\theta\right) \right\} \left\{ \theta^{-\sum_{i=1}^k n_i} \exp\left(-\sum_{i=a+1}^k T_i/\theta\right) \right\} \\ &= \theta^{-\sum_{i=1}^k n_i} \exp\left(-\sum_{i=1}^k T_i/\theta\right) \end{aligned} \quad (4.1)$$

Resolvendo a equação $\frac{dl}{d\theta} = 0$, em que $\frac{dl}{d\theta}$ é a derivada primeira do logaritmo da Equação 4.1, encontramos o estimador de máxima verossimilhança de θ , ou seja,

$$\hat{\theta} = \frac{\sum_{i=1}^k T_i}{\sum_{i=1}^k n_i} \quad (4.2)$$

O intervalo de confiança para o parâmetro θ é obtido considerando que os j sistemas foram truncados por falha, assim, $2T_{j,n_j}/\theta$ segue uma distribuição qui-quadrado com $2n_j$ graus de liberdade. Seja a soma das k variáveis aleatórias com distribuição qui-quadrado, assim, $\frac{2}{\theta} \sum_{i=1}^k T_i$ também é uma variável aleatória de distribuição qui-quadrado com $2 \sum_{i=1}^k n_i$ graus de liberdade. Então, o intervalo de confiança de $(1 - \alpha)\%$ de confiança é dado por

$$\frac{2 \sum_{i=1}^k T_{i,n_i}}{\chi_{\alpha/2}^2 \left(2 \sum_{i=1}^k n_i \right)} < \theta < \frac{2 \sum_{i=1}^k T_{i,n_i}}{\chi_{1-\alpha/2}^2 \left(2 \sum_{i=1}^k n_i \right)} \quad (4.3)$$

Considere os sistemas são diferentes, o tempo entre as falhas e as estimativas de $\theta_1, \theta_2, \dots, \theta_k$ também. Primeiro, deve-se realizar um teste de igualdade dos parâmetros e no caso de $k = 2$ sistemas truncados por falha testamos a hipótese $H_0 : \theta_1 = \theta_2$ versus $H_0 : \theta_1 \neq \theta_2$ sendo que o EMV de θ_i é dado por

$$\theta_i = \frac{T_{i,n_i}}{n_i}, \quad i = 1, 2$$

A estatística do teste $\frac{2T_{i,n_i}}{\theta_i}$ tem distribuição qui-quadrada com $2n_i$ graus de liberdade e, sob a hipótese nula verdadeira, $F = \frac{\hat{\theta}_1}{\hat{\theta}_2} \sim F(2n_1, 2n_2)$ e rejeitamos H_0 se $F < F_{1-\alpha/2}(2n_1, 2n_2)$ ou $F > F_{\alpha/2}(2n_1, 2n_2)$. Considere k sistemas truncados por falha, por tempo ou pelos dois, a hipótese testada é $H_0 : \theta_1 = \theta_2 = \dots = \theta_k$ e utiliza-se o teste de razão de verossimilhança. A estatística do teste é dada por $-2 \log \left(\frac{L(\theta_0)}{L(\hat{\theta})} \right)$ em que $L(\theta_0)$ é a verossimilhança aplicada no valor do parâmetro sob a hipótese nula e $L(\hat{\theta})$ é a verossimilhança aplicada na estimativa do parâmetro. Essa estatística tem aproximadamente distribuição qui-quadrado com $k - 1$ graus de liberdade (PAWITAN, 2001) e, valores altos dessa estatística levam à rejeição da hipótese nula de igualdade dos θ 's.

Processo lei de potência

Suponha k sistemas reparáveis idênticos com suas falhas modeladas por um processo lei de potência. Seja t_{ij} o tempo de falha da j -ésima falha para o i -ésimo sistema observado até o tempo T_i , em que $T_i = t_{i,n_i}$ se o sistema for truncado por falha e $T_i > t_{i,n_i}$ se o sistema for truncado por tempo. Seja n_i o número de falhas observadas para o i -ésimo sistema e $N = n_1 + n_2 + \dots + n_k$ então a função de verossimilhança para os k sistemas é dada por:

$$\begin{aligned} L(\beta, \theta) &= \prod_{i=1}^k \left\{ \left[\prod_{j=1}^{n_i} \lambda(t_{i,j}) \right] \exp \left[- \int_0^{T_i} \lambda(x) dx \right] \right\} \\ &= \prod_{i=1}^k \left\{ \left[\prod_{j=1}^{n_i} \frac{\beta}{\theta} \left(\frac{t_{i,j}}{\theta} \right)^{\beta-1} \right] \exp \left[- \left(T_i / \theta \right)^\beta \right] \right\}. \end{aligned} \quad (4.4)$$

Condicionando a função de verossimilhança ao número de falhas (truncamento por tempo) ou ao tempo da n -ésima falha (truncamento por falha), temos que os EMVs são dados por

$$\hat{\beta} = \frac{M}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(T_i/t_{i,j})} \quad \text{e} \quad \hat{\theta} = \left(\frac{\sum_{i=1}^k T_i^{\hat{\beta}}}{M} \right)^{1/\hat{\beta}},$$

em que, $M = \sum_{i=1}^k m_i$ e

$$m_i = \begin{cases} n_i - 1, & \text{se temos um truncamento por falha} \\ n_i, & \text{se temos um truncamento por tempo} \end{cases}$$

O intervalo de confiança para β é dado por:

$$\frac{\chi_{1-\alpha/2}^2(2M)\hat{\beta}}{2M} < \beta < \frac{\chi_{\alpha/2}^2(2M)\hat{\beta}}{2M}$$

Para construção da função de verossimilhança considera-se que os sistemas são idênticos e, portanto, seus parâmetros são iguais, mas é necessário realizar um teste de hipótese para verificar essa suposição. O teste de razão de verossimilhança apresentado por Bartlett (1937) é usado para testar a hipótese $H_0 : \beta_1 = \beta_2 = \dots = \beta_k$ versus a hipótese de que pelo menos dois β 's diferem. Altos valores para a estatística do teste $-\frac{2LR}{a} \sim \chi^2(k-1)$, em que $LR = \log\left(\frac{L(\beta_0)}{L(\hat{\beta})}\right)$ e $a = 1 + \frac{1}{6(k-1)}\left(\sum_{i=1}^k \frac{1}{m_i} - \frac{1}{M}\right)$, levam a rejeição da hipótese H_0 . Rigdon e Basu (2000) sugerem que se a hipótese nula for rejeitada e os sistemas não são idênticos, é necessário analisá-los separadamente usando as técnicas apresentadas anteriormente no Capítulo 3. Caso seja possível assumir que todos os sistemas têm o mesmo parâmetro β e diferentes parâmetros θ 's, pode-se usar a seguinte função de verossimilhança

$$L(\beta, \theta) = \prod_{i=1}^k \left\{ \left[\prod_{j=1}^{n_i} \frac{\beta}{\theta_i} \left(\frac{t_{i,j}}{\theta_i} \right)^{\beta-1} \right] \exp \left[- \left(T_{i/\theta_i} \right)^{\beta} \right] \right\}, \quad (4.5)$$

e os estimadores de máxima verossimilhança (EMV) para os k parâmetros de escala e para o β são

$$\hat{\theta}_i = \frac{T_i}{n_i^{1/\hat{\beta}}} \quad \text{e} \quad \hat{\beta} = \frac{M}{\sum_{i=1}^k \sum_{j=1}^{n_i} \log(T_i/t_{ij})}.$$

Observe que o EMV de β é igual ao caso em que se considera k sistemas idênticos.

Modelo de Regressão

Em confiabilidade o modelo de regressão frequentemente utilizado é o modelo de taxas de falhas proporcionais, ou modelo de regressão de Cox (COX, 1972). Esse modelo é considerado semi-paramétrico, uma vez que a função taxa de falha $\lambda(t) = \lambda_0(t)g(\mathbf{x}'\beta)$ apresenta o componente paramétrico $g(\mathbf{x}'\beta)$ e o componente não-paramétrico $\lambda_0(t)$. Um ponto importante para esse modelo é que a violação da suposição de proporcionalidade das taxas de falha pode gerar vícios na estimação dos coeficientes do modelo (STRUTHERS; KALBFLEISCH, 1986) e, um método alternativo é utilizar o modelo de regressão paramétrico, que consiste em acrescentar um modelo de regressão nos parâmetros da distribuição usada para a modelagem de sistemas reparáveis.

Considere o tempo de falha da variável aleatória T e um vetor de covariáveis denotado por $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$. Para a análise de sistemas reparáveis o parâmetro de forma θ , com as covariáveis por meio da função de ligação log-linear $\theta_i = \exp(\mathbf{x}_i^T \beta_1)$, em que $\beta_1 = (\beta_{10}, \dots, \beta_{1p})^T$ denota o vetor de coeficientes da regressão e $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$. O modelo de regressão Weibull pode ser aplicado no processo de Poisson homogêneo, pois a distribuição exponencial é um caso particular dessa distribuição e, considerando a Equação 4.1 temos que a nova função de verossimilhança para o modelo de regressão é dado por

$$L(\beta_1) = \exp(\mathbf{x}_i^T \beta_1)^{-\sum_{i=1}^k n_i} \exp \left(- \sum_{i=1}^k T_i / \exp(\mathbf{x}_i^T \beta_1) \right). \quad (4.6)$$

Para o processo lei de potência com k sistemas idênticos e considerando a Equação 4.7, temos que a função de verossimilhança para o modelo de regressão Weibull para o processo lei de potência é dado por

$$L(\beta, \beta_1) = \prod_{i=1}^k \left\{ \prod_{j=1}^{n_i} \frac{\beta}{\exp(\mathbf{x}_i^T \beta_1)} \left(\frac{t_{i,j}}{\exp(\mathbf{x}_i^T \beta_1)} \right)^{\beta-1} \right\} \exp \left[- \left(T_i / \exp(\mathbf{x}_i^T \beta_1) \right)^\beta \right]. \quad (4.7)$$

Maximizando a função de verossimilhança é possível encontrar os estimadores para β e β_1 .

4.3 Resultados e Discussão

Os dados analisados referem-se a quatro usinas pertencentes a um mesmo grupo, que na safra de 2014/15 gastaram R\$8.589.230,87 reais apenas com reparos em 20 colhedoras de cana-de-açúcar da empresa, sendo que os gastos com reparos com a Lâmina Chopper corresponderam a R\$469.784,19 reais, ou seja, 5,5% dos gastos totais com reparos, e por isso ela se tornou objeto de estudo. O evento de interesse é a quebra da lâmina Chopper e o tempo foi contado em dias. As covariáveis são tempo até a falha t_i , Estado onde a colhedora está localizada x_{1i} , São Paulo ou Paraná e agente causador da falha x_{2i} , desgaste natural ou outros. Os dados foram truncados por tempo, após 250 dias de safra, sendo que a Máquina que menos falhou apresentou 13 falhas e a que mais falhou, 22 falhas. A distribuição das falhas de cada Máquina (Figura 4.1) ao longo do tempo dá indicativos que o modelo PPH não é o melhor, uma vez que os tempos entre as falhas não é constante.

Figura 4.1. Tempos de falha dos 20 sistemas truncados por tempo.

O teste de razão de verossimilhanças para verificar a igualdade dos parâmetros β 's, cuja hipótese é $H_0 : \beta_1 = \dots = \beta_{20}$, apresenta o valor pra estatística do teste de $-\frac{2LR}{a} = 2,088$ e p-valor de 0,998 e, considerando o nível de significância $\alpha = 0,05$, não temos evidência estatística de que os β 's são diferentes. Assim, a função de verossimilhança para o modelo de regressão Weibull é a apresentada na Equação 4.5 e a função de verossimilhança para o modelo de regressão exponencial é apresentada na Equação 4.6. Através de métodos computacionais foram determinados as EMVs de β e β_1 , assim como o erro padrão e p-valor para o teste de significância das covariáveis (Tabela 4.1), para o modelo de regressão Weibull e para o modelo de regressão exponencial, juntamente com as estatísticas AIC e BIC.

Os parâmetros β_{10} e β_{12} não são significativos (Tabela 4.1) e a variável Agente causador da falha não influencia na durabilidade das Lâminas Chopper. Os critérios de rejeição AIC e BIC são favoráveis ao modelo de Regressão Weibull e como a variável x_{i2} não foi significativa, novas estimativas para os parâmetros foram calculadas desconsiderando essa variável (Tabela 4.2).

Tabela 4.1. EMVs, SEs, p -valor e estatísticas de qualidade de ajuste para os modelos de regressão ajustados para os dados de colhedoras de cana-de-açúcar considerando todas as covariáveis.

Weibull					
Parâmetros	Estimativa	SE	p-valor	AIC	BIC
β	0,798	0,093	-	179,156	177,315
β_{10}	-0,631	0,607	0,299	-	-
β_{11}	0,666	0,326	0,041	-	-
β_{12}	-0,160	0,644	0,803	-	-
Exponencial					
Parâmetro	Estimativa	SE	p-valor	AIC	BIC
β	(1)	(-)	-	181,242	180,015
β_{10}	-0,223	0,447	0,617	-	-
β_{11}	0,746	0,259	0,004	-	-
β_{12}	0,183	0,488	0,706	-	-

Tabela 4.2. EMVs, SEs, p -valor e estatísticas para os modelos de regressão ajustados para os dados de colhedoras de cana-de-açúcar considerando apenas a covariável Estado.

Weibull					
Parâmetros	Estimativa	SE	p-valor	AIC	BIC
β	0,904	0,037	-	704,557	707,544
β_{10}	-0,953	0,176	<0,0001	-	-
β_{11}	0,803	0,138	<0,0001	-	-
Exponencial					
Parâmetro	Estimativa	SE	p-valor	AIC	BIC
β	(1)	(-)	-	708,870	710,862
β_{10}	-0,649	0,106	<0,0001	-	-
β_{11}	0,862	0,123	<0,0001	-	-

O modelo de regressão Weibull foi que melhor ajustou aos dados (p -valor foi 0,042) sendo que os parâmetros β_{10} e β_{11} foram significativos e os Estados do Paraná e São Paulo fazem diferença para os tempos de falha. Na data da coleta de dados, em São Paulo não se queimava a cana-de-açúcar para a colheita, mas no Paraná sim, e o fato de ter palha na cana durante o processo de corte influencia no tempo do sistema. O parâmetro $\beta < 1$ significa que o modelo está melhorando ao longo do tempo e isso se deve ao fato das Máquinas trabalharem fortemente no início da safra e no final, dependendo da localização do canavial onde elas se encontram isso é reduzido e o impacto nas falhas é maior no início da safra.

4.4 Conclusão

O modelo de regressão Weibull foi o que melhor se adequou aos dados das colhedoras de cana-de-açúcar. A significância dos parâmetros de regressão mostrou que o Estado onde a Usina está localizada influencia na durabilidades dessas peças e usar apenas o processo lei de potência para modelar esses dados não é capaz de incorporar a informação desta variável no modelo. A estimativa do parâmetro β mostrou uma melhora na durabilidade dos sistemas ao longo da safra indicando que o grau de exigência das colhedoras de cana-de-açúcar diminui ao longo da safra.

REFERÊNCIAS

- AKAIKE, H., A new look at the statistical model identification. IEEE transactions on automatic control, v. 19, p. 716-723, 1974.
- ASCHER, H.; FEINGOLD, H. Repairable systems reliability. Lecture notes in statistics, v. 7, 1984.
- BARLOW, R.; HUNTER, L. Optimum preventive maintenance policies. Operations Research, v. 8, n. 1, p. 90-100, 1960.
- BARTLETT, M. S. Properties of sufficiency and statistical tests. Proceedings of Royal Society of London A, v. 160, p. 268-82, 1937.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. [S.l.]: Edgard Blucher, 2006.
- COX, D., Regression models and life-tables. Journal of the Royal Statistical Society - Series B, v. 34, p. 187-220, 1972.
- CROW, L. H. Reliability analysis for complex, repairable systems. [S.l.], 1974.
- PAWITAN, YUDI. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, 2001.
- RIGDON, S. E.; BASU, A. P. Statistical Methods for the Reliability of Repairable Systems. [S.l.]: Wiley Series in Probability and Statistics, 2000.
- SCHWARZ, G. et al., Estimating the dimension of a model. The annals of statistics, v. 6, p. 461-464, 1978.
- STRUTHERS, C. A.; KALBFLEISCH, J. D. Misspecified Proportional Hazards Models. Biometrika, p. 363-369, 1986.

APÊNDICES

Apêndice I

Função verossimilhança para os modelos GOLLFW e FW usadas no Capítulo 2.

Distribuição GOLLFW

```
verGOLLFW <- function(t,censur,par){
  alpha <- par[1]
  theta <- par[2]
  a <- par[3]
  b <- par[4]
  g <- exp((a*t - b/t)-exp(a*t - b/t))*(a+b/(t^2))
  G <- 1-exp(-exp(a*t - b/t))
  f <- (alpha*theta*g*(G^(alpha*theta-1))*(1-G^theta)^(alpha-1)) / (((G^(alpha*theta)) + (1-G^theta)^(alpha))^2)
  FF <- (G^(alpha*theta)) / (((G^(alpha*theta)) + (1-G^theta)^(alpha)))
  lv <- censur*(log(f)) + (1-censur)*(log(1-FF))
  logv <- -sum(lv)
  return(logv) }
```

Distribuição FW

```
verFW <- function(t,censur,par){
  alpha <- 1
  theta <- 1
  a <- par[1]
  b <- par[2]
  g <- exp((a*t - b/t)-exp(a*t - b/t))*(a+b/(t^2))
  G <- 1-exp(-exp(a*t - b/t))
  f <- (alpha*theta*g*(G^(alpha*theta-1))*(1-G^theta)^(alpha-1)) / (((G^(alpha*theta)) + (1-G^theta)^(alpha))^2)
  F <- (G^(alpha*theta)) / (((G^(alpha*theta)) + (1-G^theta)^(alpha)))
  lv <- censur*(log(f)) + (1-censur)*(log(1-F))
  logv <- -sum(lv)
  return(logv) }
```

Apêndice II

Análise Gráfica apresentada nos Capítulos 3 e 4.

Gráfico do n° de falhas acumuladas ao longo do tempo

```
par(mfrow=c(1,2))
plot(t1, n1, xlab = 't', ylab = 'N(t)', xlim = c(0,30), ylim = c(0,20), lwd=1, pch=16)
plot(t2, n2, xlab = 't', ylab = 'N(t)', xlim = c(0,70), ylim = c(0,30), lwd=1, pch=16)
plot(t3, n3, xlab = 't', ylab = 'N(t)', xlim = c(0,30), ylim = c(0,20), lwd=1, pch=16)
plot(t4, n4, xlab = 't', ylab = 'N(t)', xlim = c(0,60), ylim = c(0,20), lwd=1, pch=16)
```


Gráfico de Duane

```

x1 <- log(t1); x2 <- log(t2); x3 <- log(t3); x4 <- log(t4);
y1 <- log(n1/t1); y2 <- log(n2/t2); y3 <- log(n3/t3); y4 <- log(n4/t4)
par(mfrow=c(2,2))
plot(x1,y1,xlab = 'log(t)', ylab = 'log(N(t)/t)', lwd=1, pch=16, main='Máquina 1')
plot(x2,y2,xlab = 'log(t)', ylab = 'log(N(t)/t)', lwd=1, pch=16, main='Máquina 2')
plot(x3,y3,xlab = 'log(t)', ylab = 'log(N(t)/t)', lwd=1, pch=16, main='Máquina 3')
plot(x4,y4,xlab = 'log(t)', ylab = 'log(N(t)/t)', lwd=1, pch=16, main='Máquina 4')

```

Gráfico TTT-Plot

```

TTT <- function(falha){
t <- sort(falha)
n <- length(falha)
den <- sum(falha)
phi <- numeric(n)
for (r in 1:n){
soma <- 0
for (i in 1:r){soma <- soma + t[i]}
phi[r] <- (soma + ((n-r)*t[r]))/den}
phi}
TTT1 <- TTT(t1); t1 <- seq(0.001,1,length=length(TTT1))
TTT2 <- TTT(t2); t2 <- seq(0.001,1,length=length(TTT2))
TTT3 <- TTT(t3); t3 <- seq(0.001,1,length=length(TTT3))
TTT4 <- TTT(t4); t4 <- seq(0.001,1,length=length(TTT4))
par(mfrow=c(1,2))
plot(t1,TTT1,ylab="w",xlab="j/n",xlim=c(0,1),ylim=c(0,1),lwd=2,lty=4,col=2,main='Máquina 1')
lines(c(1,0),c(1,0),lwd=2)
plot(t2,TTT2,ylab="w",xlab="j/n",xlim=c(0,1),ylim=c(0,1),lwd=2,lty=4,col=2, main='Máquina 2')
lines(c(1,0),c(1,0),lwd=2)
plot(t3,TTT3,ylab="w",xlab="j/n",xlim=c(0,1),ylim=c(0,1),lwd=2,lty=4,col=2, main='Máquina 3')
lines(c(1,0),c(1,0),lwd=2)
plot(t4,TTT4,ylab="w",xlab="j/n",xlim=c(0,1),ylim=c(0,1),lwd=2,lty=4,col=2, main='Máquina 4')
lines(c(1,0),c(1,0),lwd=2)

```

Gráfico dos tempos de falha dos sistemas

```

require(ggplot2)
ggplot(data=dados) +
geom_point(mapping = aes(x = dados$Ti, y = dados$Trator, group = dados$Trator, shape = dados$Pintar)) +
geom_line(mapping = aes(x = dados$Ti, y = dados$Trator, group = dados$Trator)) +
scale_shape_manual(values=c(19,3)) +
theme_bw() +
xlim(c(0, 260)) +
theme(legend.position="none") +
labs(x = "Tempos de falha", y = "Máquinas")

```

Apêndice III

Métodos de Estimação usados nos Capítulos 3 e 4.

Função Verossimilhança

```
require(bbmle)
dados <- TratSelecionados
# obtendo o número de unidades distintas:
n <- max(dados$Trator)
# separando os tempos observados (falhas e censuras):
t <- split(dados$Tempo,dados$Trator)
# obtendo o número de obsevações por unidade:
m <- tapply(dados$Evento,dados$Trator, sum)
# obtendo o último tempo de acompanhamento:
y <- subset(dados$Tempo, dados$Evento==0)
data <- list(y=y, m=m, t=t)
#### escrevendo a função de verossimilhança Weibull ####
loglik <- function(beta, theta, data){
  l <- rep(0,n)
  for(i in 1:n){
    if(m[i]>0){
      l[i] <- - m[i]*beta*log(theta) + m[i]*log(beta) + (beta-1)*sum(log(t[[i]][1:m[i]])) - (y[i]/theta)^beta
    } else { l[i] <- -(y[i]/theta)^beta }
  }
  return(-sum(l))
}
ajuste <- mle2(loglik, start=list(beta=1,theta=1), data=list(t=t, m=m, y=y), method="L-BFGS-B",
lower=list(beta=1e-32, theta=1e-32), upper=list(beta=Inf,theta=Inf))
(par <- ajuste@coef)
(sd.par <- sqrt(diag(ajuste@vcov)))
(li <- par - qnorm(0.975)*sd.par)
(ls <- par + qnorm(0.975)*sd.par)
(l.vero1 <-loglik(par[1],par[2],data))
d <-2; n <-length(y)
(AIC_W <-2*l.vero1+2*d)
(BIC_W <-2*l.vero1+d*log(n))
#### escrevendo a função de verossimilhança Exponencial (beta=1) ####
loglik <- function( theta, data){
  l <- rep(0,n)
  for(i in 1:n){
    if(m[i]>0){l[i] <- -m[i]*log(theta) - (y[i]/theta)}
    else{l[i] <- -(y[i]/theta)}
  }
  return(-sum(l))
}
ajuste <- mle2(loglik, start=list(theta=0.5), data=list(t=t, m=m, y=y), method="L-BFGS-B", lower=list(theta=1e-32), upper=list(theta=Inf))
(par <- ajuste@coef)
(sd.par <- sqrt(diag(ajuste@vcov)))
(li <- par - qnorm(0.975)*sd.par)
(ls <- par + qnorm(0.975)*sd.par)
```

```
(l.vero2 <- loglik(par[1], data))
n <- length(y)
(AIC_E <- -2*l.vero2 + 2*d)
(BIC_E <- -2*l.vero2 + d*log(n))
```

Apêndice IV

Testes de Hipótese para sistemas reparáveis usada no Capítulo 4.

Teste de igualdade dos thetas - PPH

```
# ordenando os dados
y <- subset(dados$Ti, dados$Evento==0)
(theta_est <- sum(y)/sum(m))
teste_theta <- (2*sum(m))*log(theta_est) - 2*sum(m*log(y/m))
pchisq(teste_theta, df=19, lower.tail=FALSE)
```

Teste de igualdade dos betas - PPNH

```
# ordenando os dados
y_Ti <- subset(dados$Ti, dados$Evento==0)
(beta_est <- m/sum(log(y_Ti/y)))
(teste_beta <- sum(m)*log(sum(m)/sum(m/beta_est)) - sum(m*log(beta_est)))
(a <- 1 + (1/6*19)*(sum(1/m)-(1/sum(m))))
pchisq(-(2*teste_beta)/a, df=19, lower.tail=FALSE)
```

ANEXOS

Anexo A

Teorema 1. *Um processo é dito processo de Poisson homogêneo (PPH) com função intensidade λ se e, somente se, os tempos de falha são variáveis aleatórias iid com distribuição exponencial de média $1/\lambda$ e a fdp conjunta de um PPH para as variáveis aleatórias $0 < T_1 < T_2 < \dots < T_n$ é dada por*

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = \lambda^n \exp(-\lambda t_n), \quad 0 < t_1 < t_2 < \dots < t_n.$$

Demonstração. Assumindo que é um PPH e os tempos de falha são variáveis aleatórias iid com distribuição exponencial de média $1/\lambda$ e $T_1 < T_2 < \dots < T_n$ os tempos de falha. Considere a seguinte definição

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = f_1(t_1)f_2(t_2|t_1)f_3(t_3|t_1, t_2) \cdots f_n(t_n|t_1, t_2, \dots, t_{n-1}),$$

sendo que $0 < t_1 < t_2 < \dots < t_n$. A função de sobrevivência para T_1 é

$$\begin{aligned} S_1(t_1) &= P(T_1 > t_1) \\ &= P(N(0, t_1] = 0) \\ &= \frac{1}{0!} \left(\int_0^{t_1} \lambda dx \right)^0 \exp \left(- \int_0^{t_1} \lambda dx \right) \\ &= \exp \left(- \int_0^{t_1} \lambda dx \right) \\ &= \exp(-\lambda t_1), \end{aligned}$$

e a fdp pode ser escrita como

$$f_1(t_1) = -S_1'(t_1) = \lambda \exp(-\lambda t_1), \quad t_1 > 0.$$

A função de sobrevivência condicional de T_2 dado $T_1 = t_1$ é:

$$\begin{aligned} S_2(t_2|t_1) &= P(T_2 > t_2|t_1) \\ &= P(N(t_1, t_2] = 0) \\ &= \exp \left(- \int_{t_1}^{t_2} \lambda dx \right) \\ &= \exp[-\lambda(t_2 - t_1)], \quad t_2 > t_1 > 0 \end{aligned}$$

e a fdp pode ser escrita como

$$f_2(t_2|t_1) = -\frac{dS_2(t_2|t_1)}{dt_2} = \lambda \exp[-\lambda(t_2 - t_1)], \quad t_2 > t_1 > 0.$$

Pode-se estender esse resultado para o caso geral, pois a propriedade de incrementos independentes garante que a fdp condicional de T_n dado $T_1 = t_1, T_2 = t_2, \dots, T_{n-1} = t_{n-1}$ é independente de t_n . Sendo assim, a fdp de T_n é dada por

$$f_n(t_n|t_1, t_2, \dots, t_{n-1}) = \lambda \exp[-\lambda(t_n - t_{n-1})], \quad t_n > t_{n-1} > 0.$$

Logo, a fdp conjunta de $T_1, T_2, \dots, T_n$ é

$$\begin{aligned} f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) &= \lambda \exp[-\lambda t_1] \lambda \exp[-\lambda(t_2 - t_1)] \cdots \lambda \exp[-\lambda(t_n - t_{n-1})] \\ &= \lambda^n \exp(-\lambda t_n), \quad t_n > t_{n-1} > \dots > t_2 > t_1 > 0 \end{aligned}$$

□

Anexo B

Teorema 2. A fdp conjunta para as variáveis aleatórias com os tempos de falha $T_1, T_2, \dots, T_n$ de um processo de Poisson não homogêneo (PPNH) com função intensidade de falha $\lambda(t)$ é:

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp \left(- \int_0^{t_n} \lambda(x) dx \right)$$

Demonstração. A fdp conjunta para os tempos $T_1, T_2, \dots, T_n$ pode ser escrita como

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = f_1(t_1) f_2(t_2|t_1) f_3(t_3|t_1, t_2) \cdots f_n(t_n|t_1, t_2, \dots, t_{n-1})$$

sendo que $0 < t_1 < t_2 < \dots < t_n$. Assim, a função de sobrevivência para T_1 é

$$\begin{aligned} S_1(t_1) &= P(T_1 > t_1) \\ &= P(N(0, t_1] = 0) \\ &= \frac{1}{0!} \left(\int_0^{t_1} \lambda(x) dx \right)^0 \exp \left(- \int_0^{t_1} \lambda(x) dx \right) \\ &= \exp \left(- \int_0^{t_1} \lambda(x) dx \right), \end{aligned}$$

e a fdp para T_1 é

$$f_1(t_1) = -S_1(t_1) = \lambda(t_1) \exp \left(- \int_0^{t_1} \lambda(x) dx \right), \quad t_1 > 0.$$

A função de sobrevivência condicional de T_2 dado $T_1 = t_1$ é

$$\begin{aligned} S_2(t_2|t_1) &= P(T_2 > t_2|t_1) \\ &= P(N(t_1, t_2] = 0) \\ &= \exp \left(- \int_{t_1}^{t_2} \lambda(x) dx \right), \quad t_2 > t_1 > 0. \end{aligned}$$

Generalizando os resultados anteriores, a função de sobrevivência condicional de T_n dado $T_1 = t_1, T_2 = t_2, \dots, T_{n-1} = t_{n-1}$ é

$$\begin{aligned} S_n(t_n|t_1, t_2, \dots, t_{n-1}) &= P(T_n > t_n|T_{n-1} = t_{n-1}) \\ &= P(N(t_{n-1}, t_n] = 0) \\ &= \exp \left(- \int_{t_{n-1}}^{t_n} \lambda(x) dx \right), \quad t_n > t_{n-1} > 0. \end{aligned}$$

Então, a fdp conjunta de $T_1, T_2, \dots, T_n$ é

$$\begin{aligned} f_{T_1, \dots, T_n}(t_1, \dots, t_n) &= \lambda(t_1) \exp \left[- \int_0^{t_1} \lambda(x) dx \right] \cdots \lambda(t_n) \exp \left[- \int_{t_{n-1}}^{t_n} \lambda(x) dx \right] \\ &= \left(\prod_{i=1}^n \lambda(t_i) \right) \exp \left(- \int_0^{t_n} \lambda(x) dx \right), \quad t_n > t_{n-1} > \dots > t_2 > t_1 > 0 \end{aligned}$$

□

Anexo C

Teorema 3. Considere um PPNH com função intensidade $\lambda(t)$ observado até o tempo t , com tempos de falha $T_1, T_2, \dots, T_{N(t)}$ em que $N(T)$ é uma variável aleatória que descreve o número de falhas no

intervalo $(0, t]$, que condicionado a $N(t) = n$ as variáveis aleatórias $T_1 < T_2 < \dots < T_n$ são distribuídas como n estatísticas de ordem com distribuição acumulada

$$G(y) = \begin{cases} 0, & y \leq 0 \\ \Lambda(y)/\Lambda(t), & 0 < y \leq t \\ 1, & y > t \end{cases}$$

em que $\Lambda(t)$ é o número de falhas esperado até o tempo t e é dado pela Equação 3.1 e $\Lambda(y)$ é o número de falhas esperado até o tempo y .

Demonstração. A fdp conjunta para $T_1 < T_2 < \dots < T_{N(t)}$ dado $N(t) = n$ é

$$\begin{aligned} f(t_1, \dots, t_n | n) &= \lim_{\Delta t_i \rightarrow 0} \frac{P(t_1 < T_1 \leq t_1 + \Delta t_1, \dots, t_n < T_n \leq t_n + \Delta t_n | n)}{\prod_{i=1}^n \Delta t_i} \\ &= \lim_{\Delta t_i \rightarrow 0} \frac{P(t_1 < T_1 \leq t_1 + \Delta t_1, \dots, t_n < T_n \leq t_n + \Delta t_n | n)}{P(N(0, t] = n) \prod_{i=1}^n \Delta t_i} \\ &= \lim_{\Delta t_i \rightarrow 0} \frac{P(N(0, t_1] = 0)P(N(t_1, t_1 + \Delta t_1] = 1)P(N(t_1 + \Delta t_1, t_2] = 0) \dots P(N(t_n, t_n + \Delta t_n] = 1)P(N(t_n + \Delta t_n, t] = 0)}{P(N(0, t] = n) \prod_{i=1}^n \Delta t_i} \\ &= \lim_{\Delta t_i \rightarrow 0} \frac{\exp[-\Lambda(t_1)](\exp[-\Lambda(t_1 + \Delta t_1) + \Lambda(t_1)])[\Lambda(t_1 + \Delta t_1) - \Lambda(t_1)] \dots \exp[-\Lambda(t_n + \Delta t_n) + \Lambda(t_n)][\Lambda(t_n + \Delta t_n) - \Lambda(t_n)]}{\frac{\exp[-\Lambda(t)][\Lambda(t)]^n}{n!} \prod_{i=1}^n \Delta t_i}. \end{aligned}$$

Assim, juntando os termos do expoente e colocando as expressões na forma $\Lambda(t_i + \Delta t_i) + \Lambda(t_i)$ pode-se escrever a densidade condicional da seguinte forma

$$\begin{aligned} &\lim_{\Delta t_i \rightarrow 0} \frac{(\exp[-\Lambda(t_1) - \Lambda(t_1 + \Delta t_1) + \Lambda(t_1) + \dots + \Lambda(t_n) - \Lambda(t) + \Lambda(t_n + \Delta t_n)]) \frac{\Lambda(t_1 + \Delta t_1) - \Lambda(t_1)}{\Delta t_1} \dots \frac{\Lambda(t_n + \Delta t_n) - \Lambda(t_n)}{\Delta t_n}}{\frac{\exp[-\Lambda(t)][\Lambda(t)]^n}{n!}} \\ &= \frac{\exp[-\Lambda(t)] \left(\lim_{\Delta t_1 \rightarrow 0} \frac{\Lambda(t_1 + \Delta t_1) - \Lambda(t_1)}{\Delta t_1} \right) \dots \left(\lim_{\Delta t_n \rightarrow 0} \frac{\Lambda(t_n + \Delta t_n) - \Lambda(t_n)}{\Delta t_n} \right)}{\frac{\exp[-\Lambda(t)][\Lambda(t)]^n}{n!}} \end{aligned}$$

Note que cada um desses limites é a derivada de Λ , então

$$f(t_1, \dots, t_n | n) = \frac{n! \prod_{i=1}^n \Lambda'(t_i)}{[\Lambda(t)]^n} = n! \prod_{i=1}^n \frac{\lambda(t_i)}{\Lambda(t)}.$$

□

Essa expressão é a função de densidade condicionada as n estatísticas de ordem para uma amostra de tamanho n com distribuição acumulada $G(y)$.