

Modelos de Sobrevivência com Base nas Distribuições Geométrica e Exponencial

Cíntia Yurie Yamachi

Orientador: Prof. Dr. Francisco Louzada Neto

Co-Orientadora: Profa. Dra. Maria Aparecida de Paiva Franco

São Carlos

Fevereiro de 2013

Modelos de Sobrevivência com Base nas Distribuições Geométrica e Exponencial

Cíntia Yurie Yamachi

Orientador: Prof. Dr. Francisco Louzada Neto

Co-Orientadora: Profa. Dra. Maria Aparecida de Paiva Franco

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

São Carlos

Fevereiro de 2013

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

Y19ms Yamachi, Cíntia Yurie.
Modelos de sobrevivência com base nas distribuições
geométrica e exponencial / Cíntia Yurie Yamachi. -- São
Carlos : UFSCar, 2013.
90 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2013.

1. Estatística. 2. Distribuição geométrica exponencial. 3.
Análise de sobrevivência. 4. Risco competitivo. I. Título.

CDD: 519.5 (20^a)

Cíntia Yurie Yamachi


Modelos de Sobrevivência com Base nas Distribuições Geométrica e Exponencial

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

Aprovado em 01 de fevereiro de 2013.


BANCA EXAMINADORA

Presidente



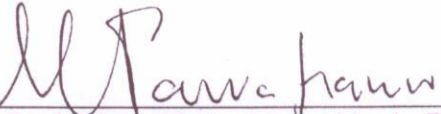
Prof. Dr. Francisco Louzada Neto (ICMC-USP/Orientador)

1º Examinador



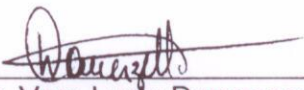
Prof. Dr. Adriano Kamimura Suzuki (ICMC-USP)

2º Examinador



Profa. Dra. Maria Aparecida de Paiva Franco
(DEs-UFSCar/Co-Orientadora)

3º Examinador



Profa. Dra. Vera Lucia Damasceno Tomazella (DEs-UFSCar)

Agradecimentos

Agradeço ao orientador Prof. Dr. Francisco Louzada Neto, a co-orientadora Profa. Dra. Maria Aparecida de Paiva Franco pela orientação e paciência, aos membros da banca: Profa. Dra Vera Lucia Damasceno Tomazella, Prof. Dr. Adriano Kamimura Suzuki pelas sugestões para melhorar a dissertação e na revisão, a Profa. Estela Maris Bereta e colegas da pós-graduação: Mari Roman, Vitor Alex Alves de Marchi pelas discussões, por terem me ajudado a corrigir alguns erros, ao PPGEs (Programa de Pós Graduação em Estatística) e ao auxílio financeiro da Capes.

Resumo

Nesta dissertação são propostos quatro modelos de distribuições de probabilidade para os tempos de vida de indivíduos em uma população. A primeira família de distribuições, a distribuição Geométrica Exponencial Complementar Exponenciada (ECEG) e é obtida via exponenciação da distribuição acumulada da distribuição Geométrica Exponencial Complementar (CEG) proposta por Louzada *et al.* (2011) a um novo parâmetro $\alpha > 0$. A segunda, é direcionada à modelagem de tempos de vida quando a população não é homogênea quanto ao risco de morte possuindo duas subpopulações: a de indivíduos não suscetíveis ao evento e a de indivíduos sob risco. Esta distribuição, distribuição Geométrica Exponencial Complementar Exponenciada na presença de longa duração (LECEG), possui o parâmetro p de longa duração que indica a proporção de indivíduos fora de risco. A terceira é a distribuição Exponencial Geométrica Exponenciada (EEG) que usa a mesma idéia de criação da ECEG, e a quarta a distribuição Exponencial Geométrica Complementar Exponenciada em N sistemas (ECEGN) que se apresenta num cenário com N sistemas funcionando independentemente e a falha ocorre quando algum sistema falhar.

Palavras-Chave: Distribuição, Geométrico Exponencial, Modelos, Sistema, Risco.

Abstract

In this dissertation we propose four models to model lifetime data. The first family of distribution is called Exponentiated Complementary Exponential Geometric distribution (ECEG) and it is obtained by exponentiation of the cumulative distribution of the Complementary Exponential Geometric distribution (CEG) proposed by Louzada *et al.* (2011) to a new parameter $\alpha > 0$. The second distribution is used to model lifetime when the population is not homogeneous about the risk of death and it has two subpopulation: one composed by individuals not susceptible by the event and other composed by individuals subjected to the risk. This model, called LECEG, has a long term parameter p related to the proportion of individuals out of risk. The third is the Exponentiated Exponential Geometric (EEG) that uses the same idea of the ECEG, and the fourth is the Exponentiated Complementary Exponential Geometric distribution under N systems (ECEGN) presented in a context of N independent working systems and the fails occurs when some of them fail.

Keywords: Distribution, Exponential Geometric, Models, Sistem, Risk.

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Conceitos básicos | 3 |
| 1.2 | Modelo de longa duração | 6 |
| 1.3 | Inferência | 8 |
| 1.4 | Objetivos do Trabalho | 9 |
| 1.5 | Organização dos Capítulos | 10 |
| 2 | Distribuição geométrica exponencial complementar exponenciada | 12 |
| 2.1 | Origem | 12 |
| 2.2 | Propriedades | 14 |
| 2.3 | Inferência | 20 |
| 2.4 | Estudo de simulação | 21 |
| 2.4.1 | Estudo das propriedades dos estimadores por M.V dos parâmetros . | 21 |
| 2.5 | Aplicação | 22 |
| 2.5.1 | Comparação dos ajustes para 3 distribuições, a dois conjuntos de dados simulados da ECEG | 22 |
| 2.5.2 | Exemplo com dados reais a distribuição ECEG e outras distribuições | 24 |
| 2.6 | Modelo de regressão geométrico exponencial complementar exponenciado . | 26 |
| 2.6.1 | Dados Reais | 28 |
| 2.6.2 | Dados simulados com covariável | 31 |
| 2.7 | Comentários Finais | 33 |
| 3 | Distribuição Geométrica Exponencial Complementar Exponenciada com Termo de Longa Duração | 34 |
| 3.1 | Origem | 34 |
| 3.2 | Propriedades | 35 |
| 3.3 | Inferência | 40 |

| | | |
|----------|--|-----------|
| 3.4 | Estudo de simulação | 41 |
| 3.4.1 | Estudo das propriedades dos estimadores por M.V. dos parâmetros | 41 |
| 3.5 | Estudo de simulação dos ajustes para quatro distribuições, a um conjunto de dados simulados da LECEG | 43 |
| 3.5.1 | Ajuste a conjunto de dados reais da distribuição LECEG e outras distribuições | 45 |
| 3.6 | Modelo de regressão geométrico exponencial exponenciado de longa duração | 47 |
| 3.6.1 | Dados Reais | 48 |
| 3.6.2 | Dados Simulados com covariável | 50 |
| 3.7 | Comentários Finais | 54 |
| 4 | Distribuição Exponencial Geométrica Exponenciada | 55 |
| 4.1 | Origem | 55 |
| 4.2 | Propriedades | 56 |
| 4.3 | Inferência | 59 |
| 4.4 | Estudo de simulação | 60 |
| 4.4.1 | Estudo das propriedades dos estimadores por M.V. dos parâmetros | 60 |
| 4.5 | Aplicação | 61 |
| 4.5.1 | Ajuste a conjunto de dados reais da distribuição EEG e outras distribuições | 61 |
| 4.6 | Comentários Finais | 64 |
| 5 | Distribuição Exponencial geométrica complementar exponenciada considerando mais de um sistema | 65 |
| 5.1 | Origem | 65 |
| 5.2 | Propriedades | 66 |
| 5.3 | Inferência | 71 |
| 5.4 | Estudo de simulação | 72 |
| 5.4.1 | Estudo das propriedades dos estimadores por M.V dos parâmetros . | 72 |
| 5.5 | Aplicação | 73 |
| 5.5.1 | Ajuste a conjunto de dados reais da distribuição ECEGN e outras distribuições | 73 |
| 5.6 | Comentários Finais | 75 |

6 Conclusão

Lista de Figuras

| | | |
|-----|--|----|
| 2.1 | Funções de risco crescente, decrescente e em forma de banheira da distribuição ECEG | 16 |
| 2.2 | Assimetria e Curtose em distribuições da distribuição ECEG | 17 |
| 2.3 | TTTplot para os dados D1 (painel esquerda) e D2 (painel direita) | 23 |
| 2.4 | Curvas de Kaplan Meier para os dados D1 (painel esquerdo) e para os dados D2 (painel direito) | 24 |
| 2.5 | TTT plot e função de sobrevivência para dados reais | 26 |
| 2.6 | Resíduos de Cox Snell <i>versus</i> sobrevivência estimada para os modelos ECEG, E padrão, CEG e E | 30 |
| 2.7 | Resíduos de Cox Snell versus Sobrevivência estimada para o modelo ECEG, CEG e E | 32 |
| 3.1 | Função de densidade e de sobrevivência da distribuição LECEG | 36 |
| 3.2 | Função de risco da distribuição LECEG | 37 |
| 3.3 | Valores da assimetria e curtose da distribuição LECEG para diversos valores | 38 |
| 3.4 | TTT plot (painel esquerdo) e Curva Kaplan Meier com as funções de sobrevivências estimadas de LECEG, LCEG, LE, LEGU e LLI (painel direito). | 44 |
| 3.5 | TTT plots de Allison e Prentice 1973. | 46 |
| 3.6 | Curva Kaplan Meier com as funções de sobrevivências estimadas de LECEG, LCEG, LE, LEGU e LLI para o conjunto de dados de Allison (painel esquerdo) e Prentice(painel direito). | 47 |
| 3.7 | Resíduos de Cox Snell <i>versus</i> Sobrevivência estimada para os modelos LECEG, LCEG e LE | 50 |
| 3.8 | Resíduos de Cox Snell versus Sobrevivência estimada dos resíduos para o modelo LECEG, LCEG e LE para os dados simulados com regressão em p | 52 |

| | | |
|-----|--|----|
| 3.9 | Resíduos de Cox Snell versus Sobrevivência estimada dos resíduos para o modelo LECEG, LCEG e LE para os dados simulados com regressão em λ | 54 |
| 4.1 | Funções de risco de EEG | 57 |
| 4.2 | TTT plot (painel esquerdo), Kaplan Meier e curva de sobrevivência ajustada para T1 (painel direito). | 63 |
| 4.3 | TTT plot (painel esquerdo), Kaplan Meier e curva de sobrevivência ajustada para T2 (painel direito). | 63 |
| 5.1 | Funções de risco da ECEGN | 67 |
| 5.2 | TTT plot (painel esquerdo) e Kaplan Meier com função de sobrevivência ajustada para cada distribuição (painel direito). | 74 |

Lista de Tabelas

| | | |
|------|--|----|
| 2.1 | Tabela com a taxa de rejeição | 21 |
| 2.2 | Resultados das simulações. | 22 |
| 2.3 | Tempos de vida dados D1 | 23 |
| 2.4 | Tempos de vida dados D2 | 23 |
| 2.5 | Valores dos critérios para a amostra simulada - Dados D1 e D2 | 24 |
| 2.6 | Tempos de vida Lawless | 25 |
| 2.7 | Estimativas dos parâmetros e critérios | 25 |
| 2.8 | Descrição das variáveis dos dados de Hosmer & Lemeshow (2008) | 28 |
| 2.9 | Valores estimados para o modelo ECEG, CEG e E | 29 |
| 2.10 | Tempos de vida | 31 |
| 2.11 | Valores da covariável x | 31 |
| 2.12 | Valores estimados para o modelo ECEG, CEG e E | 31 |
| 3.1 | Resultados das simulações. | 42 |
| 3.2 | Dados gerados da distribuição LECEG | 43 |
| 3.3 | Estimativas dos parâmetros aos modelos LECEG, LCEG, LE, LEGU e LLI | 44 |
| 3.4 | Conjunto de dados de Allison | 45 |
| 3.5 | Conjunto de dados de Prentice | 45 |
| 3.6 | Estimativas dos parâmetros dos conjuntos de dados de Allison aos modelos LECEG, LCEG, LE, LEGU e LLI | 45 |
| 3.7 | Estimativas dos parâmetros dos conjuntos de dados de Prentice aos modelos LECEG, LCEG, LE, LEGU e LLI | 46 |
| 3.8 | Variáveis do livro Kalbfleisch & Prentice (1980) | 48 |
| 3.9 | Valores estimados para o modelo LECEG, LCEG e LE | 49 |
| 3.10 | Tempos de vida | 50 |
| 3.11 | Valores da covariável x | 51 |

| | | |
|------|--|----|
| 3.12 | Valores estimados para o modelo LECEG, LCEG e LE | 51 |
| 3.13 | Tempos de vida | 52 |
| 3.14 | Valores da covariável x | 53 |
| 3.15 | Valores estimados para o modelo LECEG, LCEG e LE | 53 |
| 4.1 | Média das EMV, probabilidade de cobertura e EQM. | 61 |
| 4.2 | Ajuste de seis distribuições ao conjunto de dados T1: estimativas dos parâmetros dos modelos E, EE, EG, ExtEG, EEG e Gama | 62 |
| 4.3 | Ajuste de seis distribuições ao conjunto de dados T2: estimativas dos parâmetros dos modelos E, EE, EG, ExtEG, EEG e Gama | 63 |
| 5.1 | Média das EMV, probabilidade de cobertura e EQM. | 73 |
| 5.2 | Ajuste de quatro distribuições ao conjunto de dados T1: estimativas dos parâmetros dos modelos ECEGN, ExtEG, E, e Lindley | 75 |
| 5.3 | Ajuste das quatro distribuições ao conjunto de dados T2: estimativas dos parâmetros dos modelos ECEGN, ExtEG, E, e Lindley | 75 |
| 1 | Valores do TRV e p-valor para cada ajuste ECEG | 87 |
| 2 | Valores do TRV e p-valor para cada ajuste | 87 |
| 3 | Valores do TRV e p-valor para cada ajuste | 88 |
| 4 | Valores do TRV e p-valor para cada ajuste | 89 |
| 5 | Valores do TRV e p-valor para cada ajuste | 89 |
| 6 | Valores do TRV e p-valor para cada ajuste | 90 |

Capítulo 1

Introdução

Distribuições de probabilidade são utilizadas em análise de sobrevivência para modelar tempos de vida. Uma distribuição bastante utilizada nas aplicações é a exponencial que possui um único parâmetro. Ela é caracterizada por possuir falta de memória, ou seja, a probabilidade condicional $P(T > s + t | T > s)$ é igual a $P(T > t)$ para todo $s \geq 0$ e $t \geq 0$ e por possuir função de risco constante.

Adamidis & Loukas (1998) propuseram a distribuição Geométrica Exponencial(EG), cuja motivação é a duração de um sistema em série. A primeira falha de alguma das causas $Y = \min(T_1, \dots, T_p)$ é a falha do sistema; em que o número de componentes segue distribuição geométrica com parâmetro θ . O tempo de duração de cada componente segue distribuição exponencial de parâmetro λ .

Em situações de risco competitivos em que se observa o menor tempo de vida sob todos os riscos $Y = \min(T_1, \dots, T_p)$, Adamidis & Loukas (1998) abordaram o cenário explicando que a situação é conhecida como 'work hardening' (endurecimento) ou 'immunity' (imunidade), ou seja, existe um aprendizado nos tempos de vida. Chahkandi & Ganjali (2009) trabalham com distribuições cuja motivação é a duração de um sistema em série em que a quantidade de componentes é assumida possuir uma das seguintes distribuições discretas: Logaritmica, Binomial, Poisson, e Binomial Negativa truncadas no zero.

Kus (2007) apresenta a distribuição Poisson-Exponencial inserida num sistema em série no qual a ocorrência da falha é a causa de falha no sistema. Tahmasbi & Rezaei (2008) com a distribuição Logaritmica Exponencial inserida num sistema em série.

A distribuição Exponencial-Poisson introduzida por Cancho *et al.* (2011) modela o tempo de duração em um sistema em paralelo, cujo número de componentes segue distribuição Poisson truncada no zero. A distribuição obtida possui taxa de risco crescente.

Louzada *et al.* (2011) introduziram a distribuição geométrica exponencial complementar, cuja motivação é a duração de um sistema em paralelo em que a falha do sistema ocorre devido a falha de todas as causas sendo a última falha a causadora da falha $Y = \max(T_1, \dots, T_p)$, em que o número de componentes segue distribuição geométrica com parâmetro θ . O tempo de duração de cada componente segue distribuição exponencial de parâmetro λ .

A distribuição Weibull-Poisson, apresentada por Hemmati *et al.* (2011) está inserida num sistema em série, é capaz de modelar taxa de risco crescente, decrescente e unimodal.

Mudholkar *et al.* (1995) apresentaram uma nova distribuição para o tempo de vida de um indivíduo a partir da Weibull elevando a função acumulada da Weibull a um novo parâmetro não negativo. Gupta & Kundu (1999) criaram outra distribuição elevando a distribuição acumulada de uma distribuição Exponencial com parâmetro de locação e escala a um novo parâmetro positivo. Barriga *et al.* (2011) apresentaram a distribuição Exponencial Potência Complementar (CEP) que consiste em elevar a distribuição acumulada da distribuição Exponencial Potência proposta por Smith & Bain (1975) a um novo parâmetro. Bakouch *et al.* (2011) criaram uma nova distribuição a partir da Lindley denotada como uma extensão da distribuição Lindley. Mais informações sobre distribuições em cenários de riscos complementares ou competitivos podem ser encontradas nos livros: Basu & Klein (1982), Lawless (2003), Crowder *et al.* (1991) e Cox & Oakes (1984).

Existem modelos em análise de sobrevivência para situações onde uma parte da população ($p\%$) não está suscetível ao evento de interesse. Na amostragem de uma população desse tipo, o pesquisador não sabe identificar quais são os indivíduos não suscetíveis dentro de sua amostra e fará uso de todas suas observações para inferir sobre a distribuição do tempo de vida em uma população. Os trabalhos mais conhecidos nesse tipo de estudo são de Boag (1949) e Berkson & Gage (1952) que estudaram um modelo misto com taxa de cura. Desde então muitos outros autores como Maller & Zhou (1995), Cancho *et al.* (2009) e Perdoná & Louzada-Neto (2011) trabalharam com modelos desse tipo. Existe um método de estimação bastante utilizado que é o método da máxima verossimilhança que consiste em maximizar o logaritmo da verossimilhança dos dados. Cancho *et al.* (2009), Maller & Zhou (1995), Louzada-Neto (1999), Boag (1949) e Roman *et al.* (2012) utilizaram essa metodologia de estimação em seus modelos.

Outros autores fazem o uso do algoritmo EM, método iterativo utilizado quando se tem dados faltantes, baseado em uma distribuição de dados hipotética completa, que consiste em dois passos: o passo E e o passo M. No passo E é encontrado a esperança condicional da função de densidade dos dados completos e no passo M os dados faltantes são substituídos pela esperança condicional.

Adamidis & Loukas (1998), Karlis (2003), Ng *et al.* (2002), Cancho *et al.* (2011), Chahkandi & Ganjali (2009), Tahmasbi & Rezaei (2008), e Hemmati *et al.* (2011) utilizaram essa metodologia de estimação em seus modelos.

Em seguida apresentamos alguns conceitos básicos necessários para o entendimento do trabalho.

1.1 Conceitos básicos

Análise de sobrevivência é o estudo de tempo de vida entre uma data inicial (início do funcionamento de um equipamento, nascimento de um indivíduo, data de diagnóstico) até a ocorrência de um evento (falha de um equipamento, morte de um indivíduo, morte de um indivíduo por uma causa específica etc). Seja T uma variável aleatória denotando este tempo de vida. Sob condições de amostragem aleatória desta variável T em uma população homogênea quanto a sobrevivência, é possível estimar o modelo de probabilidade que governa a variável aleatória T e, uma vez escolhido este modelo, estimar os seus parâmetros.

Na coleta de dados para estimação nem sempre é possível observar o tempo de ocorrência T do evento para cada indivíduo seja por ele ter saído do estudo ou sofrido outro evento. No entanto é possível observar por quanto tempo (T^*) o indivíduo esteve sob observação. Em resumo é possível observar em cada indivíduo T ou T^* . Se for observado T^* em um indivíduo diz-se que foi obtido uma observação censurada de T .

Usualmente são considerados 3 mecanismos possíveis de censura:

Censura tipo I: O estudo é realizado até um período de tempo pré fixado. Nos indivíduos que não tenham apresentado o evento, será observado a variável $Y = C$ e nos demais será observado $Y = T$.

Censura tipo II: Acompanha-se N indivíduos até se obter uma quantidade pré fixada m de indivíduos que apresentem o evento. Isto resultará em m observações de T e $N - m$ observações de C .

Censura tipo III ou aleatória: É comum supor que existe um mecanismo de censura que ocorre num tempo aleatório C para cada indivíduo e que só é possível observar o menor dos valores T ou T^* , isto é, observa-se $Y = \min(T, C)$. A estimação dos parâmetros da distribuição de T será feita a partir de uma amostra das observações de Y , em n indivíduos. A variável indicadora δ_i é representada por:

$$\delta_i = \begin{cases} 1, & \text{se } Y = T; \text{ para } i=1,2,\dots,n \\ 0, & \text{se } Y = C, \end{cases} \quad (1.1)$$

A distribuição de probabilidade da variável T é o objeto de interesse em toda a análise de sobrevivência. Ela pode ser especificada pela função de sobrevivência

$$S(t) = P(T > t)$$

ou por alguma função derivada da mesma.

A função acumulada é definida por:

$$F(t) = P(T \leq t) = \int_0^t f(u)du,$$

ou pode ser encontrada como $F(t) = 1 - S(t)$.

A função de densidade de probabilidade é encontrada como:

$$f(t) = -\frac{dS(t)}{dt},$$

ou

$$f(t) = \frac{dF(t)}{dt}.$$

A função de risco ou taxa de falha pode ser encontrada a partir da função densidade e de sobrevivência $h(t) = f(t)/S(t)$ ou pela seguinte definição:

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

Muitas vezes o tempo de vida observável T de um indivíduo é o seu tempo de sobrevivência quando ele está sujeito a diversas causas de falha ou de morte. Existem dois modelos bastante difundidos para o efeito das diversas causas. O primeiro deles chamado de modelo de riscos complementares (C) admite que $T_C = \max(T_1, T_2, \dots, T_p)$, $j = 1, \dots, p$, sendo cada T_j o tempo de ocorrência até a falha pela causa j . Neste modelo

o indivíduo só apresenta o evento final se ocorrerem as falhas por todas as p causas. Esse modelo está associado ao tempo de duração de um circuito em paralelo, que funciona enquanto um de seus componentes estiver funcionando. No modelo de riscos competitivos (RC) $T_{RC} = \min(T_1, T_2, \dots, T_p)$ o que significa que o indivíduo morre quando ocorre a primeira causa de morte e isto corresponde ao tempo de vida de um circuito em série. Sob hipótese de independência de T_1, \dots, T_p e admitindo que as variáveis sejam identicamente distribuídas é imediato provar que em riscos complementares (T_C):

$$\begin{aligned} F_{T_C}(t) &= P(T_C \leq t) \\ &= P(T_1 \leq t, T_2 \leq t, \dots, T_p \leq t) \\ &= (P(T_C \leq t))^p \\ &= (F_{T_j}(t))^p. \end{aligned}$$

Em riscos competitivos vale a seguinte relação:

$$\begin{aligned} S(T_{RC}(t)) &= P(T_{RC} > t) \\ &= P(T_1 > t, T_2 > t, \dots, T_p > t) \\ &= P(T_1 > t)P(T_2 > t) \dots P(T_p > t) \\ &= (S_{T_j}(t))^p. \end{aligned}$$

Uma maneira de verificar a forma da densidade são medidas de assimetria γ_1 e de curtose γ_2 . Outra medida utilizada para verificar o tempo médio de vida médio restante de um indivíduo que sobreviveu até t é através do tempo de vida média residual. Essas medidas são encontradas como:

$$\begin{aligned} \mu_3 &= E[(Y - E(Y))^3] = \int_{-\infty}^{+\infty} (y - \mu)^3 f(y) dy, \\ \sigma^3 &= (E([(X - \mu)^2]))^{\frac{3}{2}}, \\ \gamma_1 &= \frac{\mu_3}{\sigma_3}, \\ \mu_4 &= E[(Y - E(Y))^4] = \int_{-\infty}^{+\infty} (y - \mu)^4 f(y) dy, \\ \sigma^4 &= (E([(X - \mu)^2]))^{\frac{4}{2}}, \\ \gamma_2 &= \frac{\mu_4}{\sigma_4} - 3. \end{aligned}$$

O tempo de vida média residual é obtido da seguinte forma:

$$\mu(y) = E(T - t_0 | T \geq t_0) = \frac{1}{S(t_0)} \int_t^\infty S(u) du.$$

Um coeficiente de assimetria (γ_1) negativo indica que a cauda da fdp no lado esquerdo é maior que no lado direito. Um coeficiente de assimetria positivo indica que a cauda da fdp no lado direito é mais longa que no lado esquerdo. O valor zero do coeficiente de assimetria indica que os dados estão bem distribuídos em torno da média, mas não necessariamente apresenta uma distribuição simétrica. A interpretação da curtose é ambigua, mas quando a função densidade for simétrica, valores altos de curtose indicam um pico maior e caudas longas, enquanto uma curtose baixa indica pico mais arredondado, curto e caudas finas.

1.2 Modelo de longa duração

Os modelos de longa duração são utilizados quando há uma grande quantidade de observações censuradas na amostra o que indica que parte da população está imune ao evento. Um dos primeiros trabalhos a utilizar essa abordagem foi o de Boag (1949) com o método de máxima verossimilhança para estimar a quantidade de sobreviventes em uma população de mulheres com câncer de mama. Outro trabalho bem reconhecido é o de Berkson & Gage (1952) que propuseram um modelo de mistura com o objetivo de estimar a proporção de curados em uma amostra de pessoas da clínica Mayo com câncer de estômago. Berkson & Gage (1952) trabalharam em um modelo de mistura de exponenciais e, para permitir a proporção de cura, usaram uma distribuição degenerada.

Recentemente na literatura há um crescente desenvolvimento de modelos mais complexos de longa duração, como por exemplo, Cancho *et al.* (2009), Perdoná & Louzada-Neto (2011), Rodrigues *et al.* (2009), Yakovlev *et al.* (1996), Farewell (1977), Chen *et al.* (1999), Ghitany *et al.* (1994), Perperoglou *et al.* (2007), Pons & Lemdani (2003), entre outros.

O modelo de longa duração é comumente usado em confiabilidade para ser ajustado em tempos de vida de componentes. Além disso pode ser utilizado com outros tipos de dados, por exemplo, Anscombe (1961) utilizou um modelo de mistura exponencial, trabalhando com dados no qual verificava se determinado produto lançado no mercado foi bem aceito através de uma pesquisa em que foi perguntado o dia em que a pessoa fez a pri-

meira compra do produto, no qual as censuras eram os indivíduos pesquisados que ainda não compraram o produto. Gulland (1955) usou o modelo de longa duração em dados de pesquisa de pesca de peixes que foram marcados, depois de um certo tempo recapturados e deixados de volta a natureza, e os censurados seriam os peixes nunca recapturados.

Este modelo pode ser utilizado em áreas como criminologia no estudo da proporção de vezes que o indivíduo retorna a prisão devido a algum delito. O primeiro trabalho com modelo considerando a proporção de imunes foi o de Pons & Lemdani (1969) que ajustaram o modelo misto exponencial aos dados de indivíduos que retornaram a prisão no período de 1949-1951.

O modelo considera uma parte da população subdividida em dois grupos indicados por OR e IR: indivíduos não suscetíveis a ocorrência do evento de interesse (OR) e indivíduos que em algum momento sofrerão o evento (IR). Em uma amostra casual de tamanho 1 desta população seja T a variável que represente o tempo até a ocorrência do evento no indivíduo amostrado e p a probabilidade do indivíduo pertencer ao grupo OR. O pesquisador não tem meios de reconhecer se o indivíduo pertence a OR ou IR. A função de sobrevivência de T , $S(t)$, pode ser escrita como uma combinação das funções de sobrevivências em cada subpopulação, indicadas por $S_{OR}(t)$ e $S_{IR}(t)$.

O modelo de mistura padrão pode ser obtido considerando uma variável Bernoulli M_i para os indivíduos sob riscos e os que não estão sob risco, ou seja,

$$M_i = \begin{cases} 1, & \text{se o indivíduo está sob risco;} \\ 0, & \text{se o indivíduo não está sob risco,} \end{cases}$$

com $P(M_i = 0) = p$ e $P(M_i = 1) = 1 - p$, para $i=1,2,\dots,n$.

Seja T a variável aleatória não negativa e contínua, representando o tempo de vida, tem-se:

$$P(T > t|M_i = 1) = S(t) \quad \text{e} \quad P(T > t|M_i = 0) = 1,$$

para $i=1,2,\dots,n$.

A função de sobrevivência da população é escrita como:

$$\begin{aligned} S_{pop}(t) &= P(T > t|M_i = 0)P(M_i = 0) + P(T > t|M_i = 1)P(M_i = 1) \\ &= p + (1 - p)S(t). \end{aligned}$$

Assim, usa-se a função de sobrevivência imprópria utilizada por Berkson & Gage (1952) dada por:

$$S_{pop}(y) = pS_{OR}(y) + (1 - p)S_{IR}(y) \quad (1.2)$$

A função (1.2) possui as seguintes propriedades:

- Se $p = 0$, então $S_{pop}(t) = S(t)$;
- $S_{pop}(t)$ é decrescente;
- $\lim_{t \rightarrow \infty} S_{pop}(t) = p$.

A função de densidade da população imprópria é dada por:

$$f_{pop}(t) = f(t)(1 - p),$$

em que $f(\cdot)$ representa a função de densidade própria relativa ao grupo de indivíduos em risco.

A função de risco é dada por:

$$h_{pop}(t) = f(t) \frac{1 - p}{p + (1 - p)S(t)}. \quad (1.3)$$

1.3 Inferência

Para a estimação dos parâmetros usa-se o método de máxima verossimilhança que utiliza o logaritmo da verossimilhança. Seja T uma variável aleatória que represente o tempo de vida de indivíduos e C a variável aleatória indicadora de falha ou censura do indivíduo no estudo. Supondo independência entre as variáveis T e C , através de uma amostra aleatória de tamanho n de $Y = \min(T, C)$, e considerando δ a função indicadora do evento $Y = T$, a função de verossimilhança é dada por:

$$L(\underline{y}, \underline{\delta}, \theta) = \prod_{i=1}^n (f(y_i, \theta))^{\delta_i} (S(y_i, \theta))^{(1-\delta_i)}. \quad (1.4)$$

O logaritmo da verossimilhança é escrito como:

$$l(\underline{y}, \underline{\delta}, \theta) = \sum_{i=1}^n \delta_i \log(f(y_i, \theta)) + \sum_{i=1}^n (1 - \delta_i) \log(S(y_i, \theta)).$$

Para estimar a forma da função de risco a partir de uma amostra de $Y = \min(T, C)$, utiliza-se a metodologia gráfica, como o TTT plot introduzido por Barlow *et al.* (1972). A versão empírica do TTT plot é escrita como $G(r/n) = [(\sum_{i=1}^r Y_{i:n}) - (n-r)Y_{r:n}] / (\sum_{i=1}^r Y_{i:n})$, $r = 1, \dots, n$ e $Y_{i:n}$ representa a i -ésima estatística de ordem da amostra. O TTT plot é um gráfico de dispersão de r/n versus $G(r/n)$. Prova-se que quando função de risco é crescente(decrecente) o TTT plot é côncavo(convexo).

Para comparação de distribuições usam-se o $-\log(L(\hat{\vartheta}))$ e dois outros critérios definidos respectivamente por: $-2\log(L(\hat{\vartheta})) + 2q$ denotado por critério de informação Akaike (Hirotsugu (1974)) e $-2\log(L(\hat{\vartheta})) + q\log(n)$ chamado de critério de informação Bayesiano (Schwarz (1978)). $\hat{\vartheta} = (\hat{\lambda}, \hat{\theta})$ é o vetor de estimativas de máxima verossimilhança sobre a distribuição estudada, q é o número de parâmetros estimados sobre a distribuição estudada e n é o tamanho amostral. A melhor distribuição ajustada é aquela que possui os menores valores $-\log(\hat{\vartheta})$, AIC e BIC.

Além dos critérios AIC e BIC, outra forma de selecionar modelos é por meio de teste de hipótese. Testam-se as seguintes hipóteses:

H_0 : O modelo de interesse é adequado. *versus* H_1 : O modelo não é adequado.

A estatística da razão de verossimilhança é dada por:

$$TRV = -2\log \left[\frac{L(\hat{\theta}_I)}{L(\hat{\theta}_G)} \right] = 2 \left[\log(L(\hat{\theta}_G)) - \log(L(\hat{\theta}_I)) \right]. \quad (1.5)$$

Em (1.5), $\log(L(\hat{\theta}_G))$ é o logaritmo de verossimilhança do modelo ajustado geral e $\log(L(\hat{\theta}_I))$ é o logaritmo da verossimilhança do modelo ajustado de interesse.

Esta estatística sob H_0 tem aproximadamente uma distribuição qui-quadrado com graus de liberdade igual a diferença do número de parâmetros $(\hat{\theta}_G$ e $\hat{\theta}_I)$ dos modelos ajustados.

1.4 Objetivos do Trabalho

O objetivo desse trabalho é apresentar uma nova distribuição obtida a partir da função acumulada da CEG introduzida por Louzada *et al.* (2011) que pode ser ajustada em um cenário de taxa de risco crescente, decrescente ou em forma de banheira chamada de Geométrica Exponencial Complementar Exponenciada (ECEG), assim como introduzir novas outras distribuições provindas da Exponencial Geométrica. A distribuição

Geométrica Exponencial Exponenciada (EEG) e a distribuição Geométrica Exponencial Complementar Exponenciada em N sistemas (ECEGN).

1.5 Organização dos Capítulos

No Capítulo 1 são abordados conceitos utilizados em análise de sobrevivência, definições dos tipos de dados, conceitos de medidas como a assimetria, curtose que serão utilizados nos capítulos seguintes, duas formas de geração de dados são explicadas, a conhecida como método da inversa e a por simulação aceitação-rejeição, conceito de risco competitivo e complementar, a estimação por maximização da verossimilhança e o teste de razão de verossimilhanças.

No Capítulo 2 apresenta-se a origem da distribuição ECEG e suas características, tais como, função distribuição acumulada, função de sobrevivência, função densidade, função quantil, função de risco, função característica, função geradora de momentos, medidas de assimetria e curtose são determinados. A seguir são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição ECEG. Apresenta-se o logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição ECEG, um estudo de simulação é realizado calculando o vício médio e a média do erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos amostrais da distribuição ECEG. A distribuição ECEG é ajustada a dados reais e simulados e, verificada sua performance por meio dos critérios AIC , BIC e o teste de Kolmogorov Smirnov. Por fim é apresentado o modelo de regressão ECEG no qual as covariáveis são incluídas no parâmetro de escala.

No Capítulo 3 apresenta-se a origem da distribuição LECEG e suas características, tais como, função de sobrevivência, função densidade, função de risco, função quantil, função característica, momentos, medidas de assimetria e curtose são determinados. A seguir são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição LECEG. Apresenta-se o logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição LECEG, um estudo de simulação é realizado calculando o vício médio e a média do erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos de amostras da distribuição LECEG. A distribuição

LECEG é ajustada a dados reais e simulados e, verificada sua performance por meio dos critérios *AIC*, *BIC* e o teste de Kolmogorov Smirnov. Assim como no Capítulo 2, no modelo de regressão LECEG é assumido a influência das covariáveis no parâmetro de escala.

No Capítulo 4 uma generalização da distribuição EG é apresentada, denotada por distribuição EEG é capaz de modelar função de risco crescente, decrescente e unimodal. Neste capítulo a origem da distribuição EEG e suas características, tais como, função distribuição acumulada, função de sobrevivência, função densidade, função quantil e função de risco. A seguir são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição EEG. Apresenta-se o logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição EEG, um estudo de simulação é realizado calculando a média das EMV e o erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos de amostrais da distribuição EEG. A distribuição EEG é ajustada a dados reais e simulados e verificada sua performance por meio dos critérios *AIC*, *BIC* e o teste de Kolmogorov Smirnov.

No Capítulo 5 uma extensão da distribuição ECEG é apresentada. Denotada por distribuição ECEGN é capaz de modelar função de risco crescente e decrescente. Neste capítulo a origem da distribuição ECEGN e suas características, tais como, função distribuição acumulada, função de sobrevivência, função densidade, função quantil e função de risco. A seguir são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição ECEGN. Apresenta-se o logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição ECEGN. A distribuição ECEGN é ajustada a dados reais e verificada sua performance por meio dos critérios *AIC*, *BIC* e o teste de Kolmogorov Smirnov.

No Capítulo 6 concluímos o trabalho descrevendo as novas distribuições apresentadas ao longo dos capítulos que podem ser utilizadas na análise de sobrevivência.

Capítulo 2

Distribuição geométrica exponencial complementar exponenciada

Este Capítulo 2 a origem da distribuição ECEG, e suas características, tais como, função acumulada, função de sobrevivência, função densidade, função quantil, função de risco, função característica, função geradora de momentos, medidas de assimetria e curtose são determinados na Seção (2.1). Em seguida são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição ECEG. O logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição ECEG é apresentado na Seção (2.3). Um estudo de simulação é realizado calculando o vício médio e a média do erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos de amostras da distribuição ECEG na Seção (2.4). A distribuição ECEG é ajustada a dados reais e simulados e, verificada sua performance por meio dos critérios $-\log(\cdot)$, AIC , BIC e o teste de Kolmogorov Smirnov na Seção (2.5). Por fim, na Seção (2.6) é apresentado o modelo de regressão ECEG no qual as covariáveis são incluídas no parâmetro de escala.

2.1 Origem

Marshall & Olkin (2007) usam a idéia de criar uma nova função distribuição acumulada elevando uma função distribuição acumulada base (F_{base}) a uma potência $\alpha > 0$, denominando α de "parâmetro de resiliência". Resiliência em física é a propriedade que alguns materiais tem de acumular energia sem entrar em ruptura, ou seja, é a capacidade do material voltar ao normal depois de ser submetidos a uma tensão. A distribuição nova

é definida por:

$$F(y) = (F_{base}(y))^\alpha. \quad (2.1)$$

Barriga *et al.* (2011) comentam a importância do parâmetro de resiliência, lembrando que a duração de um sistema em paralelo com n componentes independentes com duração aleatória de distribuição $F(\cdot)_{base}$, é uma variável aleatória com distribuição $F(y) = (F_{base}(y))^n$. Logo, o modelo com $\alpha > 0$ não necessariamente inteiro, generaliza o modelo conhecido de duração em um sistema em paralelo.

A distribuição geométrica exponencial complementar (CEG) introduzida por Louzada *et al.* (2011) possui um parâmetro de escala (λ) e um de forma (θ), e sua função distribuição acumulada é dada por:

$$F_{base}(y) = \frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1 - \theta) + \theta} I_{[0, \infty)}(y). \quad (2.2)$$

A distribuição Geométrica exponencial complementar exponenciada (ECEG) é obtida considerando como função base $F(\cdot)_{base}$ em (2.1) e a função distribuição acumulada da distribuição CEG (2.2) apresentada por Louzada et al. (2011). Desse modo, a função acumulada ECEG é escrita como:

$$F(y) = \left(\frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1 - \theta) + \theta} \right)^\alpha I_{[0, \infty)}(y). \quad (2.3)$$

A função de sobrevivência ou confiabilidade é dada por:

$$S(y) = I_{(-\infty, 0]}(y) + \left(1 - \left(\frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1 - \theta) + \theta} \right)^\alpha \right) I_{[0, \infty)}(y). \quad (2.4)$$

A função de densidade da ECEG é dada por:

$$f(y) = \alpha \left(\frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1 - \theta) + \theta} \right)^{\alpha-1} \left(\frac{\theta \lambda e^{-\lambda y}}{(e^{-\lambda y} (1 - \theta) + \theta)^2} \right) I_{[0, \infty)}(y). \quad (2.5)$$

Em (2.5), α e θ são os parâmetros de forma e λ é o parâmetro de escala. Para $\alpha = 1$ a densidade é reduzida a CEG, para $\theta = 1$ a densidade coincide com a da exponencial exponenciada (EE) proposta por Nadarajah & Kotz (2006), e para $\alpha = \theta = 1$ temos a distribuição exponencial.

2.2 Propriedades

O p -ésimo quantil da distribuição geométrica exponencial complementar exponenciada pode ser encontrada pelo inverso da acumulada $F(Q(p)) = p$ e é dado por:

$$Q(p) = F^{-1}(p) = -\lambda^{-1} \ln \left(\frac{1 - p^{\frac{1}{\alpha}}}{\frac{p^{\frac{1}{\alpha}}}{\theta} - p^{\frac{1}{\alpha}} + 1} \right) I_{[0,1]}(p). \quad (2.6)$$

A partir de (2.6), a mediana pode ser encontrada substituindo p por 0,5, dada por:

$$Q(0,5) = F^{-1}(0,5) = -\lambda^{-1} \ln \left(\frac{1 - 0,5^{\frac{1}{\alpha}}}{\frac{0,5^{\frac{1}{\alpha}}}{\theta} - 0,5^{\frac{1}{\alpha}} + 1} \right). \quad (2.7)$$

A função de risco é escrita como:

$$h(y) = \frac{\theta^{\alpha} (1 - e^{-\lambda y})^{\alpha-1} \alpha \lambda e^{-\lambda y}}{(e^{-\lambda y} (1 - \theta) + \theta)^{\alpha+1} - \theta^{\alpha} (1 - e^{-\lambda y})^{\alpha} (e^{-\lambda y} (1 - \theta) + \theta)} I_{[0,\infty)}(y). \quad (2.8)$$

Para $y \rightarrow 0$, $\lim_{y \rightarrow 0} h(y) = 0$ e para $y \rightarrow \infty$, $\lim_{y \rightarrow \infty} h(y) = \frac{0}{0}$, conseqüentemente a função de risco não existe quando o denominador é zero, isso acontece quando $S(y) \rightarrow 0$. Apesar dessa limitação assumindo alguns valores aos parâmetros, foi verificado que a função de risco é capaz de modelar taxa de risco crescente para $\alpha \geq 0,9$, $\lambda < 1$, decrescente quando $\alpha < 1$, $\lambda < 1$ e $0,9 \leq \theta \leq 1$, e banheira para $\alpha = 0,9$, $\lambda < 0,01$. A Figura 2.1 mostra as formas de risco para diferentes valores de α , λ e θ .

Proposição 2.2.1. *Para uma variável aleatória Y com distribuição ECEG a função característica é dada por:*

$$\Phi_Y(y) = \alpha \theta^{\alpha} \Gamma\left(\frac{-iy}{\lambda} + 1\right) \sum_{k=0}^{\infty} \binom{-\alpha - 1}{k} \frac{(\theta - 1)^k \Gamma(\alpha + k)}{\Gamma(\alpha + 1 - \frac{iy}{\lambda} + k)} I_{[0,\infty)}(y), \quad (2.9)$$

em que $i = \sqrt{-1}$.

Demonstração. Fazendo:

$$\begin{aligned} \Phi_Y(y) &= \int_0^{\infty} e^{ity} f(y) dy \\ &= \int_0^{\infty} e^{yit} \alpha \left(\frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1 - \theta) + \theta} \right)^{\alpha-1} \frac{\theta \lambda e^{-\lambda y}}{(e^{-\lambda y} (1 - \theta) + \theta)^2} dy. \end{aligned}$$

Considere $z = 1 - e^{-\lambda y}$, temos que:

$$\Phi_Y(y) = \int_0^1 (1-z)^{-\frac{it}{\lambda}+1-1} z^{\alpha-1} (1+z(\theta-1))^{-\alpha-1} dz. \quad (2.10)$$

□

Comparando a integral (2.10) com (2), temos: $n = \frac{-it}{\lambda} + 1$, $b = (\theta-1)$, $m = 1$, $p = \alpha$, $l = -(\alpha + 1)$ e fazendo as substituições apropriadas termina a prova.

Proposição 2.2.2. Se Y tem distribuição ECEG, então:

$$E(Y) = -\frac{\theta^\alpha}{\lambda} \left(\frac{-H_\alpha}{\theta^\alpha} + {}_2F_1^{(0,0,1,0)}(v) \right) \quad e$$

$$V(Y) = \frac{6\theta^\alpha \left({}_2F_1^{(0,0,2,0)}(v) - \theta^\alpha {}_2F_1^{(0,0,1,0)}(v)^2 \right) - 6\psi^{(1)}(\alpha+1) + \pi^2}{6\lambda^2}, \quad (2.11)$$

em que:

$${}_2F_1^{(0,0,1,0)}(v) = {}_2F_1^{(0,0,1,0)}(\alpha, \alpha+1, \alpha+1; 1-\theta) = \sum_{k=0}^{\infty} \frac{(\alpha)_k (\alpha+1)_k}{k!}$$

$$\times \frac{\partial \frac{1}{(\alpha+1)_k}}{\partial (\alpha+1)} (1-\theta)^k,$$

$${}_2F_1^{(0,0,2,0)}(\alpha, \alpha+1, \alpha+1; 1-\theta) = \sum_{k=0}^{\infty} \frac{(\alpha)_k (\alpha+1)_k}{k!} \frac{\partial^2 \frac{1}{(\alpha+1)_k}}{\partial (\alpha+1)^2} (1-\theta)^k, \quad e$$

$${}_2F_1(\alpha, \alpha+1, \alpha+1; 1-\theta) = \frac{\sum_{k=0}^{\infty} (\alpha)_k (\alpha+1)_k}{(\alpha+1)_k} \frac{(1-\theta)^k}{k!}.$$

Demonstração. Fazendo $y = 0$ em (2.9) e considerando que $E(Y^r) = \Phi_Y(0)^{(r)}/i^r$, tem-se o primeiro momento quando a função característica é derivada uma vez e dividida por i e para encontrar o segundo momento a função característica é derivada duas vezes e dividida por -1 . Dessa forma, com o primeiro e segundo momento a variância pode ser encontrada usando a definição $V(Y) = E(Y^2) - (E(Y))^2$. □

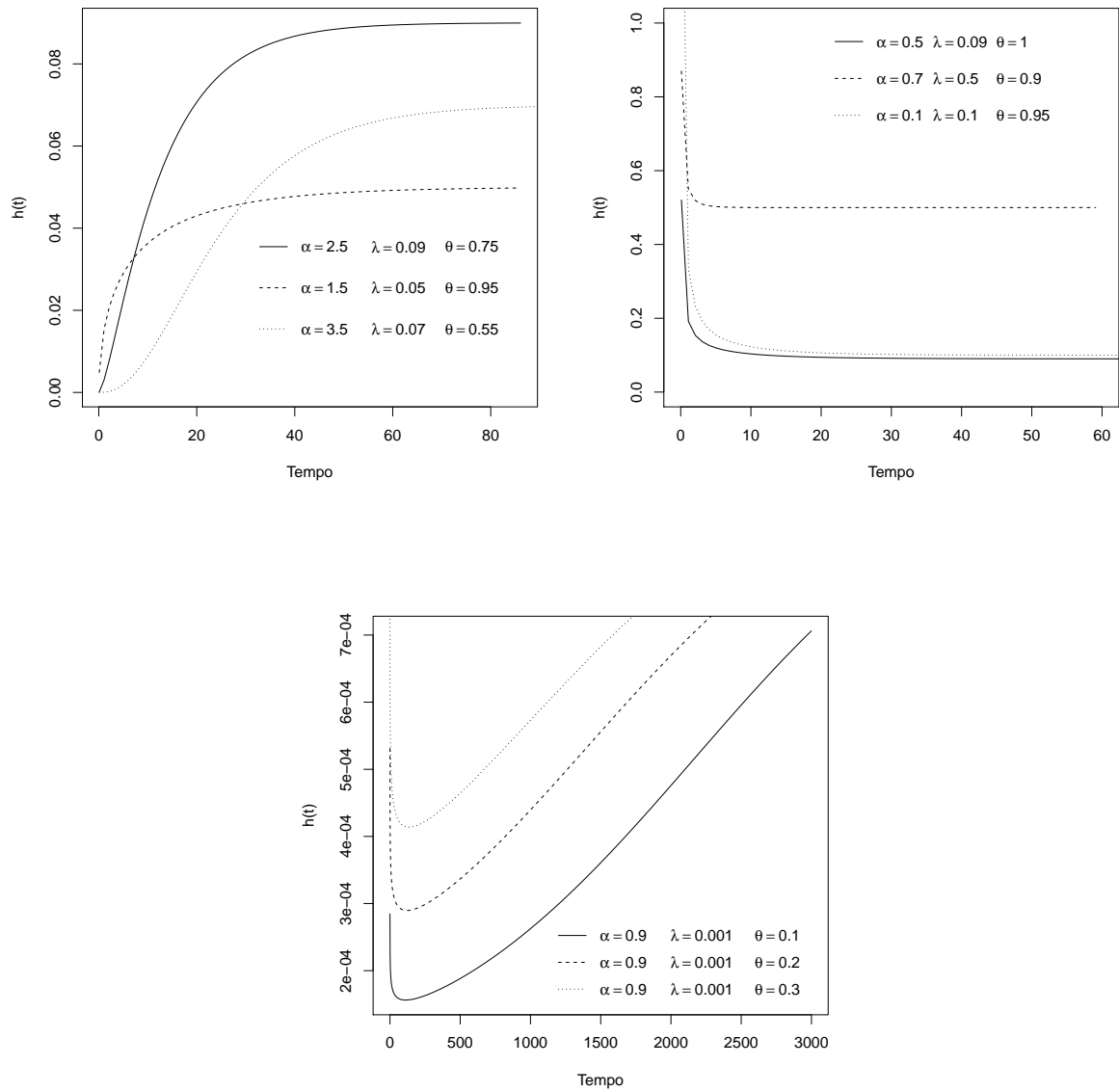


Figura 2.1: Funções de risco crescente, decrescente e em forma de banheira da distribuição ECEG

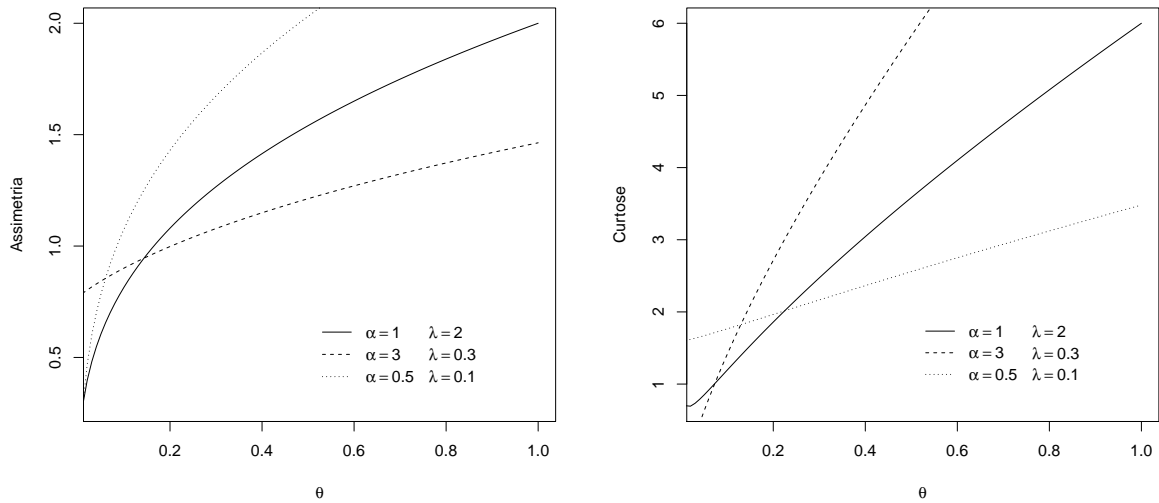


Figura 2.2: Assimetria e Curtose em distribuições da distribuição ECEG

Por meio de (2.10) os momentos da variável aleatória Y podem ser obtidos e desse modo, medidas que precisam de momentos para serem contruídos, como as medidas de curtose e assimetria são encontrados. Como pode ser visto na Figura 2.2 com o aumento do parâmetro de forma θ a assimetria da distribuição aumenta indicando que a função densidade passa a concentrar os dados no lado esquerdo.

Proposição 2.2.3. *A função densidade da k -ésima estatística de ordem é dada por:*

$$f_{k:n}(y) = \frac{\theta^{\alpha k} \alpha \lambda e^{-\lambda y}}{B(k, n-k+1)} \left(\frac{(1-e^{-\lambda y})^{\alpha k-1}}{(e^{-\lambda y}(1-\theta) + \theta)^{\alpha k+1}} \right) \times \left(1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^{\alpha} \right)^{n-k} I_{[0, \infty)}(y). \quad (2.12)$$

Demonstração. Usando a definição de função de densidade da k -ésima estatística de ordem, tem-se:

$$f_{k:n}(y) = \frac{1}{B(k, n-k+1)} f(y) (F(y))^{k-1} (S(y))^{n-k}, \quad (2.13)$$

em que $B(k, n-k+1) = \frac{(n-k)!(k-1)!}{n!}$.

Substituindo a função de densidade, a função de sobrevivência e função acumulada em (2.13), tem-se:

$$\begin{aligned}
f_{k:n}(y) &= \frac{1}{B(k, n-k+1)} \alpha \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^{\alpha-1} \left(\frac{\theta \lambda e^{-\lambda y}}{(e^{-\lambda y}(1-\theta) + \theta)^2} \right) \\
&\quad \times \left(\left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right)^{k-1} \left(1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right)^{n-k} \\
&= \frac{\theta^{\alpha k} \alpha \lambda e^{-\lambda y}}{B(k, n-k+1)} \frac{(1-e^{-\lambda y})^{\alpha k-1}}{(e^{-\lambda y}(1-\theta) + \theta)^{\alpha k+1}} \left(1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right)^{n-k} I_{[0,\infty)}(y).
\end{aligned}$$

□

Proposição 2.2.4. Para uma variável aleatória Y com distribuição ECEG, o r -ésimo momento da i -ésima estatística de ordem é dado por:

$$\begin{aligned}
E[Y_{i:n}^r] &= r! \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{p=n-i+1}^n \frac{(-1)^{p-n+i-1} (-p)_k \theta^{\alpha k} (1)_l (\alpha k)_m}{\lambda^r k! l! m! n!} \\
&\quad \times \frac{(1-\theta)^m (-(\alpha k + l + m))_n}{(n+1)^r} \binom{p-1}{n-i} \binom{n}{p}. \quad (2.14)
\end{aligned}$$

Demonstração. O r -ésimo momento da i -ésima estatística de ordem é obtido usando a definição:

$$E[Y_{i:n}^r] = r \sum_{p=n-i+1}^n (-1)^{p-n+i-1} \binom{p-1}{n-i} \binom{n}{p} \int_0^\infty y^{r-1} [S(y)]^p dy. \quad (2.15)$$

Resolvendo a integral:

$$\int_0^\infty y^{r-1} [S(y)]^p dy = \int_0^\infty y^{r-1} \left(1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right)^p dy. \quad (2.16)$$

Considere $u = 1 - e^{-\lambda y}$ e usando a equação (1) do Apêndice A, temos que:

$$\begin{aligned}
\int_0^\infty y^{r-1} [S(y)]^p dy &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} (1)_l \frac{(-p)_k \theta^{\alpha k}}{k! \lambda^r} (1-\theta)^m \int_0^1 (-\ln(1-u))^{r-1} u^{\alpha k + l + m} du \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (1)_l \frac{(-p)_k \theta^{\alpha k}}{k! \lambda^r} (1-\theta)^m \frac{(-(\alpha k + l + m))_n}{n!} \frac{(r-1)!}{(n+1)^r}. \quad (2.17)
\end{aligned}$$

Substituindo (2.17) em (2.15):

$$\begin{aligned}
E[Y_{i:n}^r] &= r! \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{p=n-i+1}^n \frac{(-1)^{p-n+i-1} (-p)_k \theta^{\alpha k} (1)_l (\alpha k)_m}{\lambda^r k! l! m! n!} \\
&\quad \times \frac{(1-\theta)^m (-(\alpha k + l + m))_n}{(n+1)^r} \binom{p-1}{n-i} \binom{n}{p}. \quad (2.18)
\end{aligned}$$

□

A distribuição de tempo de vida residual de uma variável aleatória Y , distribuída como ECEG, tem função de sobrevivência:

$$S(y) = P[Y > y + t | Y > t] \\ = I_{(-\infty, 0]} + \left(\frac{1 - \left(\frac{\theta(1 - e^{-\lambda(y+t)})}{e^{-\lambda(y+t)}(1 - \theta) + \theta} \right)^\alpha}{1 - \left(\frac{\theta(1 - e^{-\lambda t})}{e^{-\lambda t}(1 - \theta) + \theta} \right)^\alpha} \right) I_{[0, \infty)}(y).$$

O tempo de vida médio residual de uma distribuição contínua com função de sobrevivência $S(y)$ é dado por:

$$\mu(y) = E(Y - y | Y > y) = \frac{1}{S(y)} \int_y^\infty S(u) du. \quad (2.19)$$

Proposição 2.2.5. *Para uma variável aleatória Y com distribuição ECEG, a média residual é dada por:*

$$\mu(y) = \frac{\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)_k \theta^{\alpha k}}{k! \lambda} \frac{(\alpha k)_l (1)_m}{l! m!} \frac{(1 - \theta)^l}{\alpha k + l + m + 1} (1 - (1 - e^{-\lambda y})^{\alpha k + l + m + 1})}{1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha}.$$

Demonstração. Resolvendo a integral:

$$\int_y^\infty S(u) du = \int_y^\infty \left(1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right) du. \quad (2.20)$$

Considere $u = 1 - e^{-\lambda y}$ e usando (1):

$$\int_y^\infty S(u) du = \int_{1 - e^{-\lambda y}}^1 \left(1 - \left(\frac{\theta u}{(1 - u)(1 - \theta) + \theta} \right)^\alpha \right) \frac{1}{\lambda(1 - u)} du \\ = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)_k \theta^{\alpha k}}{k! \lambda} \frac{(\alpha k)_l (1)_m}{l! m!} \frac{(1 - \theta)^l}{\alpha k + l + m + 1} (1 - (1 - e^{-\lambda y})^{\alpha k + l + m + 1}). \quad (2.21)$$

Substituindo (2.21) em (2.19), a média residual é encontrada. □

2.3 Inferência

Sabendo que a função de verossimilhança é escrita como (1.4), o logaritmo da verossimilhança é dado por:

$$l\left(\alpha, \lambda, \theta; \underline{y}, \underline{\delta}\right) = \sum_{i=1}^n \delta_i \log \left(\alpha \left(\frac{\theta (1 - e^{-\lambda y_i})}{e^{-\lambda y_i} (1 - \theta) + \theta} \right)^{\alpha-1} \left(\frac{\theta \lambda e^{-\lambda y_i}}{(e^{-\lambda y_i} (1 - \theta) + \theta)^2} \right) \right) - \sum_{i=1}^n \delta_i \log \left(1 - \left(\frac{\theta (1 - e^{-\lambda y_i})}{e^{-\lambda y_i} (1 - \theta) + \theta} \right)^{\alpha} \right) + \sum_{i=1}^n \log \left(1 - \left(\frac{\theta (1 - e^{-\lambda y_i})}{e^{-\lambda y_i} (1 - \theta) + \theta} \right)^{\alpha} \right). \quad (2.22)$$

Os estimadores de máxima verossimilhança são encontrados via maximização do logaritmo da verossimilhança, em que δ_i é a variável indicadora definida em (1.1). No presente trabalho foi utilizado a função *optim* do sistema R (Team (2008)). Foram calculados os critérios *AIC*, *BIC* e o teste de Kolmogorov Smirnov para comparar os ajustes.

As EMV dos parâmetros são soluções das equações:

$$\begin{aligned} \frac{\sum_{i=1}^n (1 - \delta_i) X_i^{\alpha} \ln(X_i)}{1 - X_i^{\alpha}} &= \frac{\sum_{i=1}^n \delta_i}{\alpha} + \sum_{i=1}^n \delta_i \ln(X_i), \\ \sum_{i=1}^n (1 - \delta_i) \alpha X_i^{-1+\alpha} y_i \theta Z_i (A_i (1 - \theta) + 1) &= \frac{\sum_{i=1}^n \delta_i}{\alpha \lambda} (\alpha (1 + 2\lambda y_i Z_i - \lambda y_i \\ &\quad + (-1 + \alpha) \lambda y_i K_i (A_i (1 - \theta) + 1))), \\ \sum_{i=1}^n (1 - \delta_i) \alpha X_i^{\alpha} \frac{(1 - A_i)}{(1 - X_i^{\alpha})} &= \sum_{i=1}^n \delta_i \left(-2 \frac{Z_i}{K_i} + \frac{1}{\theta} + (-1 + \alpha) \left(-\frac{Z_i}{K_i} + \frac{1}{\theta} \right) \right), \end{aligned}$$

em que $X_i = \frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$, $Z_i = \frac{e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta)+\theta}$, $K_i = \frac{e^{-\lambda y_i}}{1-e^{-\lambda y_i}}$ e $A_i = \frac{1-e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta)+\theta}$.

A matriz de informação de Fisher é dada por:

$$I_F(\alpha, \lambda, \theta) = - \left(\begin{array}{ccc} \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \theta} \\ \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \theta} \\ \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \theta} \end{array} \right) \Bigg|_{(\alpha, \lambda, \theta) = (\hat{\alpha}, \hat{\lambda}, \hat{\theta})} \quad (2.23)$$

Os elementos de (2.23) estão no Apêndice A.

2.4 Estudo de simulação

2.4.1 Estudo das propriedades dos estimadores por M.V dos parâmetros

Um estudo de simulação foi feito para estimar a probabilidade de rejeição da hipótese $H_0: \alpha = 1$, $H_1: \alpha \neq 1$, quando se usa o teste da razão de verossimilhanças. O nível de significância fixado nos testes considerados foi de 0,05. Na hipótese nula testa-se a hipótese dos dados virem de uma distribuição CEG contra a hipótese da distribuição dos dados serem da ECEG. O vício e a raiz do erro quadrático médio das estimativas por máxima verossimilhança de cada parâmetro da distribuição ECEG foram também estimados para cada simulação assim como a média do vício e da raiz do erro quadrático médio estimada.

Para isto, foram considerados 5 cenários para os verdadeiros valores dos parâmetros, e sob cada um destes cenários foram simuladas 1000 observações independentes de T e de C de tamanhos amostrais 30, 50, 100, 500 e 1000, sendo T uma observação da distribuição ECEG, e C uma observação de uma Bernoulli(0,90), independentes entre si.

Tabela 2.1: Tabela com a taxa de rejeição

| Parâmetros | 30 | 50 | 100 | 500 | 1000 |
|---|-------|-------|-------|-------|-------|
| $\alpha = 5,000; \lambda = 3,000; \theta = 0,500$ | 0,415 | 0,658 | 0,934 | 0,999 | 0,999 |
| $\alpha = 2,500; \lambda = 9,000; \theta = 0,750$ | 0,221 | 0,425 | 0,769 | 0,999 | 0,999 |
| $\alpha = 20,000; \lambda = 3 \times 10^{-4}; \theta = 0,500$ | 0,516 | 0,742 | 0,957 | 0,999 | 0,999 |
| $\alpha = 1,000; \lambda = 3,000; \theta = 0,750$ | 0,034 | 0,023 | 0,031 | 0,041 | 0,039 |
| $\alpha = 2,500; \lambda = 9,000; \theta = 1,000$ | 0,189 | 0,434 | 0,791 | 0,999 | 0,999 |

Na Tabela 2.1 tem-se as probabilidades de rejeição estimadas para a hipótese nula dos dados virem de uma distribuição CEG. Para todos os cenários a probabilidade de rejeição cresce com o aumento do tamanho amostral. Apenas para $\alpha = 1$ a probabilidade de rejeição é baixa pois aceita-se a hipótese nula na maioria dos casos para todos os tamanhos amostrais.

Tabela 2.2: Resultados das simulações.

| n | Parâmetros | Cenários | | | | | | | | | |
|------|------------|----------|--------|--------|-------|-----------------------|-----------------------|--------|-------|--------|-------|
| | | C1 | | C2 | | C3 | | C4 | | C5 | |
| | | V. M. | EQM | V. M. | EQM | V. M. | EQM | V. M. | EQM | V. M. | EQM |
| 30 | α | -1,944 | 35,084 | -1,189 | 2,260 | -8,732 | 877,794 | -0,028 | 0,150 | 0,129 | 1,038 |
| | λ | -0,290 | 16,328 | -0,569 | 7,065 | $-2,0 \times 10^{-5}$ | $1,1 \times 10^{-8}$ | -0,403 | 1,591 | -0,855 | 8,107 |
| | θ | -0,162 | 0,189 | 0,047 | 0,131 | -0,137 | 0,188 | 0,099 | 0,136 | 0,243 | 0,170 |
| 50 | α | -1,355 | 16,632 | -0,014 | 0,935 | -5,719 | 465,264 | -0,019 | 0,073 | 0,201 | 0,589 |
| | λ | 0,025 | 0,298 | -0,107 | 3,891 | $-5,9 \times 10^{-6}$ | $3,0 \times 10^{-9}$ | -0,102 | 0,680 | -0,339 | 3,775 |
| | θ | -0,168 | 0,170 | 0,039 | 0,109 | -0,135 | 0,183 | 0,054 | 0,101 | 0,216 | 0,133 |
| 100 | α | -0,929 | 8,352 | -0,029 | 0,472 | -3,358 | 235,822 | -0,003 | 0,037 | 0,188 | 0,307 |
| | λ | 0,101 | 0,148 | 0,246 | 1,640 | $5,9 \times 10^{-6}$ | $1,1 \times 10^{-9}$ | 0,078 | 0,325 | 0,129 | 1,532 |
| | θ | -0,148 | 0,143 | -0,001 | 0,078 | -0,131 | 0,155 | 0,038 | 0,073 | 0,163 | 0,083 |
| 500 | α | -0,712 | 3,235 | -0,017 | 0,135 | -1,376 | 87,063 | -0,002 | 0,009 | 0,154 | 0,081 |
| | λ | 0,215 | 0,076 | 0,654 | 0,747 | $1,5 \times 10^{-5}$ | $4,1 \times 10^{-10}$ | 0,223 | 0,120 | 0,524 | 0,562 |
| | θ | -0,161 | 0,093 | -0,042 | 0,037 | -0,114 | 0,098 | 0,006 | 0,030 | 0,083 | 0,024 |
| 1000 | α | 0,661 | 2,258 | -0,041 | 0,084 | -4,446 | 81,725 | -0,005 | 0,004 | 0,133 | 0,044 |
| | λ | 0,231 | 0,068 | 0,73 | 0,719 | $1,8 \times 10^{-5}$ | $4,2 \times 10^{-10}$ | 0,262 | 0,103 | 0,613 | 0,507 |
| | θ | -0,155 | 0,071 | -0,062 | 0,029 | -0,221 | 0,110 | -0,009 | 0,017 | 0,057 | 0,012 |

Em que:

*V. M. : Vício Médio;

*EQM: Erro quadrático médio;

*C1: $\alpha = 5,000$; $\lambda = 3,000$; $\theta = 0,500$;

*C2: $\alpha = 2,500$; $\lambda = 9,000$; $\theta = 0,750$;

*C3: $\alpha = 20,000$; $\lambda = 3 \times 10^{-4}$; $\theta = 0,500$;

*C4: $\alpha = 1,000$; $\lambda = 3,000$; $\theta = 0,750$;

*C5: $\alpha = 2,500$; $\lambda = 9,000$; $\theta = 1,000$;

Na Tabela 2.2 apresentamos o erro quadrático médio (EQM) e o vício médio (V.M.), no qual o erro quadrático médio diminui à medida que o tamanho amostral aumenta, ou seja, as estimativas dos parâmetros melhoram. Este fato é verificado para todos os cenários.

2.5 Aplicação

Nessa seção a metodologia é usada em 3 conjuntos de dados, dois deles artificiais e um relacionado a tumor mamário em ratas.

2.5.1 Comparação dos ajustes para 3 distribuições, a dois conjuntos de dados simulados da ECEG

Dois conjuntos de dados da distribuição ECEG foram gerados, ambos de tamanho amostral 20, e com 10% de censura. Para os dados D1 foi considerado $(\alpha, \lambda, \theta) = (4, 232; 0, 701; 0, 500)$ e para os dados D2, $(\alpha, \lambda, \theta) = (0, 300; 5 \times 10^{-4}; 0, 975)$. A distribuição ECEG foi ajustada aos dados e comparada com as distribuições CEG e Exponencial

(E) cujas funções são dadas por:

$$f_{CEG}(y) = \lambda \theta (e^{-\lambda y} (1 - \theta) + \theta)^{-2} I_{[0, \infty)}(y),$$

$$f_E(y) = \frac{1}{\theta} e^{-\frac{1}{\theta} y} I_{[0, \infty)}(y).$$

Os dados D1 e D2 encontram-se na Tabela 2.3 e 2.4:

Tabela 2.3: Tempos de vida dados D1

| | | | | | | | | | | | |
|-----|-------|-------|--------|-------|--------|-------|-------|-------|-------|-------|-------|
| y | 5,900 | 6,500 | 6,800* | 6,900 | 6,900* | 7,200 | 7,200 | 7,500 | 8,200 | 8,300 | 9,100 |
| | 9,100 | 9,100 | 9,200 | 9,200 | 9,600 | 9,600 | 9,700 | 9,700 | 9,800 | | |

*observação censurada

Tabela 2.4: Tempos de vida dados D2

| | | | | | | | | |
|-----|----------|----------|----------|----------|---------|---------|---------|----------|
| y | 0,200 | 0,800 | 2,000* | 2,500 | 2,800 | 5,300 | 5,700 | 10,300 |
| | 58,300 | 77,100 | 441,700 | 441,900 | 482,100 | 589,500 | 595,100 | 1507,100 |
| | 1693,300 | 2114,000 | 2685,400 | 4127,300 | | | | |

*observação censurada

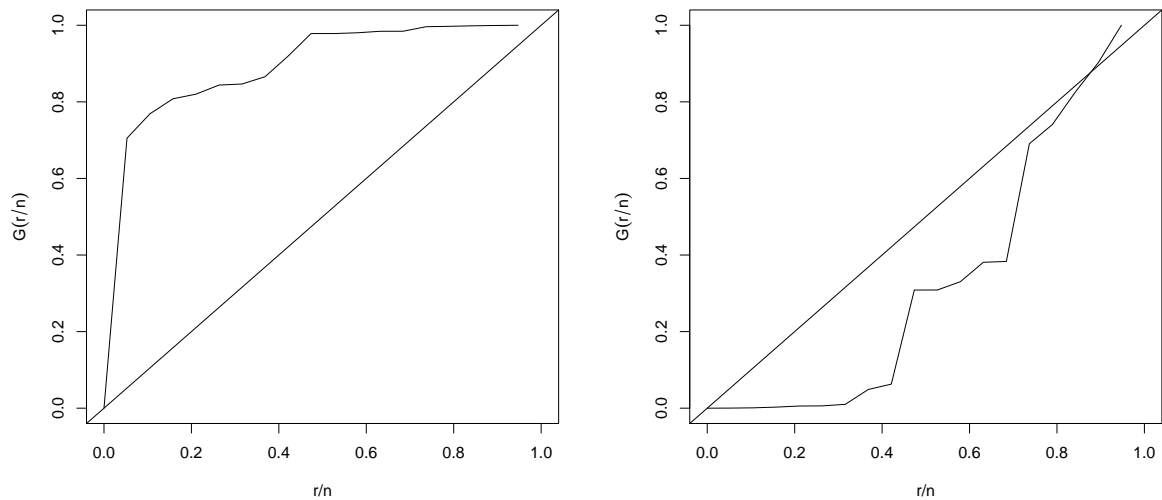


Figura 2.3: TTTplot para os dados D1 (painel esquerda) e D2 (painel direita)

Na Figura 2.3 inferimos que D1 possui uma função de risco crescente, e D2 uma função de risco decrescente.

A melhor distribuição é aquela que possuir o menor $-\log(\cdot)$, AIC , BIC e estatística de Kolmogorov Smirnov. Na Tabela 2.5 observa-se que a ECEG se ajusta melhor em relação a CEG e a E, para os dois conjunto de dados gerados. A Figura 2.4 mostra

as funções de sobrevivência das distribuições analisadas sobre a curva de Kaplan Meier. Observa-se que a distribuição ECEG se ajusta melhor que as particulares CEG e E.

Tabela 2.5: Valores dos critérios para a amostra simulada - Dados D1 e D2

| | Dados D1 | | | Dados D2 | | |
|-------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | ECEG | CEG | E | ECEG | CEG | E |
| $-\log(L(\vartheta_g))$ | 24,722 | 29,995 | 57,934 | 129,941 | 145,556 | 145,555 |
| AIC | 55,444 | 63,990 | 117,869 | 265,881 | 295,112 | 293,110 |
| BIC | 58,431 | 65,981 | 118,865 | 268,869 | 297,104 | 294,106 |
| Estatística KS | 0,144 (0,801) | 0,155 (0,721) | 0,423 (0,001) | 0,159 (0,686) | 0,376 (0,006) | 0,376 (0,006) |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p-valor do teste de kolmogorov smirnov

As estimativas dos parâmetros α , λ e θ da distribuição ECEG para D1 são dados respectivamente por 0,029; 24,943 e $4,6 \times 10^{-107}$, e para D2 por 0,269; 8×10^{-4} e 0,314. Os desvios padrões destes parâmetros para D1 são, respectivamente, 0,018; 14,595 e $6,6 \times 10^{-105}$; e 0,084; 5×10^{-4} e 0,495.

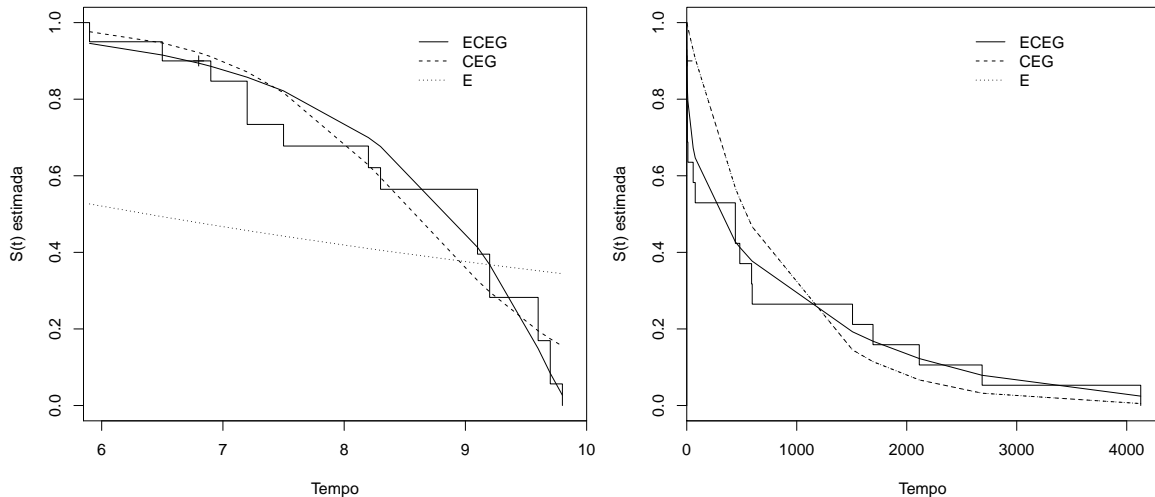


Figura 2.4: Curvas de Kaplan Meier para os dados D1 (painel esquerdo) e para os dados D2 (painel direito)

2.5.2 Exemplo com dados reais a distribuição ECEG e outras distribuições

O conjunto de dados de Lawless (2003), consiste de observações de um estudo de 9 meses sobre o efeito do carminogenese DES e DMBA na indução do tumor mamário em ratas e se encontra na Tabela 2.6.

Tabela 2.6: Tempos de vida Lawless

| | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|-----|-----|-----|
| y | 57* | 67* | 88 | 94 | 100 | 107 | 113 | 123 | 123 | 125 | 129 | 129 | 129 |
| | 136 | 143 | 144 | 191 | 191 | 192 | 211 | 218 | 266* | 266* | | | |

*observação censurada

A distribuição ECEG foi ajustada aos dados e o ajuste comparado com a CEG e a E pelos critérios $-\log(\cdot)$, AIC , BIC e estatística de Kolmogorov Smirnov.

Tabela 2.7: Estimativas dos parâmetros e critérios

| Parâmetros | ECEG | | CEG | | E | |
|-------------------------|---------|--------|---------|-------|---------|--------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 22,582 | 14,928 | - | - | - | - |
| θ | 0,992 | 0,218 | 0,011 | 0,010 | 175,894 | 40,353 |
| λ | 0,024 | 0,004 | 0,030 | 0,006 | - | - |
| $-\log(L(\vartheta_g))$ | 102,679 | | 105,979 | | 117,227 | |
| AIC | 211,358 | | 215,959 | | 236,455 | |
| BIC | 214,765 | | 218,230 | | 237,591 | |
| Estatística KS | 0,182 | | 0,196 | | 0,346 | |
| | (0,424) | | (0,337) | | (0,008) | |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p-valor do teste de kolmogorov smirnov

Pela Tabela 2.7 a distribuição ECEG se ajusta melhor em relação a CEG, E de acordo com todos os critérios considerados. As estimativas para a ECEG indicam que as ratas tinham um tempo médio de vida de 0,024, com uma proporção de ocorrência de morte em 99% dos casos, e uma grande quantidade de células afetadas pelo tumor.

A Figura 2.5 mostra o TTT plot (painel esquerdo), no qual inferimos que a função de risco é crescente, e a curva de sobrevivência mais próxima da curva Kaplan Meier é a ECEG (painel direito).

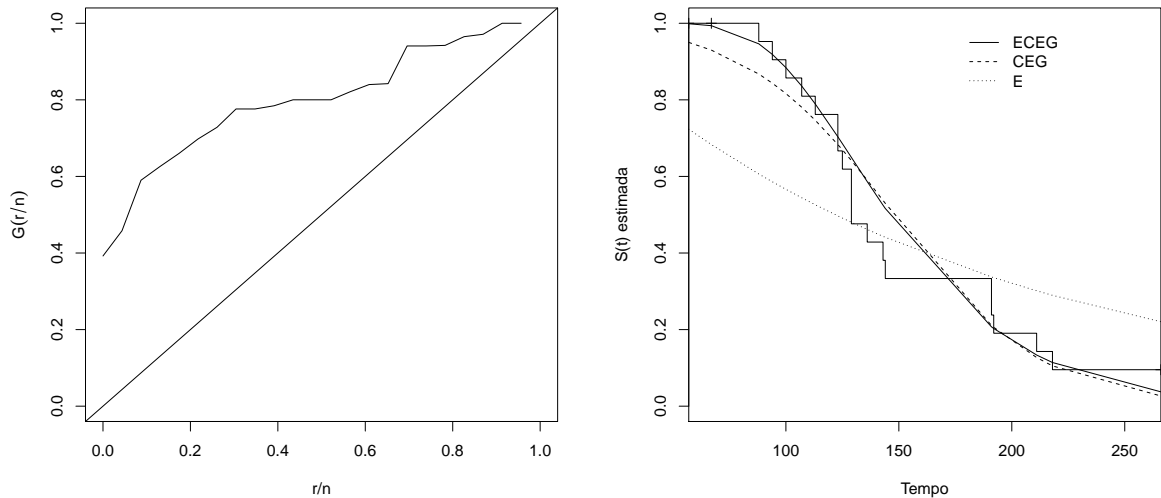


Figura 2.5: TTT plot e função de sobrevivência para dados reais

2.6 Modelo de regressão geométrico exponencial complementar exponenciado

Nesta seção, um conjunto de dados com covariáveis é ajustado ao modelo ECEG, CEG e E, os critérios de medidas $-\log(\cdot)$, AIC , BIC e gráficos dos resíduos de Cox Snell pela sobrevivência estimada são utilizados para verificar a performance da ECEG em relação as particulares.

A função de sobrevivência de um modelo com covariáveis é escrito como:

$$S(y_i, \alpha, \beta, \theta) = I_{(-\infty, 0]}(y) + \left(1 - \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i}(1 - \theta) + \theta}\right)^\alpha\right) I_{(0, \infty]}(y).$$

O logaritmo da verossimilhança é dado por:

$$\begin{aligned} l(\alpha, \beta, \theta; \underline{y}, \underline{\delta}) &= \sum_{i=1}^n \delta_i \log \left(\alpha \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i}(1 - \theta) + \theta} \right)^{\alpha-1} \left(\frac{\theta \lambda(x_i) e^{-\lambda(x_i)y_i}}{(e^{-\lambda(x_i)y_i}(1 - \theta) + \theta)^2} \right) \right) \\ &\quad - \sum_{i=1}^n \delta_i \log \left(1 - \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i}(1 - \theta) + \theta} \right)^\alpha \right) \\ &\quad + \sum_{i=1}^n \log \left(1 - \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i}(1 - \theta) + \theta} \right)^\alpha \right), \end{aligned} \quad (2.24)$$

em que $\lambda(x) = e^{x'\beta}$.

Os estimadores de máxima verossimilhança podem ser encontrados via maximização do logaritmo da verossimilhança. O *optim* da rotina do sistema R (Team (2008))

pode ser utilizado como maximização se considerarmos o $-\log(\cdot)$.

Processos automáticos de escolha de covariáveis como o *forward*, *backward* e *stepwise* são comumente utilizados em modelagem para verificar quais possíveis covariáveis influenciam na variável resposta. O processo de seleção de variáveis é o *forward*, o qual utiliza TRV no processo de seleção e consiste nos seguintes passos:

Passo 1: Ajustar o modelo com apenas o intercepto e anotar o logaritmo da verossimilhança, depois ajustar todos os possíveis modelos com uma covariável e realizar o teste de razão de verossimilhança para verificar quais ajustes foram significativos. Se tivermos mais de um modelo significativo, permanecer com aquele que possuir o menor *AIC*, *BIC* e $-\log(\cdot)$.

Passo 2: Ajustar o modelo com 2 covariáveis considerando a covariável do modelo escolhido no passo 1, e verificando aquele com os menores critérios *AIC*, *BIC* e $-\log(\cdot)$.

Passo 3: Ajustar os possíveis modelos com a adição de mais uma covariável sempre considerando as covariáveis escolhida no passo anterior até não sobrar mais nenhuma covariável, lembrando sempre de permanecer com aquele com o menor *AIC*, *BIC* e $-\log(\cdot)$.

Passo 4 e demais: Ajustar os possíveis modelos com a adição de interação entre covariáveis sempre considerando as covariáveis do modelo escolhida no passo anterior testando todas as interações possíveis, lembrando sempre de permanecer com aquele com o menor *AIC*, *BIC* e $-\log(\cdot)$.

A análise de resíduos é comumente utilizada para verificar a adequabilidade do modelo. Os resíduos de Cox Snell introduzidos por Cox & Snell (1968) são definidos por:

$$\hat{e}_i = \hat{\Lambda}(y_i|x_i), i = 1, \dots, n$$

De acordo com Lawless (2003) os resíduos devem seguir uma distribuição exponencial padrão se o modelo ajustado for adequado. O gráfico \hat{e}_i versus $\hat{\Lambda}(e_i)$ deve ser uma reta de inclinação 1 caso o modelo exponencial seja adequado. Uma outra forma de verificar a qualidade do modelo ajustado utilizada no livro de Colosimo & Giolo (2006), é através do gráfico da curva de sobrevivência dos resíduos encontradas por Kaplan-Meier e pelo modelo exponencial padrão. Quanto mais próximas as curvas estiverem melhor a adequabilidade do modelo ajustado.

Os resíduos Martingal servem para verificar a forma funcional de determinada

covariável, sendo definidas por:

$$\hat{m}_i = \delta_i - \hat{e}_i,$$

em que δ_i é o vetor indicador de falha e \hat{e}_i são os resíduos de Cox Snell, $i=1, \dots, n$.

Os resíduos *deviance* servem para verificar a acurácia do modelo ajustado para cada indivíduo em estudo e são definidos como:

$$\hat{d}_i = \text{sin}(\hat{m}_i)[-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{1/2}, i = 1, \dots, n.$$

Se os gráficos dos resíduos *deviance* ou martingal *versus* o tempo apresentarem um comportamento aleatório em torno de zero indicam a adequabilidade do modelo ajustado.

Informações sobre seleção de variáveis, assim como sobre os tipos de resíduos que podem ser utilizados para verificação da adequação de modelos podem ser encontrados em Draper & Smith (1998), Hosmer & Lemeshow (1999) e Colosimo & Giolo (2006).

Para os dados com covariáveis o teste de Kolmogorov Smirnov testa se a curva de distribuição Kaplan Meier dos resíduos do modelo ajustado está próxima da curva de distribuição da exponencial padrão.

Para os dados sem covariáveis o teste de Kolmogorov Smirnov testa se a curva de distribuição Kaplan Meier dos tempos de vida está próxima da curva de distribuição do modelo ajustado.

2.6.1 Dados Reais

Os dados são de Hosmer & Lemeshow (2008) e consiste em 1000 observações de pessoas com síndrome coronária aguda. A Tabela 2.8 apresenta as variáveis deste conjunto de dados e suas respectivas descrições.

Tabela 2.8: Descrição das variáveis dos dados de Hosmer & Lemeshow (2008)

| Variável | Descrição |
|----------|--|
| Y | Tempo de acompanhamento em dias |
| c | Se houve morte durante o acompanhamento |
| X1 | Se o paciente fez revascularização |
| X2 | Dias até a revascularização |
| X3 | Tempo total de permanência no hospital em dias |
| X4 | Idade em anos |
| X5 | Pressão sistólica do sangue em mmHg |
| X6 | Se houve desvio no segmento ST |

Ajustando o modelo ECEG aos dados e realizando o processo de seleção *forward*, as variáveis significativas a 5% foram: X1, X2, X3, X4, X5, as interações entre X1 e X4 e entre X1, X3, e X6.

O mesmo processo para a CEG e a E foram realizados.

As variáveis significativas para a CEG foram: X1, X2, X3, X4, X5, as interações entre X1 e X4 e entre X1, X3 e X6.

Para o modelo exponencial as covariáveis significativas foram: X1, X2, X3, X4, X5, X6 e as interações entre X1 e X4 e entre X1,X2 e X6.

Tabela 2.9: Valores estimados para o modelo ECEG, CEG e E

| Parâmetros | ECEG | | CEG | | E | |
|-------------------------|----------|-------|--------------------------|-------|--------------------------|-------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 0,529 | 0,043 | - | - | - | - |
| θ | 0,035 | 0,016 | 0,449 | 0,099 | - | - |
| β_0 | -0,446 | 0,371 | -1,384 | 0,472 | -2,177 | 0,532 |
| β_1 | 0,005 | 0,003 | 0,009 | 0,004 | 0,013 | 0,005 |
| β_2 | -0,029 | 0,001 | -0,029 | 0,001 | -0,031 | 0,001 |
| β_3 | -7,133 | 0,742 | -6,852 | 0,660 | -7,297 | 0,741 |
| β_4 | -0,027 | 0,007 | -0,033 | 0,009 | -0,005 | 0,001 |
| β_5 | -0,004 | 0,001 | -0,005 | 0,001 | -0,031 | 0,010 |
| β_6 | 0,042 | 0,009 | 0,040 | 0,107 | 0,149 | 0,124 |
| β_7 | 0,042 | 0,013 | 0,041 | 0,008 | 0,048 | 0,010 |
| β_8 | - | - | 0,038 | 0,013 | 0,057 | 0,022 |
| $-\log(L(\vartheta_g))$ | 1687,360 | | 1719,650 | | 1724,724 | |
| AIC | 3394,719 | | 3459,301 | | 3467,447 | |
| BIC | 3443,797 | | 3508,378 | | 3511,617 | |
| Estatística KS | 0,038 | | 0,089 | | 0,102 | |
| | (0,096) | | ($1,9 \times 10^{-7}$) | | ($1,3 \times 10^{-9}$) | |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p -valor do teste de kolmogorov smirnov

Assim, por meio da Tabela 2.9, o parâmetro λ estimado para o modelo ECEG é escrito como:

$$\lambda(\underline{x}) = e^{-0,446+0,005X4-0,029X2-7,133X1-0,027X3-0,004X5+0,042X1X4+0,042X1X3X6}.$$

O risco médio de morrer das pessoas que fizeram revascularização e tiveram desvio no segmento ST é maior quando a idade e o tempo de permanência é grande, em relação àquelas com pouca idade e com tempo de permanência menor no hospital. O aumento de uma unidade na pressão sanguínea aumenta o risco de morrer em 0,996.

A Figura 2.6 apresenta o gráfico dos resíduos de Cox Snell *versus* a sobrevivência dos resíduos de Cox Snell ajustado para o modelo ECEG, CEG e E sobreposto sobre a

curva de sobrevivência Kaplan Meier dos resíduos. Verifica-se que o ajuste pelo modelo ECEG possui uma adequabilidade melhor do que os outros por estar mais próximo da curva de sobrevivência Kaplan Meier dos resíduos, o que se confirma graficamente e pelo teste de Kolmogorov Smirnov. Este teste foi realizado na comparação da curva de Kaplan Meier dos resíduos do modelo ajustado e dos resíduos do modelo ajustado na sobrevivência da exponencial padrão, o p -valor ao nível de significância de 0,05 para os resíduos do modelo ECEG indicam que a curva de sobrevivência dos resíduos da exponencial padrão está próxima da curva de Kaplan Meier, infere-se que os resíduos Cox Snell do modelo ECEG seguem distribuição exponencial, ou seja, o modelo ECEG além de possuir baixos critérios possui uma boa adequabilidade.

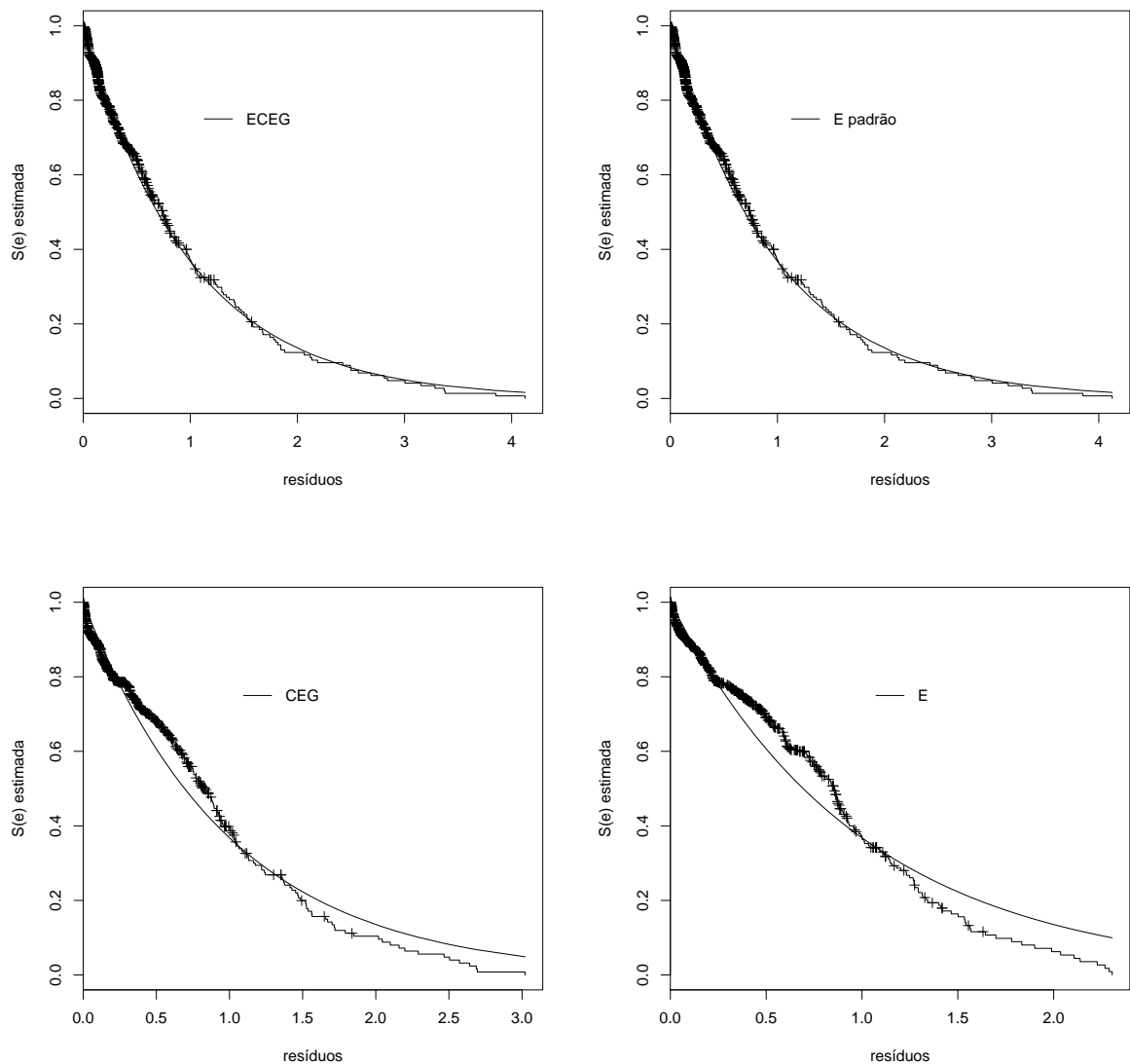


Figura 2.6: Resíduos de Cox Snell *versus* sobrevivência estimada para os modelos ECEG, E padrão, CEG e E

2.6.2 Dados simulados com covariável

Nesta seção um conjunto de dados gerados pelo método da inversa com parâmetros $\alpha = 10, 5$; $\theta = 0, 8$ e $\lambda = 5$ é analisado. A Tabela 2.10 contém os dados gerados de y e na Tabela 2.11 os dados de x .

Tabela 2.10: Tempos de vida

| | | | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|-------|--------|--------|
| y | 0,409* | 0,543* | 0,427* | 0,463 | 1,216 | 0,829 | 0,510* | 1,066 | 0,407 | 0,402 |
| | 0,769 | 0,938* | 0,488 | 0,491 | 1,347* | 0,224* | 0,471 | 0,814 | 0,454 | 0,569 |
| | 0,645* | 0,951* | 0,380* | 0,600* | 0,696 | 0,780 | 0,356 | 0,333 | 0,625* | 0,687 |
| | 0,483 | 0,720 | 0,526 | 0,464* | 1,072 | 0,554 | 0,499 | 0,612 | 0,669 | 0,411* |
| | 0,369 | 0,367 | 0,356 | 0,612 | 0,958 | 0,731 | 0,980 | 0,557 | 0,574 | 0,651 |
| | 0,488 | 0,635* | 0,581 | 0,720 | 0,753* | 0,602 | 1,358* | 1,007 | 0,577 | 0,527 |

*observação censurada

Tabela 2.11: Valores da covariável x

| | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| x | 0,627 | 0,601 | 0,353 | 0,710 | 0,731 | 0,816 | 1,086 | 0,421 | 0,473 | 0,789 |
| | 0,290 | 0,702 | 0,831 | 0,695 | 1,088 | 0,918 | 0,721 | 0,420 | 0,360 | 0,825 |
| | 0,605 | 0,718 | 0,439 | 0,672 | 0,396 | 1,150 | 1,437 | 0,590 | 1,011 | 0,523 |
| | 0,848 | 0,627 | 0,811 | 0,512 | 0,434 | 0,800 | 0,710 | 0,555 | 0,874 | 0,653 |
| | 0,757 | 0,649 | 0,881 | 0,832 | 0,527 | 0,588 | 1,518 | 0,413 | 0,858 | 0,535 |
| | 1,207 | 0,422 | 0,765 | 0,489 | 0,527 | 0,766 | 0,726 | 1,014 | 0,811 | 0,851 |

*observação censurada

Tabela 2.12: Valores estimados para o modelo ECEG, CEG e E

| Parâmetros | ECEG | | CEG | | E | |
|-------------------------|---------|-------|---------|--------|------------------------|-------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 14,123 | 5,018 | - | - | - | - |
| θ | 0,999 | 0,011 | 0,013 | 0,002 | - | - |
| β_0 | 1,540 | 0,158 | 1,864 | 0,355 | 0,100 | 0,226 |
| β_1 | -0,007 | 0,049 | 0,024 | 49,764 | -0,231 | 0,597 |
| $-\log(L(\vartheta_g))$ | 8,001 | | 13,604 | | 37,994 | |
| AIC | 24,002 | | 33,209 | | 79,988 | |
| BIC | 32,379 | | 39,492 | | 84,177 | |
| Estatística KS | 0,091 | | 0,125 | | 0,254 | |
| | (0,696) | | (0,303) | | $(8,0 \times 10^{-4})$ | |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p -valor do teste de kolmogorov smirnov

Pela Tabela 2.12 observa-se que as estimativas estão próximas dos valores gerados e que o modelo de regressão ECEG com covariável presente no parâmetro λ apresenta o melhor ajuste para os dados gerados pelo método da inversa de acordo com os critérios apresentados.

De acordo com a Figura 2.7 o modelo de regressão ECEG apresenta adequabilidade melhor em relação a CEG e E, pois possui a sobrevivência dos resíduos de Cox Snell próxima da curva de sobrevivência Kaplan Meier dos resíduos e devido ao fato dos resíduos de Cox Snell seguirem distribuição exponencial padrão. O teste de Kolmogorov Smirnov foi realizado na comparação da curva de Kaplan Meier dos resíduos do modelo ajustado e dos resíduos do modelo ajustado na sobrevivência da exponencial padrão, o p -valor alto para os resíduos do modelo ECEG indicam que a curva de sobrevivência dos resíduos da exponencial padrão está próxima da curva de Kaplan Meier, infere-se que os resíduos Cox Snell do modelo ECEG seguem distribuição exponencial, ou seja, o modelo ECEG além de possuir baixos critérios possui uma boa adequabilidade.

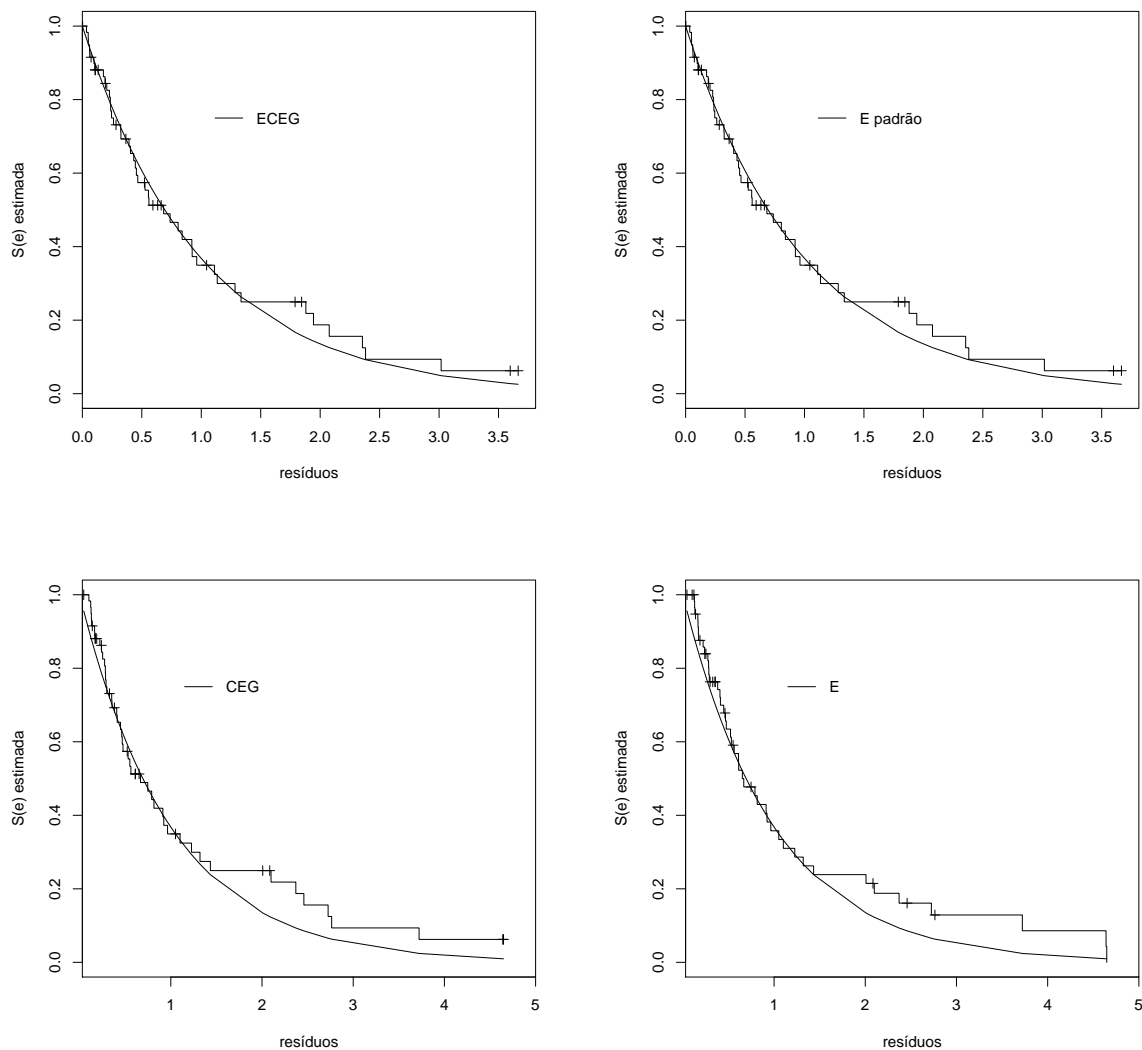


Figura 2.7: Resíduos de Cox Snell versus Sobrevivência estimada para o modelo ECEG, CEG e E

2.7 Comentários Finais

Nesse capítulo foi mostrado propriedades da ECEG, uma generalização da CEG, distribuição que apresenta diferentes taxas de risco dependendo dos valores dos parâmetros. Este modelo obteve uma melhor performance em conjunto de dados simulados e reais através dos critérios $-\log(\cdot)$, AIC , BIC e pelo teste de Kolmogorov Smirnov em comparação com os casos particulares, CEG e E. O mesmo foi verificado quando o estudo foi realizado colocando covariáveis no parâmetro λ .

Capítulo 3

Distribuição Geométrica Exponencial Complementar Exponenciada com Termo de Longa Duração

Este capítulo apresenta a distribuição LECEG e suas características, tais como: função de sobrevivência, função densidade, função de risco, função quantil, função característica, momentos, e medidas de assimetria e curtose, são determinados na Seção (3.1). Em seguida são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição LECEG na Seção (3.2). O logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição LECEG é apresentado na Seção (3.3). Na Seção (3.4) um estudo de simulação é realizado calculando o vício médio e a média do erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos de amostras da distribuição LECEG. Na Seção (3.5) a distribuição LECEG é ajustada a dados reais e simulados e, verificada sua performance através dos critérios $-\log(\cdot)$, AIC , BIC e o teste de Kolmogorov Smirnov. Por fim, na Seção (3.6) é apresentado o modelo de regressão LECEG no qual as covariáveis são incluídas no parâmetro de escala.

3.1 Origem

Em análise de sobrevivência é comum uma parte da população estar subdividida em dois grupos indicados por OR e IR: indivíduos não suscetíveis a ocorrência do evento de interesse(OR) e indivíduos que em algum momento sofrerão o evento(IR).

Em uma amostra casual de tamanho 1 desta população seja Y a variável que

represente o tempo até a ocorrência do evento no indivíduo amostrado e p a probabilidade do indivíduo pertencer ao grupo OR. A função de sobrevivência de Y , $S(y)$ pode ser escrita como uma combinação das funções de sobrevivências em cada subpopulação, indicadas por $S_{OR}(y)$ e $S_{IR}(y)$, dadas por

$$\begin{aligned} S(y) &= pP(Y > y|OR) + (1-p)P(Y > y|IR) \\ &= pS_{OR}(y) + (1-p)S_{IR}(y). \end{aligned} \quad (3.1)$$

Esta função de sobrevivência é imprópria, pois, $S_{OR}(y) = 1$ para todo y pertencente ao grupo fora de risco e portanto $\lim_{y \rightarrow \infty} S(y) = p$.

Neste capítulo estudaremos o modelo (3.1), usando como função S_{IR} a função de sobrevivência da ECEG dada em (2.4), que é reproduzida abaixo:

$$S_{IR}(y) = I_{(-\infty, 0]}(y) + \left(1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha\right) I_{[0, \infty)}(y). \quad (3.2)$$

3.2 Propriedades

No contexto da seção anterior, supondo que Y siga a distribuição LECEG, sua função de sobrevivência, de densidade de probabilidade e função de risco são dadas por:

$$\begin{aligned} S(y) &= I_{(-\infty, 0]}(y) + \left(p + (1-p) \left(1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha\right)\right) I_{[0, \infty)}(y); \\ f(y) &= -\frac{dS(y)}{dy} = \frac{(1-p)\alpha}{(1 - e^{-\lambda y})} \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha \left(\frac{\lambda e^{-\lambda y}}{e^{-\lambda y}(1 - \theta) + \theta}\right) I_{[0, \infty)}(y); \\ h(y) &= \frac{\frac{(1-p)\alpha}{(1 - e^{-\lambda y})} \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha \left(\frac{\lambda e^{-\lambda y}}{e^{-\lambda y}(1 - \theta) + \theta}\right)}{p + (1-p) \left(1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha\right)} I_{[0, \infty)}(y), \end{aligned} \quad (3.3)$$

em que λ é o parâmetro de escala, α e θ parâmetros de forma, e p o parâmetro de longa duração. A função de sobrevivência também pode ser escrita como:

$$S(y) = I_{(-\infty, 0]}(y) + \left(1 - (1-p) \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta}\right)^\alpha\right) I_{[0, \infty)}(y). \quad (3.4)$$

A Figura 3.1 mostra a *f.d.p.* da LECEG e função de sobrevivência para alguns valores do vetor $\phi = (\alpha, \lambda, \theta)$, com $p = 0, 100$ e $\alpha = 0, 800; 0, 600; 0, 200; 1, 500; 1, 700; 1, 800$, $\theta = 0, 500; 0, 700; 0, 800$ e $\lambda = 2, 000$.

A função quantil é dada por:

$$Q(u) = -\log \left(\frac{\theta \left(1 - \left(\frac{u}{1-p} \right)^{\frac{1}{\alpha}} \right)}{\left(\frac{u}{1-p} \right)^{\frac{1}{\alpha}} (1-\theta) + \theta} \right) \frac{1}{\lambda} I_{[0,1-p]}(u). \quad (3.5)$$

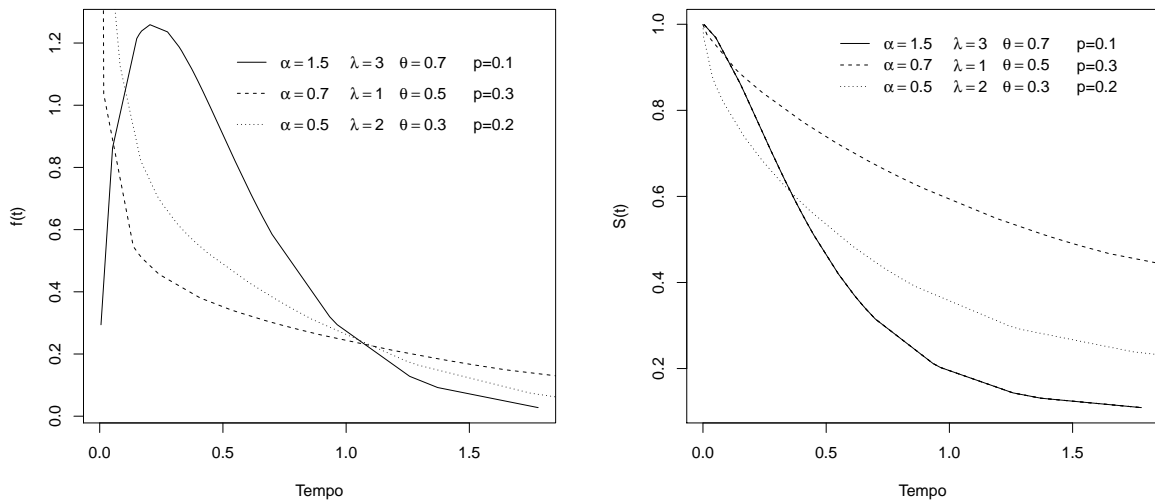


Figura 3.1: Função de densidade e de sobrevivência da distribuição LECEG

A mediana da distribuição LECEG é dada por:

$$Q(0,5) = \log \left(\frac{\theta \left(1 - \left(\frac{0.5}{1-p} \right)^{\frac{1}{\alpha}} \right)}{\left(\frac{0.5}{1-p} \right)^{\frac{1}{\alpha}} (1-\theta) + \theta} \right) \frac{1}{\lambda}. \quad (3.6)$$

A função de risco é decrescente ou unimodal, assumindo alguns valores aos parâmetros, foi observado o seguinte:

- Se $\alpha < 1$, a função de risco é decrescente;
- Se $\alpha > 1$, a função de risco é unimodal.

A Figura 3.2 corrobora com as afirmações acima com $\alpha = 0,200; 0,600; 0,800; 1,500; 1,700; 1,800$
 $\lambda = 2,000$, $p = 0,100$ e $\theta = 0,200; 0,500; 0,700; 0,800$.

Proposição 3.2.1. *A função característica de uma variável aleatória Y com distribuição LECEG é dada por:*

$$\Phi_Y(y) = (1-p) \alpha \theta^\alpha \Gamma \left(\frac{-iy}{\lambda} + 1 \right) \sum_{k=0}^{\infty} \binom{-\alpha-1}{k} \frac{(\theta-1)^k \Gamma(\alpha+k)}{\Gamma(\alpha+1 - \frac{iy}{\lambda} + k)} I_{[0,\infty)}(y). \quad (3.7)$$

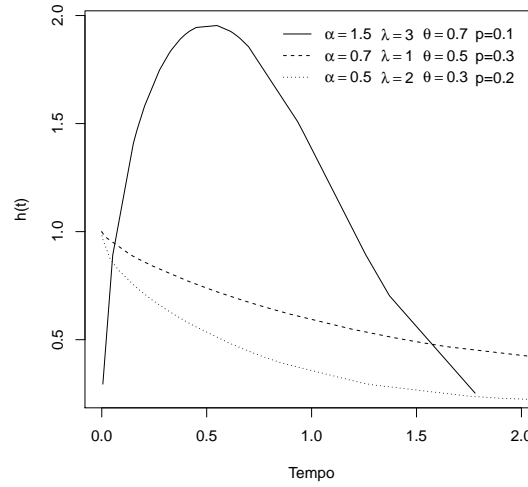


Figura 3.2: Função de risco da distribuição LECEG

Demonstração. Usando a Transformação de Fourier $\Phi_Y(y) = \int_0^\infty e^{ity} f(t) dt$, tem-se que:

$$\Phi_Y(y) = \int_0^\infty e^{ity} \frac{(1-p)\alpha}{(1-e^{-\lambda t})} \left(\frac{\theta(1-e^{-\lambda t})}{e^{-\lambda t}(1-\theta) + \theta} \right)^\alpha \left(\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}(1-\theta) + \theta} \right) dt.$$

Considera-se $u = 1 - e^{-\lambda y}$, obtendo:

$$\Phi_Y(y) = (1-p) \alpha \theta^\alpha \int_0^1 u^{\alpha-1} (1-u)^{\frac{-iy}{\lambda}} (1+u(\theta-1))^{-\alpha-1} du. \quad (3.8)$$

Resolvendo a integral (3.8) a função característica é encontrada. \square

Por meio da função característica, os momentos podem ser obtidos.

Proposição 3.2.2. *A média e a variância de uma variável aleatória Y com distribuição LECEG são dadas por:*

$$\begin{aligned} E(Y) &= \frac{(p-1) \left(\theta^\alpha {}_2F_1(0, 0, 1, 0) (\alpha, \alpha+1, \alpha+1, 1-\theta) - H_\alpha \right)}{\lambda}; \\ V(Y) &= \left(6\theta^\alpha \left(2p H_{\alpha 2} F_1(0, 0, 1, 0) (\alpha, \alpha+1, \alpha+1, 1-\theta) \right. \right. \\ &\quad \left. \left. - {}_2F_1(0, 0, 2, 0) (\alpha, \alpha+1, \alpha+1, 1-\theta) \right) \right. \\ &\quad \left. - 6(p-1) \theta^{2\alpha} {}_2F_1(0, 0, 1, 0) (\alpha, \alpha+1, \alpha+1, 1-\theta)^2 \right. \\ &\quad \left. + 6\psi^{(1)}(\alpha+1) - 6p(H_\alpha)^2 - \pi^2 \right) \frac{\alpha(p-1)\Gamma(\alpha)}{6\lambda^2\Gamma(\alpha+1)}, \end{aligned} \quad (3.9)$$

$$\text{em que } {}_2F_1(0, 0, 1, 0)(a, b, c; z) = - \sum_{k=0}^{\infty} \frac{(a)_k (b)_k \psi(c+k) z^k}{(c)_k k!} + \psi(c) {}_2F_1(a, b, c; z) e$$

$${}_2F_1(0, 0, 2, 0)(a, b, c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{k!} \frac{\partial^n \frac{1}{(c)_k}}{\partial c^2} z^k.$$

Demonstração. Fazendo $y = 0$ em (3.7) e considerando que $E(Y^r) = \Phi_Y(0)^{(r)}/i^r$, $r = 1$ e $r = 2$, temos o primeiro e segundo momento derivando a função característica e dividindo por i para encontrar o primeiro momento e por -1 para encontrar o segundo momento, a partir dos quais a variância pode ser obtida. \square

A partir dos momentos centrados na origem pode-se encontrar os momentos centrados na média e a partir deles obter medidas como a curtose e a assimetria que determinam a forma da função densidade.

Como pode ser visto na Figura 5.1, com o aumento do parâmetro de forma θ para o caso em que $p = 0, 1$, o aumento do parâmetro de forma θ faz diminuir a assimetria, ou seja, a função densidade passa a concentrar os dados no lado direito e produz uma cauda longa a esquerda da densidade. Para os outros valores de p , o decaimento é mais suave.

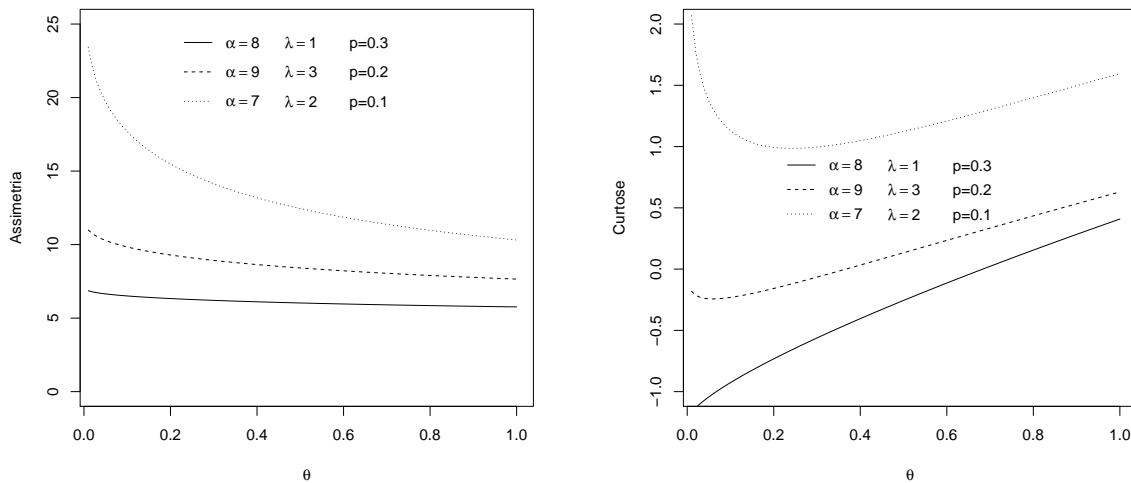


Figura 3.3: Valores da assimetria e curtose da distribuição LECGE para diversos valores

Proposição 3.2.3. Seja Y_1, Y_2, \dots, Y_n variáveis aleatórias *i.i.d* tais que $Y_k \sim LECGE(p, \lambda, \theta)$, para $k = 1, 2, \dots, n$. A *f.d.p.* da k -ésima estatística de ordem, $Y_{k:n}$, é dada por:

$$f_{k:n}(y) = \frac{(1-p)^k \alpha \theta^{\alpha k} \lambda (1 - e^{-\lambda y})^{\alpha k - 1} e^{-\lambda y}}{B(k, n-k+1) (e^{-\lambda y} (1-\theta) + \theta)^{\alpha k + 1}} \times \left(1 - (1-p) \left(\frac{\theta (1 - e^{-\lambda y})}{e^{-\lambda y} (1-\theta) + \theta} \right)^\alpha \right)^{n-k} I_{[0, \infty)}(y).$$

Demonstração. Por definição:

$$f_{k:n}(y) = \frac{1}{B(k, n-k+1)} f(y) (F(y))^{k-1} (S(y))^{n-k}, \quad (3.10)$$

em que $B(k, n - k + 1) = \frac{(n-k)!(k-1)!}{n!}$.

Usando a definição temos que:

$$\begin{aligned}
f_{k:n}(y) &= \frac{1}{B(k, n - k + 1)} f(y) (F(y))^{k-1} (S(y))^{n-k} \\
&= \frac{(1-p)\alpha}{B(k, n - k + 1)(1 - e^{-\lambda y})} \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \left(\frac{\lambda e^{-\lambda y}}{e^{-\lambda y}(1 - \theta) + \theta} \right) \\
&\quad \times \left((1-p) \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right)^{k-1} \left(1 - (1-p) \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right)^{n-k} \\
&= \frac{(1-p)^k \alpha \theta^{\alpha k} \lambda (1 - e^{-\lambda y})^{\alpha k - 1} e^{-\lambda y}}{B(k, n - k + 1) (e^{-\lambda y}(1 - \theta) + \theta)^{\alpha k + 1}} \\
&\quad \times \left(1 - (1-p) \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right)^{n-k} I_{[0, \infty)}(y).
\end{aligned}$$

□

Proposição 3.2.4. *A função característica da k -ésima estatística de ordem de uma variável Y com distribuição LECEG é dada por:*

$$\begin{aligned}
\Phi_{Y_{k:n}}(y) &= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{(-n+k)_r}{r!} \frac{(1-p)^{r+k} \theta^{\alpha r + \alpha k} \alpha}{C(k, n - k + 1)} \binom{-\alpha k - \alpha r - 1}{s} \\
&\quad \times \frac{(\theta - 1)^s \Gamma(\alpha k + \alpha r + s)}{\Gamma\left(\alpha k + \alpha r + 1 + s \frac{-iy}{\lambda}\right)} \Gamma\left(\frac{-iy}{\lambda} + 1\right) I_{[0, \infty)}(y). \tag{3.11}
\end{aligned}$$

Demonstração. Usando a Transformação de Fourier $\Phi_{Y_{k:n}}(y) = \int_0^{\infty} e^{ity} f_{k:n}(t) dt$

$$\begin{aligned}
\Phi_{Y_{k:n}}(y) &= \frac{(1-p)^k \alpha \theta^{\alpha k} \lambda}{C(k, n - k + 1)} \int_0^{\infty} \frac{e^{ity - \lambda t} \left((1 - e^{-\lambda t})^{\alpha k - 1} \right)}{(e^{-\lambda t}(1 - \theta) + \theta)^{\alpha k + 1}} \\
&\quad \times \left(1 - (1-p) \left(\frac{\theta(1 - e^{-\lambda t})}{e^{-\lambda t}(1 - \theta) + \theta} \right)^\alpha \right)^{n-k} dt \\
&= \frac{(1-p)^{k+r} \alpha \theta^{\alpha k + \alpha r}}{C(k, n - k + 1)} \sum_{r=0}^{\infty} \frac{(-n+k)_r}{r!} \\
&\quad \times \int_0^1 (1-u) \frac{-iy}{\lambda} u^{\alpha k + \alpha r - 1} (1 - u(1 - \theta))^{-\alpha k - \alpha r - 1} du \\
&= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{(-n+k)_r}{r!} \frac{(1-p)^{r+k} \theta^{\alpha r + \alpha k} \alpha}{C(k, n - k + 1)} \binom{-\alpha k - \alpha r - 1}{s} \\
&\quad \times \frac{(\theta - 1)^s \Gamma(\alpha k + \alpha r + s)}{\Gamma\left(\alpha k + \alpha r + 1 + s \frac{-iy}{\lambda}\right)} \Gamma\left(\frac{-iy}{\lambda} + 1\right) I_{[0, \infty)}(y). \tag{3.12}
\end{aligned}$$

□

Proposição 3.2.5. *A média da k -ésima estatística de ordem da LECEG é dada por:*

$$E_{k:n}(Y) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{(1-p)^{k+r} s\alpha (-1+\theta)^s \theta^{k\alpha+r\alpha} \Gamma(1+n) \Gamma(-k\alpha-r\alpha)}{\lambda \Gamma(k) \Gamma(1+r) \Gamma(1-k+n) \Gamma(1+s) \Gamma(-s-k\alpha-r\alpha) \Gamma(1+k\alpha+r\alpha)} \\ \times \frac{\Gamma(1+s+k\alpha+r\alpha) (k-n)_r \psi^0(1+k\alpha+r\alpha)}{\lambda \Gamma(k) \Gamma(1+r) \Gamma(1-k+n) \Gamma(1+s) \Gamma(-s-k\alpha-r\alpha) \Gamma(1+k\alpha+r\alpha)}.$$

Demonstração. Fazendo $y = 0$ em (3.11) e considerando que $E(Y_{k:n}^r) = \Phi_{Y_{k:n}}(0)^{(r)}/i^r$, o primeiro momento é obtido ao derivarmos a função característica uma vez e dividirmos por i . □

3.3 Inferência

Nesta seção mostra-se o logaritmo da verossimilhança utilizado para estimar os parâmetros da distribuição. Assumindo que os tempos de vida são independentes identicamente distribuídos do mecanismo de censura, os estimadores de máxima verossimilhança dos parâmetros podem ser obtidos pela maximização do logaritmo da função de verossimilhança dada por:

$$l(\alpha, \lambda, \theta, p; \underline{y}, \underline{\delta}) = \sum_{i=1}^n \delta_i \left[\log \left(\frac{(1-p)\alpha}{(1-e^{-\lambda y_i})} \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^\alpha \left(\frac{\lambda e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta) + \theta} \right) \right) \right] \\ + \sum_{i=1}^n (1-\delta_i) \log \left(1 - (1-p) \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^\alpha \right) \\ = \sum_{i=1}^n \delta_i \log((1-p)\alpha\lambda\theta^\alpha) - \lambda \sum_{i=1}^n \delta_i y_i + \sum_{i=1}^n \delta_i (\alpha-1) \log(1-e^{-\lambda y_i}) \\ - \sum_{i=1}^n \delta_i (\alpha+1) \log(e^{-\lambda y_i}(1-\theta) + \theta) \\ + \sum_{i=1}^n (1-\delta_i) \log \left(1 - (1-p) \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^\alpha \right). \quad (3.13)$$

As EMV são encontradas a partir das equações:

$$\begin{aligned}
& \sum_{i=1}^n \frac{(1-p)X_i^\alpha \ln(X_i)(1-\delta_i)}{1-(1-p)X_i^\alpha} + \sum_{i=1}^n \ln(e^{-\lambda y_i}(1-\theta) + \theta)\delta_i = \sum_{i=1}^n \ln(1 - e^{-\lambda y_i})\delta_i \\
& + \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \ln(\theta) \right); \\
& \sum_{i=1}^n \frac{\delta_i}{\lambda} + \sum_{i=1}^n K_i(-1 + \alpha)\delta_i y_i + \sum_{i=1}^n Z_i(1 + \alpha)(1 - \theta)\delta_i y_i = \sum_{i=1}^n \delta_i y_i \\
& + \frac{\sum_{i=1}^n (1-p)\alpha X_i^{-1+\alpha}(1-\delta_i)y_i \theta Z_i(A_i(1-\theta) + 1)}{1-(1-p)X_i^\alpha}; \\
& \frac{\alpha}{\theta} \sum_{i=1}^n \delta_i = \frac{\sum_{i=1}^n (1-p)\alpha X_i^{-1+\alpha} \left(-\frac{X_i^2}{\theta} + A_i \right) (1-\delta_i)}{1-(1-p)X_i^\alpha} + \sum_{i=1}^n (1+\alpha)A_i\delta_i; \\
& \frac{\sum_{i=1}^n X_i^\alpha(1-\delta_i)}{1-(1-p)X_i^\alpha} = \frac{\sum_{i=1}^n \delta_i}{1-p}. \tag{3.14}
\end{aligned}$$

A matriz de informação de Fisher é dada por:

$$I_F(\alpha, \theta, \lambda, p) = - \left(\begin{array}{cccc} I_{\alpha,\alpha} & I_{\alpha,\theta} & I_{\alpha,\lambda} & I_{\alpha,p} \\ I_{\theta,\alpha} & I_{\theta,\theta} & I_{\theta,\lambda} & I_{\theta,p} \\ I_{\lambda,\alpha} & I_{\lambda,\theta} & I_{\lambda,\lambda} & I_{\lambda,p} \\ I_{p,\alpha} & I_{p,\theta} & I_{p,\lambda} & I_{p,p} \end{array} \right) \Bigg|_{(\alpha,\theta,\lambda,p)=(\hat{\alpha},\hat{\theta},\hat{\lambda},\hat{p})}, \tag{3.15}$$

Os elementos de (3.15) se encontram no Apêndice A.

É possível encontrar as estimativas dos parâmetros usando a função *optim* do sistema R (Team (2008)).

3.4 Estudo de simulação

3.4.1 Estudo das propriedades dos estimadores por M.V. dos parâmetros

Foi conduzido um estudo de simulação para verificar a performance das estimativas dos parâmetros da distribuição estudada. Foram gerados tempos de vida da LECEG que tem densidade $f(y)$ de parâmetros $\alpha = 1, 3$, $\lambda = 1, 1$ e $\theta = 0, 9$ e $p = 0, 1$, pelo algoritmo de rejeição. Este algoritmo considera uma função auxiliar $g(y)$, satisfazendo $g(y) \geq f(y)$, se $c = \int_{-\infty}^{\infty} g(y)dy$. Então $h(y)$ definida por $h(y) = g(y)/c$ é uma função densidade de probabilidade.

Passos do algoritmo:

- 1) Gere uma variável aleatória Y com densidade $h(y)$;
- 2) Gere um número aleatório U com distribuição Uniforme(0,1) independente de Y ;
- 3) Aceite o valor Y se $U \leq \frac{f(y)}{g(y)}$;
- 4) Volte ao passo 1 até conseguir os elementos da amostra desejados;

Nesta simulação escolheu-se $h(y)$ como a função densidade da distribuição de Weibull de parâmetro de escala e de forma, respectivamente, iguais a 1,2 e 1,1.

Pode-se aceitar o valor U como sendo uma amostra da distribuição LECEG quando:

$$U \leq (1,287(0,9(1 - e^{-1,1y}))^{1,3}(e^{-1,1y}0,1 + 0,9)^{-2,3} e^{-1,1y}(1 - e^{-1,1y})^{-1}/(1,1/(1,2)y^{0,1}/(1,2)^{0,1}e^{-(y/1,2)^{1,1}})).$$

Nas simulações foi considerada uma proporção de cura $p = 0,10$. Os indicadores de censura foram gerados da distribuição Bernoulli com probabilidade de sucesso iguais a 0,80; 0,70 e 0,60. Foram simuladas 1000 amostras censuradas de cada tamanho n , para $n = 30, 50, 100, 200$ e 500 da distribuição LECEG e foram calculados o vício médio e a média da raiz quadrada do erro quadrático médio para cada amostra.

Tabela 3.1: Resultados das simulações.

| | | Cenário | | | | | | | |
|-----|------------|----------------|-------|-----------------|-------|----------------|-------|-----------|--|
| | | $\alpha = 1,3$ | | $\lambda = 1,1$ | | $\theta = 0,9$ | | $p = 0,1$ | |
| | | 20% | | 30% | | 40% | | | |
| n | Parâmetros | V. M. | EQM | V. M. | EQM | V. M. | EQM | | |
| 30 | α | -0,008 | 0,252 | -0,013 | 0,297 | -0,037 | 0,458 | | |
| | λ | -0,091 | 0,192 | -0,061 | 0,254 | -0,100 | 1,461 | | |
| | θ | 0,201 | 0,162 | 0,229 | 0,184 | 0,256 | 0,209 | | |
| | p | 0,090 | 0,008 | 0,076 | 0,008 | 0,054 | 0,146 | | |
| 50 | α | 0,023 | 0,138 | 0,018 | 0,165 | 0,021 | 0,181 | | |
| | λ | 0,005 | 0,099 | 0,077 | 0,120 | 0,128 | 0,170 | | |
| | θ | 0,170 | 0,127 | 0,185 | 0,141 | 0,210 | 0,162 | | |
| | p | 0,093 | 0,009 | 0,086 | 0,008 | 0,073 | 0,123 | | |
| 100 | α | 0,039 | 0,066 | 0,047 | 0,071 | 0,054 | 0,081 | | |
| | λ | 0,087 | 0,045 | 0,172 | 0,073 | 0,238 | 0,127 | | |
| | θ | 0,127 | 0,087 | 0,140 | 0,095 | 0,164 | 0,116 | | |
| | p | 0,096 | 0,009 | 0,091 | 0,008 | 0,081 | 0,096 | | |
| 200 | α | 0,037 | 0,032 | 0,044 | 0,035 | 0,052 | 0,039 | | |
| | λ | 0,141 | 0,037 | 0,230 | 0,072 | 0,317 | 0,123 | | |
| | θ | 0,078 | 0,051 | 0,091 | 0,059 | 0,106 | 0,068 | | |
| | p | 0,098 | 0,009 | 0,095 | 0,009 | 0,088 | 0,065 | | |
| 500 | α | 0,025 | 0,014 | 0,030 | 0,015 | 0,035 | 0,016 | | |
| | λ | 0,173 | 0,036 | 0,268 | 0,078 | 0,359 | 0,137 | | |
| | θ | 0,037 | 0,024 | 0,039 | 0,028 | 0,049 | 0,033 | | |
| | p | 0,098 | 0,009 | 0,099 | 0,009 | 0,093 | 0,039 | | |

*V. M.: Vício Médio

*EQM: Erro quadrático médio

Pela Tabela 3.1, observa-se que o erro quadrático médio das estimativas de α , λ ,

θ e p diminuem a medida que n cresce.

3.5 Estudo de simulação dos ajustes para quatro distribuições, a um conjunto de dados simulados da LECEG

Uma amostra censurada de tamanho 30 da distribuição LECEG foi gerada pelo método da rejeição, e com probabilidade de 40% de censura. Os parâmetros escolhidos da LECEG foram $(\alpha, \lambda, \theta, p) = (0,500; 2,900; 0,600; 0,100)$ que corresponde a uma LECEG com função de risco decrescente. Na Tabela 3.2 são apresentados os valores da amostra obtidos.

Tabela 3.2: Dados gerados da distribuição LECEG

| | | | | | | | | | |
|-----|--------|-------|--------|-------|-------|--------|--------|--------|-------|
| y | 0,005* | 0,009 | 0,009 | 0,016 | 0,019 | 0,019 | 0,024* | 0,034 | 0,037 |
| | 0,041* | 0,053 | 0,086 | 0,097 | 0,130 | 0,132 | 0,139* | 0,140 | 0,168 |
| | 0,266 | 0,325 | 0,357* | 0,380 | 0,542 | 0,553* | 0,567* | 0,608* | 0,659 |
| | 1,170* | 1,812 | 2,005 | | | | | | |

*observação censurada

Abaixo funções de densidade da Exponencial, Lindley, CEG, EGU considerando que na amostra existem pessoas imunes ao evento, ou seja, funções de densidade considerando o parâmetro p .

$$f_{LE}(y) = p + \lambda e^{-\lambda y} (1 - p) I_{[0, \infty)},$$

$$f_{LLI}(y) = p + (1 - p) (y + 1) \frac{\lambda^2 e^{-\lambda y}}{\lambda + 1} I_{[0, \infty)},$$

$$f_{LCEG}(y) = p + (1 - p) \left(\frac{\theta \lambda e^{-\lambda y}}{(e^{-\lambda y} (1 - \theta) + \theta)^2} \right) I_{[0, \infty)},$$

$$f_{LEGU}(y) = p + (1 - p) \alpha \left(1 - e^{-e^{-\left(\frac{y - \mu}{\sigma}\right)}} \right)^{\alpha - 1} \frac{e^{-\left(\frac{y - \mu}{\sigma}\right)} - \left(\frac{y - \mu}{\sigma}\right)}{\sigma} I_{[0, \infty)}.$$

A Tabela 3.3 possui as estimativas dos parâmetros para as distribuições LECEG, LCEG, LE, LEGU e LLI. Para a LECEG as estimativas indicam que o tempo médio de vida das pessoas é de 1,17, com uma proporção de acontecimento do evento de 0,99, uma baixa proporção de curados e o comportamento dos tempos de vida caracterizam

Tabela 3.3: Estimativas dos parâmetros aos modelos LECEG, LCEG, LE, LEGU e LLI

| Parâmetros | LECEG | | LCEG | | LE | | LEGU | | LLI | |
|-------------------------------|----------------------|--------------------|----------------------|-------|----------------------|-------|----------------------|----------------------|--------------------|-------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 0,528 | 0,123 | - | - | - | - | 0,121 | 0,025 | - | - |
| λ | 1,171 | 0,448 | 2,112 | 0,450 | 2,112 | 0,450 | - | - | 2,725 | 0,487 |
| θ | 0,999 | 0,005 | 0,999 | 0,002 | - | - | - | - | - | - |
| σ | - | - | - | - | - | - | 0,059 | 4×10^{-4} | - | - |
| μ | - | - | - | - | - | - | -0,016 | 0,007 | - | - |
| p | $4,7 \times 10^{-6}$ | 8×10^{-4} | $6,6 \times 10^{-5}$ | 0,005 | $6,1 \times 10^{-5}$ | 0,005 | $1,6 \times 10^{-9}$ | $8,9 \times 10^{-6}$ | 7×10^{-4} | 0,022 |
| $-\log(L(\hat{\vartheta}_g))$ | 1,289 | | 5,551 | | 5,550 | | 9,717 | | 6,329 | |
| AIC | 10,578 | | 17,102 | | 15,101 | | 27,435 | | 16,659 | |
| BIC | 16,183 | | 21,305 | | 17,904 | | 33,040 | | 19,462 | |
| Estatística KS | 0,127 | | 0,235 | | 0,235 | | 0,225 | | 0,243 | |
| | (0,711) | | (0,072) | | (0,072) | | (0,095) | | (0,056) | |

* $\hat{\vartheta}_g$ indica os parâmetros de uma distribuição
e em parênteses o p -valor do teste de kolmogorov smirnov

uma função de risco decrescente pois a estimativa de α foi menor do que 1. A distribuição LECEG é a melhor pelos critérios $-\log(\cdot)$, AIC, BIC e a estatística de Kolmogorov Smirnov, em relação as particulares LCEG, LE e as usuais LEGU e LLI. O teste de Kolmogorov Smirnov foi realizado entre a curva de sobrevivência dos tempos de vida de Kaplan Meier e cada função de sobrevivência estimada. Um alto p -valor indica que as curva dos tempos de vida Kaplan Meier e de sobrevivência ajustada estão muito próximos. Neste caso a LECEG é a mais adequada por possuir baixos critérios e alto p -valor.

A Figura 3.4 mostra o TTT plot (painel esquerdo) indicando que os tempos de vida se comportam de uma forma decrescente e as curvas ajustadas de cada modelo sobrepostas a curva de Kaplan Meier (painel direito) evidenciam a LECEG como a melhor ajustada.

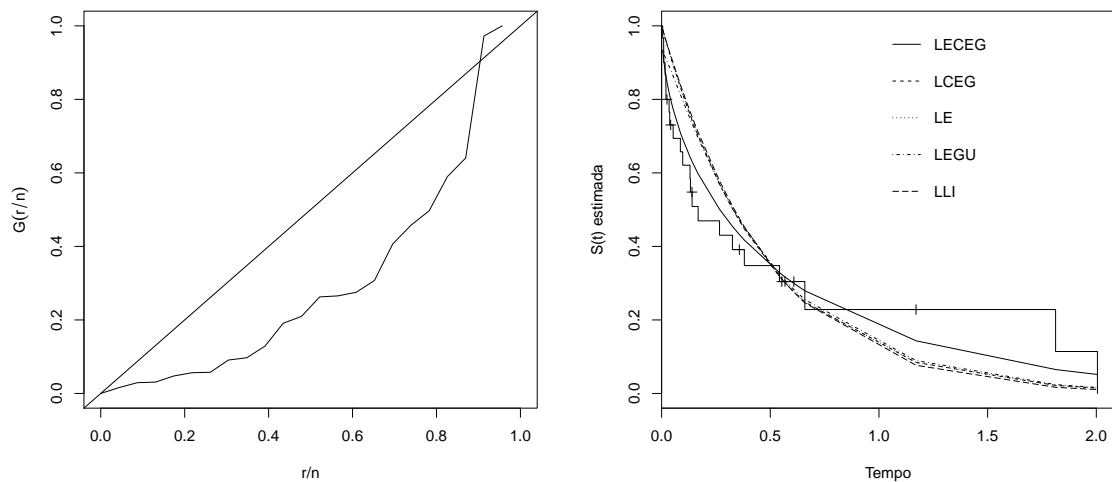


Figura 3.4: TTT plot (painel esquerdo) e Curva Kaplan Meier com as funções de sobrevivências estimadas de LECEG, LCEG, LE, LEGU e LLI (painel direito).

3.5.1 Ajuste a conjunto de dados reais da distribuição LECEG e outras distribuições

Nesta seção, comparamos a distribuição LECEG com os casos particulares (LCEG e LE), e também com a exponenciada da Gumbel com termo de longa duração (LEGU) e a Lindley com o termo de longa duração (LLI) em dois conjunto de dados.

A idéia é mostrar a aplicabilidade da distribuição proposta no ajuste de dados de sobrevivência. Os dados de Prentice (1973) e Allison (1995) encontram-se respectivamente, nas Tabelas 3.5 e 3.4.

O primeiro conjunto de dados de Allison (Allison (1995)) consiste em 25 tempos de vida com aproximadamente 32% de censura.

O segundo conjunto de dados de Prentice (Prentice (1973)) consiste em 40 tempos de vida de pacientes sob tratamento com 7,5% de censura.

Tabela 3.4: Conjunto de dados de Allison

| | | | | | | | | | | | | | |
|-----|-------|-------|-----|------|-------|-------|-----|----|-------|------|------|----|----|
| y | 8 | 180 | 632 | 852* | 52 | 2240* | 220 | 63 | 195 | 76 | 70 | 8 | 13 |
| | 1990* | 1976* | 18 | 700 | 1296* | 1460* | 210 | 63 | 1328* | 1296 | 365* | 23 | |

*observação censurada

Tabela 3.5: Conjunto de dados de Prentice

| | | | | | | | | | | | | | | |
|-----|------|-----|-----|----|-----|-----|-----|-----|------|-----|-----|-----|-----|----|
| y | 411 | 126 | 118 | 82 | 8 | 25* | 11 | 54 | 153 | 16 | 56 | 21 | 287 | 10 |
| | 8 | 12 | 177 | 12 | 200 | 250 | 100 | 999 | 231* | 991 | 1 | 201 | 44 | 15 |
| | 103* | 2 | 20 | 51 | 18 | 90 | 84 | 164 | 19 | 43 | 340 | 231 | | |

*observação censurada

A Figura 3.5 mostra os TTT plots para os conjunto de dados considerados, implicando em função de risco decrescente para os ambos casos.

Tabela 3.6: Estimativas dos parâmetros dos conjuntos de dados de Allison aos modelos LECEG, LCEG, LE, LEGU e LLI

| Parâmetros | Dados Allison | | | | | | | | | |
|-------------------------|--------------------|--------------------|------------------|--------------------|------------------|--------------------|----------------------|--------------------|------------------|--------------------|
| | LECEG | | LCEG | | LE | | LEGU | | LLI | |
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 0,4776 | 0,193 | - | - | - | - | 0,0647 | 0,015 | - | - |
| λ | 6×10^{-4} | 7×10^{-4} | 0,001 | 5×10^{-4} | 0,0011 | 5×10^{-4} | - | - | 0,003 | 6×10^{-4} |
| θ | 0,691 | 1,550 | 0,999 | 0,010 | 0,999 | 0,010 | - | - | - | - |
| σ | - | - | - | - | - | - | 62,470 | 0,154 | - | - |
| μ | - | - | - | - | - | - | -30,375 | 30,504 | - | - |
| p | 5×10^{-4} | 0,011 | 0,023 | 0,168 | 0,024 | 0,165 | $8,7 \times 10^{-7}$ | 2×10^{-4} | 0,134 | 0,096 |
| $-\log(L(\vartheta_g))$ | 128,501 | | 132,669 | | 132,669 | | 134,756 | | 148,036 | |
| AIC | 265,003 | | 271,339 | | 269,338 | | 277,512 | | 300,072 | |
| BIC | 269,878 | | 274,996 | | 271,775 | | 282,387 | | 302,510 | |
| Estatística KS | 0,148 (0,641) | | 0,258 (0,070) | | 0,258 (0,071) | | 0,249 (0,088) | | 0,340 (0,006) | |

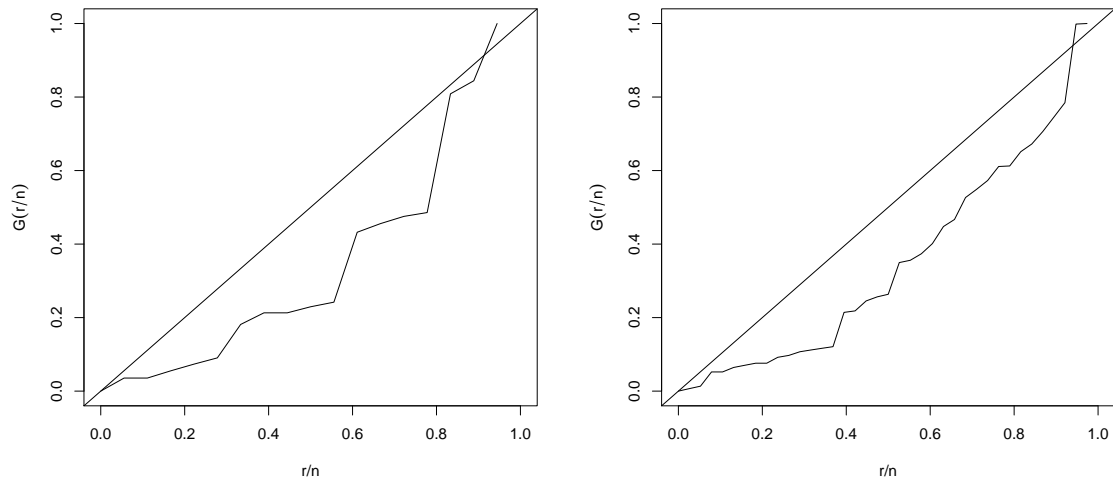


Figura 3.5: TTT plots de Allison e Prentice 1973.

Tabela 3.7: Estimativas dos parâmetros dos conjuntos de dados de Prentice aos modelos LECEG, LCEG, LE, LEGU e LLI

| Dados Prentice | | | | | | | | | | |
|-------------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-------|--------------------|--------|---------------------------------|--------------------|
| Parâmetros | LECEG | | LCEG | | LE | | LEGU | | LLI | |
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 0,630 | 0,124 | - | - | - | - | 0,172 | 0,030 | - | - |
| λ | 0,004 | 0,001 | 0,006 | 0,001 | 0,0063 | 0,001 | - | - | 0,012 | 0,001 |
| θ | 0,999 | 0,011 | 0,999 | 0,003 | - | - | - | - | - | - |
| σ | - | - | - | - | - | - | 27,177 | 0,071 | - | - |
| μ | - | - | - | - | - | - | 2,812 | 10,515 | - | - |
| p | $1,7 \times 10^{-5}$ | 7×10^{-4} | $1,1 \times 10^{-5}$ | 5×10^{-4} | $4,5 \times 10^{-5}$ | 0,001 | 1×10^{-4} | 0,001 | $2,3 \times 10^{-5}$ | 8×10^{-4} |
| $-\log(L(\vartheta_g))$ | 220,919 | | 223,922 | | 223,923 | | 229,550 | | 244,441 | |
| AIC | 449,838 | | 453,844 | | 451,846 | | 467,101 | | 492,883 | |
| BIC | 456,594 | | 458,911 | | 455,223 | | 473,857 | | 496,261 | |
| Estatística KS | 0,111 (0,700) | | 0,208 (0,061) | | 0,208 (0,060) | | 0,194 (0,097) | | 0,316 (6×10^{-4}) | |

As Tabelas 3.6 e 3.7 apresentam, respectivamente, as estimativas dos parâmetros para LECEG, LCEG, LE, LEGU e LLI para o conjunto de dados Allison e Prentice. As estimativas para α evidenciam uma função de risco decrescente para os tempos de vida, para ambos os casos temos baixo tempo de vida médio das pessoas e poucos casos de pessoas imunes ao evento. Os critérios $-\log(\cdot)$, AIC, BIC e o teste de Kolmogorov Smirnov indicam que a distribuição LECEG se ajusta melhor em relação as outras.

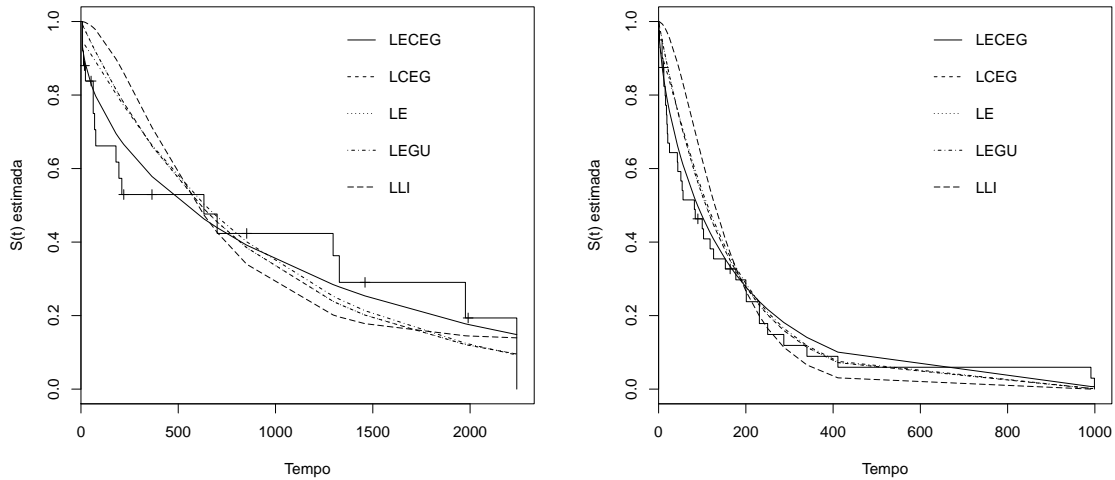


Figura 3.6: Curva Kaplan Meier com as funções de sobrevivências estimadas de LECEG, LCEG, LE, LEGU e LLI para o conjunto de dados de Allison (painel esquerdo) e Prentice (painel direito).

Esses resultados estão corroborados na Figura 3.6, que mostra as funções de sobrevivências sobrepostas na função de sobrevivência de Kaplan Meier, sendo a LECEG a mais próxima.

3.6 Modelo de regressão geométrico exponencial exponenciado de longa duração

Para o caso em que a população em estudo não for homogênea quanto a um conjunto de covariáveis que podem influenciar o tempo de sobrevivência de cada indivíduo, o modelo LECEG pode ser generalizado admitindo que a influencia das covariáveis se dê sobre o parâmetro de escala. A função de sobrevivência de uma variável aleatória Y com distribuição LECEG dado o vetor de valores de p covariáveis x pode ser modelada como (3.16):

$$S(y_i, \beta, \theta, \alpha, x_i) = 1 - (1 - p) \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i}(1 - \theta) + \theta} \right)^\alpha, \quad (3.16)$$

em que $\lambda(x) = e^{x'\beta}$.

A partir da verossimilhança dada em (1.4), o logaritmo da verossimilhança é escrito

como:

$$\begin{aligned}
 l(\alpha, \theta, \beta, p; \underline{y}, \underline{\delta}) &= \sum_{i=1}^n \delta_i \log((1-p) \alpha \theta^\alpha \lambda(x_i)) + (\alpha+1) \sum_{i=1}^n \delta_i \log(1 - e^{-\lambda(x_i)y}) \\
 &- \lambda(x_i) \sum_{i=1}^n \delta_i y_i + (-\alpha-1) \sum_{i=1}^n \delta_i \log(e^{-\lambda(x_i)y} (1-\theta) + \theta) \\
 &+ \sum_{i=1}^n (1-\delta_i) \log \left(1 - (1-p) \left(\frac{\theta(1 - e^{-\lambda(x_i)y_i})}{e^{-\lambda(x_i)y_i} (1-\theta) + \theta} \right)^\alpha \right). \quad (3.17)
 \end{aligned}$$

em que $\lambda(\underline{x}) = e^{\underline{x}'\beta}$.

3.6.1 Dados Reais

Os dados, extraídos são do livro Kalbfleisch & Prentice (1980), podem ser encontrados no site

<http://lib.stat.cmu.edu/datasets/stanford>. São ajustados os modelo LECEG, LCEG e LE para este conjunto de dados e calculadas as medidas *AIC* e *BIC*.

A Tabela 3.8 apresenta as variáveis deste conjunto de dados e suas respectivas descrições.

Tabela 3.8: Variáveis do livro Kalbfleisch & Prentice (1980)

| Variável | Descrição |
|----------|---|
| Y | Tempo de sobrevivência |
| c | Estado(Vivo/Morto) |
| X1 | Anos de aceitação |
| X2 | Idade |
| X3 | Se o paciente fez transplante |
| X4 | Se o paciente passou por algum tipo de cirurgia antes |

Por meio da Tabela 3.9 para o modelo LECEG a estimativa de λ é dada por $\hat{\lambda}(\underline{x}) = e^{8,71-32,33*X3-1,06*X4-0,17*X1+0,44*X1*X3}$. Assim, para o modelo LECEG, o risco médio de morrer das pessoas que não realizaram cirurgia é 2,89 vezes maior com relação aqueles que já realizaram cirurgia. As pessoas que realizaram transplante tem um risco médio de morrer maior quando o tempo de aceitação é grande em comparação com um tempo de aceitação pequeno.

Tabela 3.9: Valores estimados para o modelo LECEG, LCEG e LE

| Parâmetros | LECEG | | LCEG | | LE | |
|-------------------------|----------------------|--------------------|-----------------------------------|-----------------------|------------------|-------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 0,828 | 0,143 | - | - | - | - |
| θ | 0,697 | 0,231 | 1,000 | $1,2 \times 10^{-10}$ | - | - |
| p | $1,7 \times 10^{-5}$ | 3×10^{-6} | 0,080 | 0,047 | 0,076 | 0,054 |
| β_0 | 8,710 | 7,385 | -5,820 | 0,784 | 13,183 | 6,492 |
| β_1 | -32,339 | 9,147 | -2,283 | 0,298 | -25,988 | 9,826 |
| β_2 | -1,067 | 0,359 | 0,046 | 0,016 | 0,054 | 0,015 |
| β_3 | -0,178 | 0,105 | -63,942 | 30,293 | -0,275 | 0,093 |
| β_4 | 0,441 | 0,130 | 0,879 | 0,421 | -1,135 | 0,483 |
| β_5 | - | - | - | - | 0,336 | 0,140 |
| $-\log(L(\vartheta_g))$ | 437,613 | | 474,210 | | 471,892 | |
| AIC | 891,226 | | 962,421 | | 957,785 | |
| BIC | 912,040 | | 980,864 | | 976,228 | |
| Estatística KS | 0,092 (0,365) | | 0,239 ($1,9 \times 10^{-5}$) | | 0,149 (0,019) | |

* ϑ_g indica os parâmetros de uma distribuição

A Figura 3.7 possui os resíduos de Cox Snell *versus* a sobrevivência dos resíduos de Cox Snell ajustada sobrepostas a curva de Kaplan Meier dos resíduos do modelo ajustado. Verifica-se que a distribuição LECEG possui uma adequabilidade melhor em relação a LCEG e LE, pois está mais próxima da curva de sobrevivência dos resíduos Kaplan Meier. O teste de Kolmogorov Smirnov comparou a curva de sobrevivência dos resíduos do modelo ajustado e a curva de sobrevivência da exponencial padrão dos resíduos do modelo ajustado. Entre os modelos ajustados apenas os resíduos da LECEG seguem distribuição exponencial devido a função de sobrevivência da exponencial padrão estar próxima da curva de Kaplan Meier indicada pelo p -valor de 0,365 que ao nível de significância de 0,05 indica a LECEG com os resíduos seguindo distribuição exponencial.

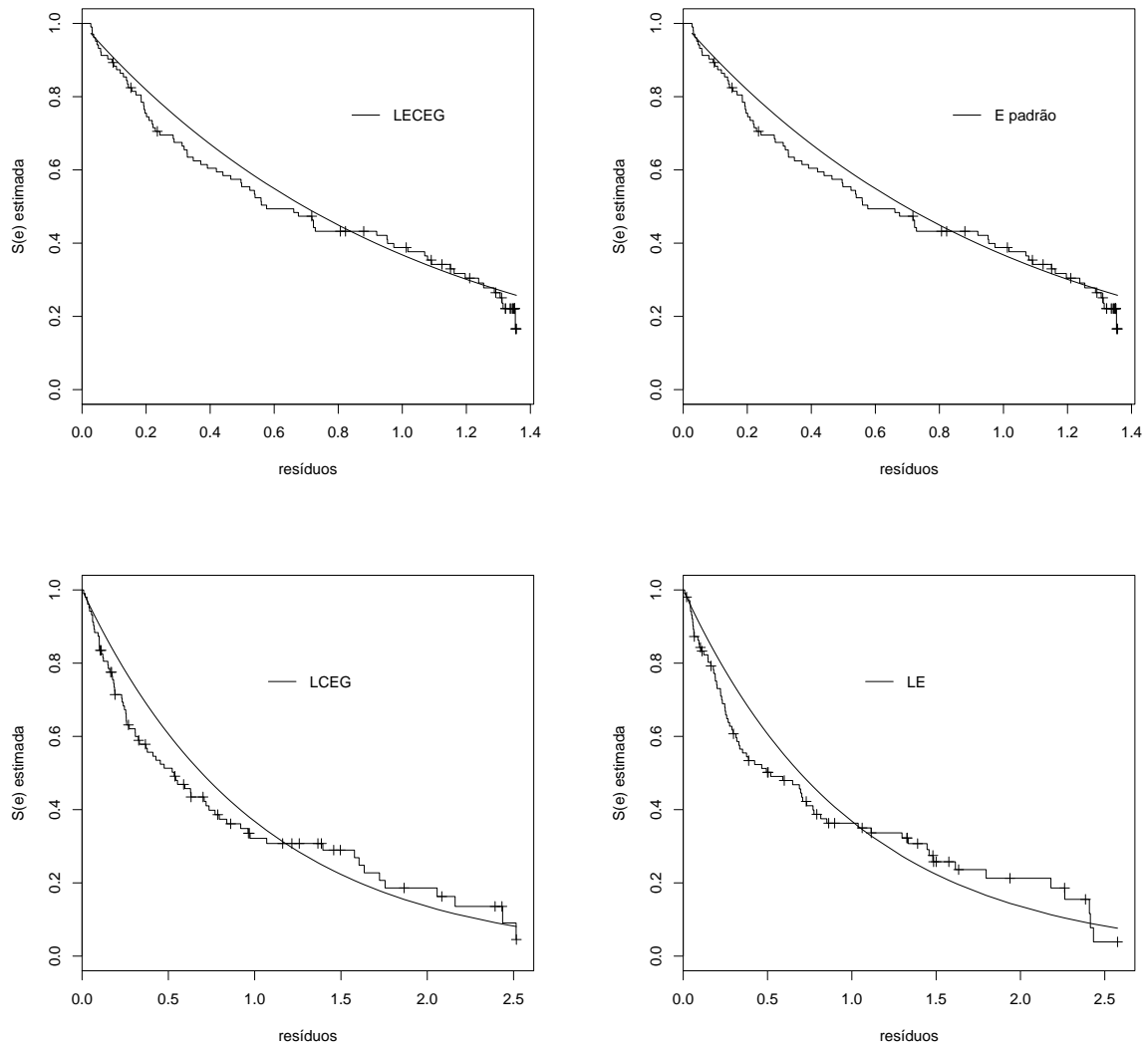


Figura 3.7: Resíduos de Cox Snell *versus* Sobrevida estimada para os modelos LECEG, LCEG e LE

3.6.2 Dados Simulados com covariável

Uma amostra de tamanho 30 foi gerada pelo método da inversa considerando $\alpha = 7,5$, $\lambda = 5,5$, $\theta = 0,9$ e $p = 0,1$. O vetor de censura foi gerado de uma Bernoulli com probabilidade de fracasso de 30%. Foi realizado o ajuste considerando a covariável no parâmetro p . A Tabela 3.10 possui os dados gerados para y e a Tabela 3.11 os dados para x .

Tabela 3.10: Tempos de vida

| | | | | | | | | | | |
|-----|--------|--------|--------|-------|-------|--------|-------|-------|-------|-------|
| y | 0,181* | 0,447 | 0,578* | 1,007 | 0,500 | 0,473 | 0,391 | 0,353 | 0,553 | 0,299 |
| | 0,266 | 0,578 | 0,623 | 0,821 | 0,633 | 0,365 | 1,061 | 0,418 | 0,476 | 0,606 |
| | 0,445 | 1,081* | 0,753 | 0,249 | 0,290 | 0,249* | 0,370 | 0,228 | 0,185 | 0,707 |

*observação censurada

Tabela 3.11: Valores da covariável x

| | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| x | 0,652 | 0,344 | 1,160 | 0,621 | 1,093 | 0,324 | 0,921 | 0,287 | 0,513 | 0,913 |
| | 0,568 | 0,462 | 0,398 | 0,380 | 0,213 | 0,524 | 0,974 | 0,546 | 0,464 | 0,288 |
| | 0,746 | 0,466 | 0,148 | 0,174 | 0,284 | 0,366 | 0,688 | 0,460 | 0,664 | 0,225 |

A Tabela 3.12 possui os critérios e as estimativas de máxima verossimilhança (EMV) para os modelos ajustados. Verifica-se que o modelo LECEG se ajusta melhor em relação a LCEG e LE para todos os critérios.

Tabela 3.12: Valores estimados para o modelo LECEG, LCEG e LE

| Parâmetros | LECEG | | LCEG | | LE | |
|-------------------------|------------------|---------|------------------|-------|------------------|--------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 8,323 | 3,454 | - | - | - | - |
| θ | 0,998 | 0,981 | 0,025 | 0,020 | - | - |
| λ | 5,479 | 0,099 | 7,431 | 1,508 | 1,712 | 0,335 |
| β_0 | -12,642 | 109,413 | -4,092 | 1,671 | -10,487 | 45,599 |
| β_1 | 0,036 | 18,243 | -0,397 | 0,407 | 0,120 | 20,117 |
| $-\log(L(\vartheta_g))$ | -4,707 | | 1,981 | | 12,019 | |
| AIC | 0,585 | | 11,963 | | 30,039 | |
| BIC | 7,591 | | 17,568 | | 34,242 | |
| Estatística KS | 0,049 (0,999) | | 0,076 (0,996) | | 0,266 (0,037) | |

* ϑ_g indica os parâmetros de uma distribuição

A Figura 3.8 possui os gráficos do resíduo de Cox Snell versus a sobrevivência ajustada dos resíduos para cada modelo sobreposta a curva da sobrevivência kaplan meier dos resíduos. Verifica-se que a LECEG possui uma adequabilidade melhor em relação a LCEG e LE, pois a curva de sobrevivência dos resíduos está mais próxima da curva de sobrevivência dos resíduos Kaplan Meier e devido aos resíduos seguirem distribuição exponencial padrão como pode ser verificado pelo teste de Kolmogorov Smirnov.

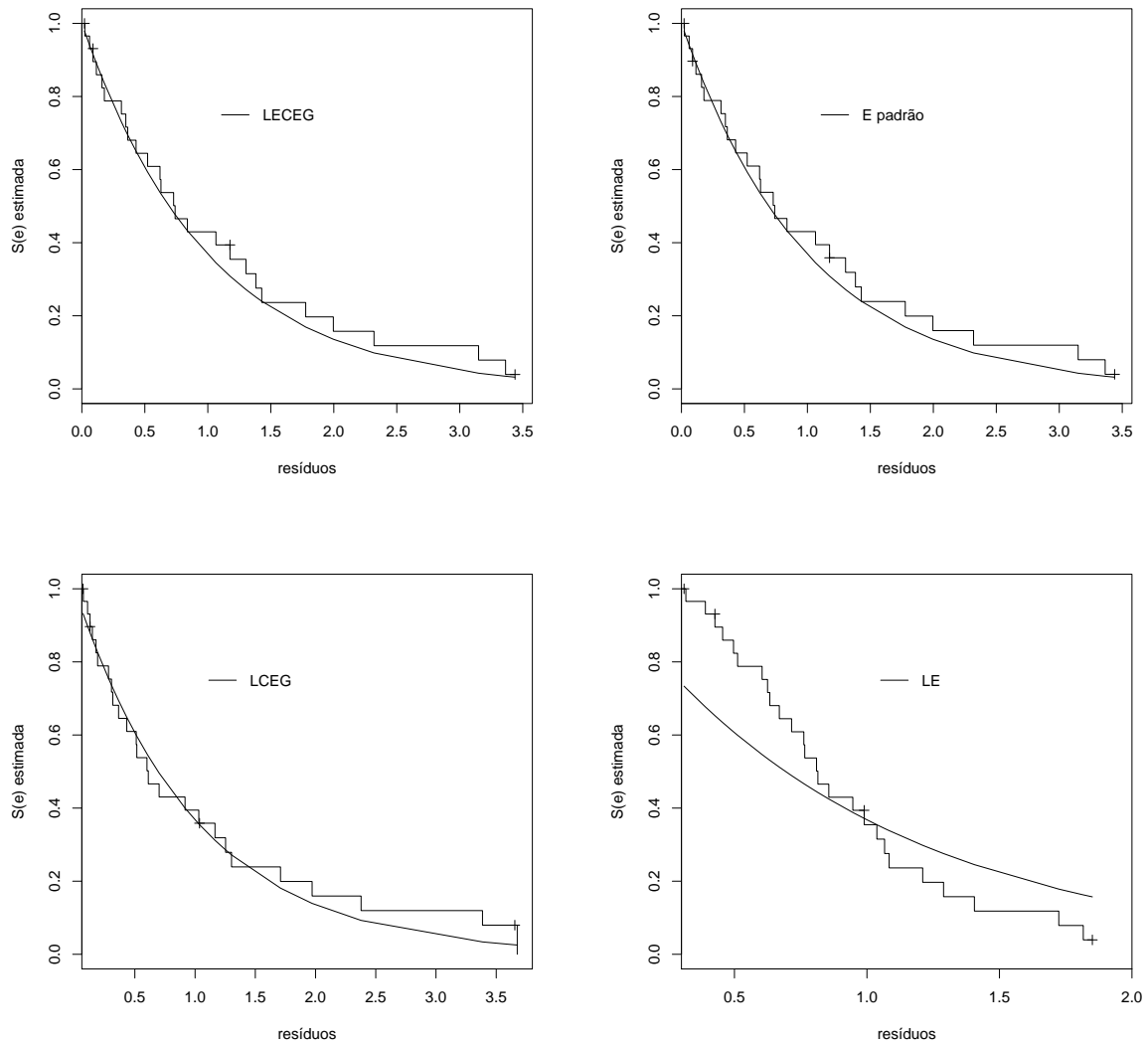


Figura 3.8: Resíduos de Cox Snell versus Sobrevivência estimada dos resíduos para o modelo LECEG, LCEG e LE para os dados simulados com regressão em p

Uma amostra de tamanho 30 foi gerada pelo método da inversa considerando $\alpha = 10,500$, $\lambda = 8,500$, $\theta = 0,700$ e $p = 0,300$. O vetor de censura foi gerado de uma Bernoulli com probabilidade de fracasso de 40%. Foi realizado o ajuste considerando a covariável no parâmetro λ .

A Tabela 3.13 possui os dados gerados para y e a Tabela 3.14 os dados gerados para x .

Tabela 3.13: Tempos de vida

| | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| y | 0,247 | 0,305 | 0,250 | 0,948 | 0,465 | 0,329 | 0,381 | 0,179 | 0,311 | 0,540 |
| | 0,238 | 0,490 | 0,225 | 0,167 | 0,292 | 0,335 | 0,208 | 0,493 | 0,365 | 0,370 |
| | 0,443 | 0,518 | 0,204 | 0,311 | 0,497 | 0,294 | 0,506 | 0,272 | 0,382 | 0,161 |

*observação censurada

Tabela 3.14: Valores da covariável x

| | | | | | | | | | | |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| x | 0,490 | 0,286 | 0,820 | 0,470 | 0,776 | 0,273 | 0,665 | 0,248 | 0,399 | 0,660 |
| | 0,435 | 0,365 | 0,322 | 0,311 | 0,195 | 0,406 | 0,699 | 0,421 | 0,367 | 0,248 |
| | 0,551 | 0,368 | 0,147 | 0,167 | 0,246 | 0,301 | 0,513 | 0,364 | 0,498 | 0,204 |

A Tabela 3.15 possui as estimativas de máxima verossimilhança (EMV) e os critérios para os modelos LECEG, LCEG e LE. Verifica-se que o modelo LECEG possui os melhores critérios.

Tabela 3.15: Valores estimados para o modelo LECEG, LCEG e LE

| parâmetros | LECEG | | LCEG | | LE | |
|-------------------------|----------------------|----------------------|-----------------------|-----------------------|----------------------|-------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 10,534 | 4,702 | - | - | - | - |
| θ | 0,999 | 0,003 | 0,018 | 0,014 | - | - |
| p | $3,4 \times 10^{-7}$ | $1,3 \times 10^{-5}$ | $1,0 \times 10^{-16}$ | $1,6 \times 10^{-14}$ | $1,0 \times 10^{-4}$ | 0,003 |
| β_0 | 2,104 | 0,197 | 2,363 | 0,196 | 0,773 | 0,236 |
| β_1 | -0,015 | 0,036 | -0,001 | 0,007 | -0,033 | 0,088 |
| $-\log(L(\vartheta_g))$ | -13,852 | | -9,011 | | 3,609 | |
| AIC | -17,705 | | -10,023 | | 13,218 | |
| BIC | -10,699 | | -4,418 | | 17,421 | |
| Estatística KS | 0,106 (0,883) | | 0,105 (0,893) | | 0,273 (0,022) | |

* ϑ_g indica os parâmetros de uma distribuição

A Figura 3.9 possui os resíduos de Cox Snell versus a sobrevivência dos resíduos sobrepostas a curva de sobrevivência de Kaplan Meier dos resíduos. Verifica-se que o modelo LECEG possui uma melhor adequabilidade do modelo pois a curva ajustada está próxima da curva de Kaplan Meier. Apesar do teste de Kolmogorov Smirnov indicar que os resíduos da distribuição LCEG na curva da exponencial padrão está mais próxima da curva Kaplan Meier dos resíduos, considerando os outros critérios pode-se dizer que a LECEG é a melhor distribuição.

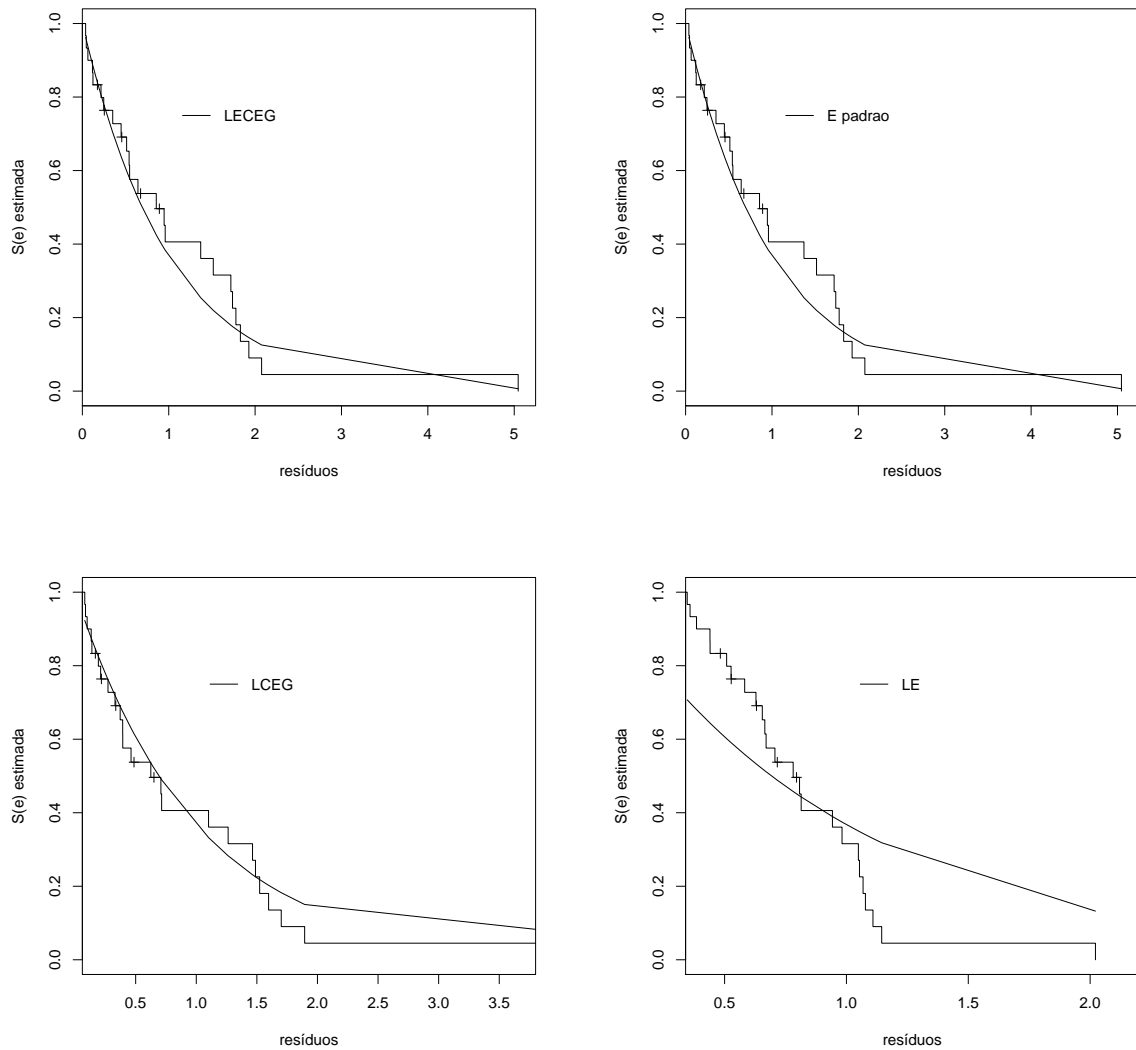


Figura 3.9: Resíduos de Cox Snell versus Sobrevivência estimada dos resíduos para o modelo LECEG, LCEG e LE para os dados simulados com regressão em λ

3.7 Comentários Finais

A distribuição LECEG se apresenta em um contexto em que possui indivíduos que não estão suscetíveis ao evento foi mostrado sua performance em dados reais através de medidas como AIC , BIC e Kolmogorov Smirnov comparando a suas particulares, LCEG e LE e as usuais LLI e LEGU. Os critérios também foram utilizados para o estudo quando covariáveis foram incluídas no parâmetro de escala, tendo como resultado a LECEG melhor do que suas particulares LCEG e LE.

Capítulo 4

Distribuição Exponencial Geométrica Exponenciada

Neste capítulo 4 apresenta-se a origem da distribuição EEG e suas características, tais como: função acumulada, função de sobrevivência, função densidade, função quantil, função de risco, função característica, e função geradora de momentos. Em seguida são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição EEG. O logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição EEG é apresentado na Seção (4.3). Um estudo de simulação é realizado calculando o vício médio e o erro quadrático médio das estimativas dos parâmetros em 1000 simulações para diferentes tamanhos de amostras da distribuição EEG na Seção (4.4). A distribuição EEG é ajustada a dados reais e verificada sua performance através dos critérios $-\log(\cdot)$, AIC, BIC e teste Kolmogorov Smirnov na Seção (4.5).

4.1 Origem

A distribuição Geométrica Exponencial (EG) introduzida por Adamidis & Loukas (1998) possui um parâmetro de escala (λ) e um de forma (θ), e sua função distribuição acumulada é dada por:

$$F_{base}(y) = \frac{1 - e^{-\lambda y}}{1 - \theta e^{-\lambda y}} I_{[0, \infty)}(y). \quad (4.1)$$

A distribuição Geométrica Exponencial Exponenciada (EEG) é obtida considerando como função base F_{base} dada em (4.1) e a função distribuição acumulada da distribuição EG (4.1) apresentada por Adamidis & Loukas (1998).

Desse modo, a função acumulada EEG é escrita como:

$$F(y) = \left(\frac{1 - e^{-\lambda y}}{1 - \theta e^{-\lambda y}} \right)^\alpha I_{[0, \infty)}(y), \quad (4.2)$$

em que $y > 0$, $\alpha > 0$, $0 < \theta < 1$ e $\lambda > 0$.

A função de sobrevivência ou confiabilidade é dada por:

$$S(y) = I_{(-\infty, 0]}(y) + \left(1 - \left(\frac{1 - e^{-\lambda y}}{1 - \theta e^{-\lambda y}} \right)^\alpha \right) I_{[0, \infty)}(y). \quad (4.3)$$

A função de densidade da EEG é dada por:

$$f(y) = \frac{\alpha \lambda (1 - \theta) e^{-\lambda y} (1 - e^{-\lambda y})^{\alpha-1}}{(1 - \theta e^{-\lambda y})^{\alpha+1}} I_{[0, \infty)}(y). \quad (4.4)$$

4.2 Propriedades

O p -ésimo quantil da distribuição geométrica exponencial exponenciada pode ser encontrada pelo inverso da acumulada $F(Q(p)) = p$ e é dado por:

$$Q(p) = F^{-1}(p) = \frac{\log \left(\frac{\theta u^{1/\alpha-1}}{u^{1/\alpha-1}} \right)}{\lambda} I_{[0,1)}(p). \quad (4.5)$$

A partir da expressão (4.5) a mediana pode ser encontrada substituindo p por 0,500, dada por:

$$Q(0,5) = F^{-1}(0,5) = \frac{\log \left(\frac{\theta u^{1/\alpha-1}}{u^{1/\alpha-1}} \right)}{\lambda}. \quad (4.6)$$

A função de risco é escrita como:

$$h(y) = \frac{\alpha \lambda (1 - \theta) e^{-\lambda y} (1 - e^{-\lambda y})^{\alpha-1}}{(1 - \theta e^{-\lambda y}) [(1 - \theta e^{-\lambda y})^\alpha - (1 - e^{-\lambda y})^\alpha]} I_{[0, \infty)}(y). \quad (4.7)$$

A função de risco (4.7) acomoda forma crescente, decrescente ou unimodal como mostra a Figura 4.1 para diferentes valores de α , λ e θ . Defina $\eta(y) = -f'(y)/f(y)$ em que $f'(y)$ é a primeira derivada da função densidade (4.4). Assim,

$$\eta(y) = \lambda - \frac{\lambda(1 - \theta)(\alpha - 1)e^{-2\lambda y}}{(1 - e^{-\lambda y})} + (\alpha + 1)\theta\lambda e^{-\lambda y}. \quad (4.8)$$

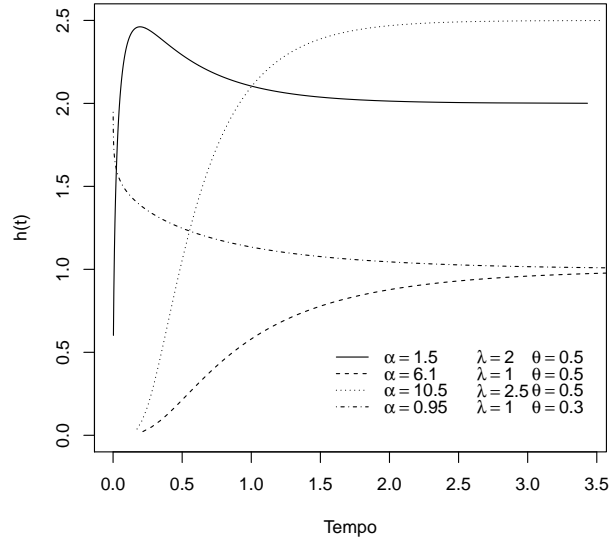


Figura 4.1: Funções de risco de EEG

A primeira derivada de $\eta(y)$ é dada por:

$$\eta'(y) = \frac{(1-\theta)(\alpha-1)e^{-2\lambda y}2\lambda^2}{(1-e^{-\lambda y})} + \frac{\lambda^2(1-\theta)(\alpha-1)e^{-3\lambda y}}{(1-e^{-\lambda y})^2} - \lambda^2\theta(\alpha+1)e^{-\lambda y}. \quad (4.9)$$

Se $\alpha < 1$, $\eta'(y) < 0$ para $y > 0$, do teorema de Glaser (1980), a função de risco (4.7) é decrescente. Para verificar que a função de risco é crescente, considere $(\alpha, \lambda, \theta) = (6, 1; 2, 5; 0, 5)$. Verifica-se a condição $\eta'(y) > 0$ para $y > 0$. De Glaser (1980), a condição de função de risco crescente é satisfeita. Numericamente se observa função crescente quando $\alpha > 1$. Para verificar que a função de risco é unimodal considere $(\alpha, \lambda, \theta) = (1, 6; 17; 0, 85)$, tem-se $y_0 = 0, 012$; tal que a primeira derivada $\eta'(y) = 0$. $\eta'(y) > 0$ para $y \in (0, y_0)$ (e.g. $\eta'(0, 012) = 0, 013$), e $\eta'(y) < 0$ para $y \in (y_0, \infty)$. (e.g. $\eta'(0, 012) = -0, 001$). Assim, $\lim_{y \rightarrow 0} f(y) = 0$. De Glaser (1980), conclui-se que a função de risco é unimodal.

Os momentos da variável aleatória Y pode ser encontrada utilizando a seguinte expressão:

$$E[Y^r] = r \int_0^{\infty} y^{r-1} S(y) dy. \quad (4.10)$$

A expressão do r -ésimo momento $\mu'_r = E(Y^r)$ da variável Y , com densidade dada em (4.4) pode ser encontrada analiticamente se considerarmos a expansão binomial (ver expansão (1)):

Proposição 4.2.1. *O r -ésimo momento da variável Y com distribuição EEG é dada por:*

$$\mu'_r = \frac{r\theta^d}{\lambda^r} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \frac{(-1)_b(-\alpha b)_c(\alpha b)_d}{b!c!d!} (c+d)^{-r} \Gamma(r). \quad (4.11)$$

Demonstração. Das equações (4.3) e (1), usando (4.10), tem-se que:

$$\begin{aligned} \mu'_r &= r \int_0^{\infty} y^{r-1} S(y) dy \\ &= r \int_0^{\infty} y^{r-1} \sum_{b=0}^{\infty} \frac{(-1)_b}{b!} \left(\left(\frac{1 - e^{-\lambda y}}{1 - \theta e^{-\lambda y}} \right)^{\alpha} \right)^b dy \\ &= r \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \frac{\theta^d}{(\lambda)^r} \frac{(-1)_b(-\alpha b)_c(\alpha b)_d}{b!c!d!} \int_0^1 u^{c+d-1} (-\ln u)^{r-1} du, \end{aligned} \quad (4.12)$$

em que $u = e^{-\lambda y}$, seja $v = -\ln u$ e fazendo a adequada substituição encontra-se:

$$\int_0^{\infty} v^{r-1} e^{-v(c+d)} = \int_0^{\infty} \frac{(c+d)^r}{\Gamma(r)} v^{r-1} e^{-v(c+d)} = \frac{\Gamma(r)}{(c+d)^r}$$

e assim se termina a prova. \square

A função de densidade da k -ésima estatística de ordem $Y_{k:n}$ é definida por:

$$f_{k:n}(y) = \frac{n!}{(k-1)!(n-k)!} (F(y))^{k-1} (1-F(y))^{n-k} f(y).$$

O r -ésimo momento da k -ésima estatística de ordem $Y_{k:n}$, é representada por:

$$E[Y_{k:n}^r] = r \sum_{p=n-k+1}^n (-1)^{p-n+k-1} \binom{p-1}{n-k} \binom{n}{p} \int_0^{\infty} y^{r-1} [S(y)]^p dy. \quad (4.13)$$

Proposição 4.2.2. *Para a variável aleatória Y com distribuição EEG, o r -ésimo momento da k -ésima estatística de ordem é dada por:*

$$\begin{aligned} E[Y_{k:n}^r] &= \frac{r}{\lambda^r} \theta^d \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \sum_{p=n-k+1}^n (-1)^{p-n+k-1+r} \binom{p-1}{n-k} \\ &\quad \times \binom{n}{p} \frac{(-p)_b(-\alpha b)_c(\alpha b)_d}{b!c!d!} \theta^d (c+d)^{-r} \Gamma(r). \end{aligned}$$

Demonstração. A partir da expressão (4.13), e utilizando as expressões (1) e (4.3), procedendo da mesma forma como pela Proposição 4.2.1, o resultado é encontrado. \square

Dado que não tenha falha antes de t , a distribuição de tempo de vida residual de uma variável aleatória Y , distribuída como EEG tem função de sobrevivência dada por:

$$S_t(y) = Pr[Y > y + t | Y > t] = \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \frac{(-1)_b(1)_c}{b!c!} \left(\frac{(1 - e^{-\lambda(y+t)})}{(1 - \theta e^{-\lambda(y+t)})} \right)^{\alpha b} \left(\frac{(1 - e^{-\lambda(y)})}{(1 - \theta e^{-\lambda(y)})} \right)^{\alpha c} \quad (4.14)$$

O tempo de vida médio residual de uma variável contínua com função de sobrevivência $S(y)$ é dado por:

$$\mu(t) = E(Y - t | Y > t) = \frac{1}{S(t)} \int_t^{\infty} S(u) du. \quad (4.15)$$

Proposição 4.2.3. *Para a variável aleatória Y com distribuição EEG, a média residual é dada por:*

$$\mu(t) = \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \sum_{e=0}^{\infty} \frac{(-1)_b (-\alpha b)_c (\alpha b)_d}{b! c! d!} \frac{\theta^d e^{-\lambda t(c+d)} (1)_e}{-\lambda (c+d) e!} \left(\frac{(1 - e^{-\lambda y})}{(1 - \theta e^{-\lambda y})} \right)^{\alpha e}. \quad (4.16)$$

Demonstração. A partir da expressão (4.15) e usando $S(y)$ dado em (4.3), tem-se que:

$$\begin{aligned} \frac{1}{S(t)} \int_t^{\infty} S(u) du &= \sum_{e=0}^{\infty} \left(\frac{(1 - e^{-\lambda y})}{(1 - \theta e^{-\lambda y})} \right)^{\alpha e} \int_t^{\infty} \left(1 - \left(\frac{1 - e^{-\lambda u}}{1 - \theta e^{-\lambda u}} \right)^{\alpha} \right) du \\ &= \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \sum_{e=0}^{\infty} \frac{(1)_e}{e!} \left(\frac{(1 - e^{-\lambda y})}{(1 - \theta e^{-\lambda y})} \right)^{\alpha e} \frac{(-1)_b (-\alpha b)_c (\alpha b)_d}{b! c! d!} \frac{\theta^d}{-\lambda} \int_0^{e^{-\lambda t}} u^{c+d-1} du. \end{aligned} \quad (4.17)$$

Utilizando a equação (1), o resultado é encontrado. \square

4.3 Inferência

Sabendo que a função de verossimilhança é escrita como em (1.4), o logaritmo da verossimilhança é dado por:

$$\begin{aligned} l(\alpha, \theta, \lambda, \underline{y}, \underline{\delta}) &= \sum_{i=1}^n \delta_i \ln \left(\frac{\alpha \lambda (1 - \theta) e^{-\lambda y_i} (1 - e^{-\lambda y_i})^{\alpha-1}}{(1 - \theta e^{-\lambda y_i})^{\alpha+1}} \right) \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \ln \left(1 - \left(\frac{1 - e^{-\lambda y_i}}{1 - \theta e^{-\lambda y_i}} \right)^{\alpha} \right), \end{aligned} \quad (4.18)$$

Os estimadores de máxima verossimilhança são encontrados via maximização do logaritmo da verossimilhança, em que δ_i é a variável indicadora assumindo um quando a observação corresponde a uma falha e zero quando a observação for um tempo de censura. No presente trabalho foi utilizado a função *optim* do sistema R (Team (2008)). Foram calculados os critérios $-\log(\cdot)$, AIC, BIC e teste de Kolmogorov Smirnov para comparar os ajustes.

As EMV $\hat{\alpha}$, $\hat{\lambda}$ e $\hat{\theta}$, são soluções das seguintes equações:

$$\begin{aligned} \sum_{i=1}^n \frac{\delta_i}{\alpha} + \sum_{i=1}^n \delta_i \ln(B_i) &= \frac{(1 - \delta_i)}{(1 - B_i^\alpha)} B_i^\alpha \ln B_i; \\ \sum_{i=1}^n \frac{\delta_i}{\lambda} + (\alpha - 1) \frac{\sum_{i=1}^n \delta_i X_i}{K_i} &= \sum_{i=1}^n \delta_i y_i + (\alpha + 1) \sum_{i=1}^n \frac{\delta_i \theta X_i}{Z_i} + \alpha(1 - \theta) \sum_{i=1}^n \frac{(1 - \delta_i)}{1 - B_i^\alpha} B_i^{\alpha-1} \frac{X_i}{Z_i^2}; \\ (\alpha + 1) \sum_{i=1}^n \frac{\delta_i e^{-\lambda y_i}}{Z_i} &= \frac{\sum_{i=1}^n \delta_i}{(1 - \theta)} + \alpha \sum_{i=1}^n \frac{(1 - \delta_i)}{(1 - B_i^\alpha)} B_i^{\alpha-1} \frac{B_i e^{-\lambda y_i}}{Z_i}. \end{aligned}$$

Em que $K_i = 1 - e^{-\lambda y_i}$, $Z_i = 1 - \theta e^{-\lambda y_i}$, $B_i = \frac{K_i}{Z_i}$ e $X_i = y_i e^{-\lambda y_i}$. A matriz de informação de Fisher para $(\alpha, \theta, \lambda)$ é dada por:

$$I_F(\alpha, \theta, \lambda) = - \left(\begin{array}{ccc} I_{\alpha, \alpha} & I_{\alpha, \theta} & I_{\alpha, \lambda} \\ I_{\theta, \alpha} & I_{\theta, \theta} & I_{\theta, \lambda} \\ I_{\lambda, \alpha} & I_{\lambda, \theta} & I_{\lambda, \lambda} \end{array} \right) \Bigg|_{(\alpha, \theta, \lambda) = (\hat{\alpha}, \hat{\theta}, \hat{\lambda})}, \quad (4.19)$$

em que os elementos de $I_F(\alpha, \theta, \lambda)$ estão dados no Apêndice A. A distribuição assintótica de $(\hat{\alpha}, \hat{\theta}, \hat{\lambda})$, para $n \rightarrow \infty$, é uma Normal com média zero e matriz de variância e covariância $I_F^{-1}(\alpha, \theta, \lambda)$.

4.4 Estudo de simulação

4.4.1 Estudo das propriedades dos estimadores por M.V. dos parâmetros

O estudo se baseia na geração de 1000 conjuntos de dados com distribuição EEG para diferentes valores de α , λ e θ para $n = 200, 300, 400$ e 500 . Para o parâmetro θ a seguinte transformação foi considerada: $\theta = \frac{e^{\theta^*}}{(1 + e^{\theta^*})}$, em que $\theta^* \in (0, 1)$ e para λ e α se considerou a transformação exponencial. Para o cálculo das variâncias foi considerado o método Delta. Os resultados estão apresentados na Tabela (4.1), no qual estão as médias das estimativas de máxima verossimilhança ($\text{Av}(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$), junto com as probabilidades de cobertura com intervalo de confiança de 95% $\text{EEG}[C(\hat{\alpha}, \hat{\lambda}, \hat{\theta})]$ e calculado as variâncias e encontrado o EQM[EQM(($\hat{\alpha}$), ($\hat{\lambda}$), ($\hat{\theta}$))]. Apesar de uma alta probabilidade de cobertura para α , o EQM e os valores médios estimados da Tabela 4.1 indicam que em algumas simulações os valores foram muito viciados. Fato que melhora a partir do tamanho amostral 200. Para λ e θ temos boas estimativas para todos os cenários e tamanhos amostrais.

Tabela 4.1: Média das EMV, probabilidade de cobertura e EQM.

| n | $(\alpha, \lambda, \theta)$ | $Av(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$ | $C(\alpha, \lambda, \theta)$ | $EQM(\hat{\alpha})$ | $EQM(\lambda)$ | $EQM(\hat{\theta})$ |
|-----|-----------------------------|---|------------------------------|----------------------|----------------|---------------------|
| 30 | (0,500;1,000;0,700) | (0,574;1,239;0,506) | (0,949;0,924;0,688) | 0,065 | 0,750 | 0,168 |
| | (0,700;1,000;0,500) | (7,772;1,000;0,420) | (0,952;0,948;0,578) | 47037,690 | 0,241 | 0,134 |
| | (1,000;0,500;0,600) | (18,051;0,503;0,503) | (0,941;0,944;0,634) | 52634,510 | 0,060 | 0,138 |
| | (2,500;1,000;0,500) | (13834610;0,927;0,537) | (0,935;0,931;0,540) | $1,9 \times 10^{17}$ | 0,100 | 0,155 |
| | (3,000;2,500;0,900) | (81631,72;2,713;0,820) | (0,887;0,937;0,736) | $4,9 \times 10^{12}$ | 2,633 | 0,067 |
| 50 | (0,500;1,000;0,700) | (0,542;1,109;0,540) | (0,962;0,887;0,744) | 0,022 | 0,475 | 0,139 |
| | (0,700;1,000;0,500) | (0,574;1,239;0,506) | (0,963;0,932;0,624) | 0,114 | 0,159 | 0,118 |
| | (1,000;0,500;0,600) | (26,522;0,479;0,510) | (0,947;0,916;0,696) | 523242,3 | 0,040 | 0,116 |
| | (2,500;1,000;0,500) | (6163,766;0,907;0,510) | (0,905;0,925;0,570) | 9531671759 | 0,067 | 0,137 |
| | (3,000;2,500;0,900) | (3888,789;2,482;0,835) | (0,882;0,932;0,792) | 5877669204 | 1,626 | 0,048 |
| 100 | (0,500;1,000;0,700) | (0,518;0,218;0,080) | (0,953;0,903;0,871) | 0,007 | 0,206 | 0,074 |
| | (0,700;1,000;0,500) | (0,738;0,917;0,440) | (0,952;0,915;0,778) | 0,021 | 0,088 | 0,077 |
| | (1,000;0,500;0,600) | (1,065;0,458;0,545) | (0,944;0,902;0,829) | 0,074 | 0,022 | 0,068 |
| | (2,500;1,000;0,500) | (51,348;0,895;0,508) | (0,904;0,896;0,725) | 345414,3 | 0,046 | 0,092 |
| | (3,000;2,500;0,900) | (296,445;2,251;0,878) | (0,906;0,903;0,867) | 20285192 | 0,780 | 0,013 |
| 200 | (0,500;1,000;0,700) | (0,506;0,918;0,637) | (0,949;0,913;0,932) | 0,003 | 0,114 | 0,042 |
| | (0,700;1,000;0,500) | (0,714;0,896;0,450) | (0,943;0,884;0,855) | 0,009 | 0,056 | 0,051 |
| | (1,000;0,500;0,600) | (1,019;0,446;0,563) | (0,936;0,878;0,894) | 0,029 | 0,0138 | 0,041 |
| | (2,500;1,000;0,500) | (2,883;0,892;0,501) | (0,911;0,859;0,806) | 2,993 | 0,446 | 0,004 |
| | (3,000;2,500;0,900) | (3,262;2,152;0,890) | (0,894;0,872;0,899) | 3,238 | 0,446 | 0,004 |
| 500 | (0,500;1,000;0,700) | (0,503;0,876;0,672) | (0,934;0,857;0,948) | 0,001 | 0,055 | 0,013 |
| | (0,700;1,000;0,500) | (0,704;0,887;0,468) | (0,939;0,831;0,940) | 0,003 | 0,032 | 0,024 |
| | (1,000;0,500;0,600) | (1,000;0,440;0,584) | (0,933;0,814;0,936) | 0,010 | 0,008 | 0,015 |
| | (2,500;1,000;0,500) | (2,570;0,895;0,502) | (0,940;0,766;0,901) | 0,276 | 0,019 | 0,027 |
| | (3,000;2,500;0,900) | (2,921;2,114;0,897) | (0,888;0,770;0,927) | 0,372 | 0,271 | 0,001 |

4.5 Aplicação

4.5.1 Ajuste a conjunto de dados reais da distribuição EEG e outras distribuições

Nesta seção o ajuste da distribuição EEG é comparada com outras distribuições em conjunto de dados extraídos da literatura. Dois conjunto de dados são analisados, um com função de risco decrescente e outro unimodal. A função de densidade da distribuição Exponencial considerada nos ajustes foi:

$$f(x) = \lambda e^{-\lambda x}$$

. A distribuição Exponencial Exponenciada proposta por Nadarajah & Kotz (2006) com função densidade dada por:

$$f(x) = \alpha \lambda e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha-1}$$

. A distribuição Geométrica Exponencial introduzida por Adamidis & Loukas (1998) com densidade:

$$f(x) = \lambda(1 - \theta)e^{-\lambda x} / (1 - \theta e^{-\lambda x})^2$$

. A Extensão da distribuição Geométrica Exponencial apresentada por Adamidis *et al.* (2005) com densidade dada por:

$$\lambda\theta e^{-\lambda x}/(1 - (1 - \theta)e^{-\lambda x})^2$$

. A Exponencial Poisson Exponenciada introduzida por Ristic & Nadarajah (2012) com densidade dada por:

$$f(x) = \alpha\beta\lambda e^{-\beta x}(1 - e^{-\beta x})^{\alpha-1}e^{-\lambda(1-e^{-\beta x})^\alpha}/(1 - e^{-\lambda})$$

. A distribuição Gama com função densidade:

$$f(x) = 1/(\alpha^\lambda\Gamma(\lambda))x^{\lambda-1}e^{-x/\alpha}.$$

O primeiro conjunto de dados T1 consiste de tempos de sobrevivência de 103 possíveis receptores de transplante de coração desde o tempo que entraram no programa. (dados encontrados no Livro Kalbfleisch & Prentice (1980) do artigo de Crowley & Hu (1977)). O segundo conjunto de dados T2 refere-se a tempos de sobrevivência de pacientes submetidos a quimioterapia combinados com radioterapia no tratamento de câncer gástrico. (dados de Stablein & Koutrouvelis (1985)).

Nas Tabelas 4.2 e 4.3 tem-se $-\log(\cdot)$, AIC, BIC, o teste de Kolmogorov Smirnov, e as estimativas para todos os ajustes considerados para os conjuntos de dados T1 e T2. As Figuras 4.2 e 4.3 mostram as funções de sobrevivências ajustados sobrepostos a curva de sobrevivência Kaplan Meier para os dois conjuntos de dados considerados.

Tabela 4.2: Ajuste de seis distribuições ao conjunto de dados T1: estimativas dos parâmetros dos modelos E, EE, EG, ExtEG, EEG e Gama

| Parâmetros | E | | EE | | EG | | ExtEG | | EEG | | Gama | |
|-------------------------|------------------|--------------------|------------------|--------------------|---------------------|-----------------------|-----------------------|-----------------------|--------------------|----------------------|-----------------------------------|--------------------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | - | - | 0,449 | 0,056 | - | - | - | - | 0,613 | 0,108 | 8×10^{-4} | 1×10^{-4} |
| λ | 0,002 | 2×10^{-4} | 0,001 | 2×10^{-4} | 2×10^{-12} | $1,7 \times 10^{-12}$ | $4,1 \times 10^{-12}$ | $5,6 \times 10^{-12}$ | 1×10^{-6} | $3,7 \times 10^{-5}$ | 0,454 | 0,115 |
| θ | - | - | - | - | 1 | $2,5 \times 10^{-10}$ | $5,7 \times 10^{-10}$ | $7,7 \times 10^{-10}$ | 0,999 | 0,012 | - | - |
| $-\log(L(\vartheta_g))$ | 529,092 | | 504,484 | | 505,002 | | 505,002 | | 501,456 | | 504,314 | |
| AIC | 1060,185 | | 1012,969 | | 1014,006 | | 1014,006 | | 1008,913 | | 1012,629 | |
| BIC | 1062,819 | | 1018,238 | | 1019,275 | | 1019,275 | | 1016,817 | | 1017,899 | |
| Estatística KS | 0,120 (0,100) | | 0,122 (0,090) | | 0,123 (0,085) | | 0,123 (0,085) | | 0,101 (0,241) | | 0,269 ($6,1 \times 10^{-7}$) | |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p -valor do teste de Kolmogorov Smirnov

Tabela 4.3: Ajuste de seis distribuições ao conjunto de dados T2: estimativas dos parâmetros dos modelos E, EE, EG, ExtEG, EEG e Gama

| Parâmetros | E | | EE | | EG | | ExtEG | | EEG | | Gama | |
|-------------------------|------------------|--------------------|--------------------|--------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|----------------------|--------------------|--------------------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | - | - | 0,644 | 0,127 | - | - | - | - | 2,699 | 1,540 | 7×10^{-4} | 2×10^{-4} |
| λ | 0,001 | 1×10^{-4} | 8×10^{-4} | 2×10^{-4} | $7,7 \times 10^{-13}$ | 7×10^{-13} | $1,3 \times 10^{-8}$ | $2,7 \times 10^{-7}$ | $5,5 \times 10^{-7}$ | $1,2 \times 10^{-5}$ | 0,645 | 0,123 |
| θ | - | - | - | - | 1 | $2,3 \times 10^{-10}$ | $4,3 \times 10^{-6}$ | $8,7 \times 10^{-5}$ | 0,999 | 0,001 | - | - |
| $-\log(L(\vartheta_g))$ | 299,962 | | 297,296 | | 290,0889 | | 290,091 | | 287,759 | | 297,020 | |
| AIC | 601,925 | | 598,593 | | 584,177 | | 584,182 | | 581,518 | | 598,040 | |
| BIC | 603,732 | | 602,207 | | 587,791 | | 587,795 | | 586,938 | | 601,653 | |
| Estatística KS | 0,242 (0,010) | | 0,209 (0,038) | | 0,114 (0,598) | | 0,114 (0,598) | | 0,067 (0,985) | | 0,202 (0,049) | |

* ϑ_g indica os parâmetros de uma distribuição e em parênteses o p -valor do teste de Kolmogorov Smirnov

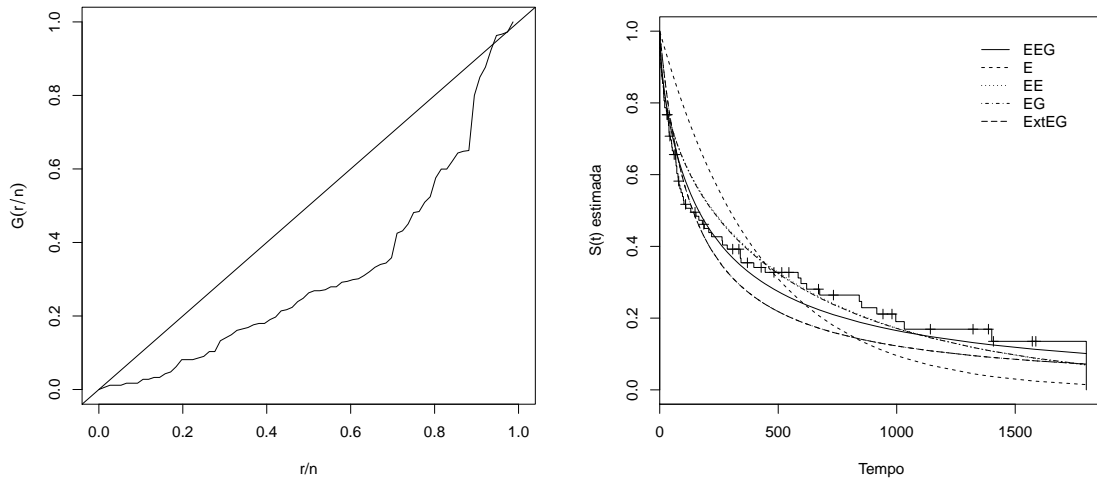


Figura 4.2: TTT plot (painel esquerdo), Kaplan Meier e curva de sobrevivência ajustada para T1 (painel direito).

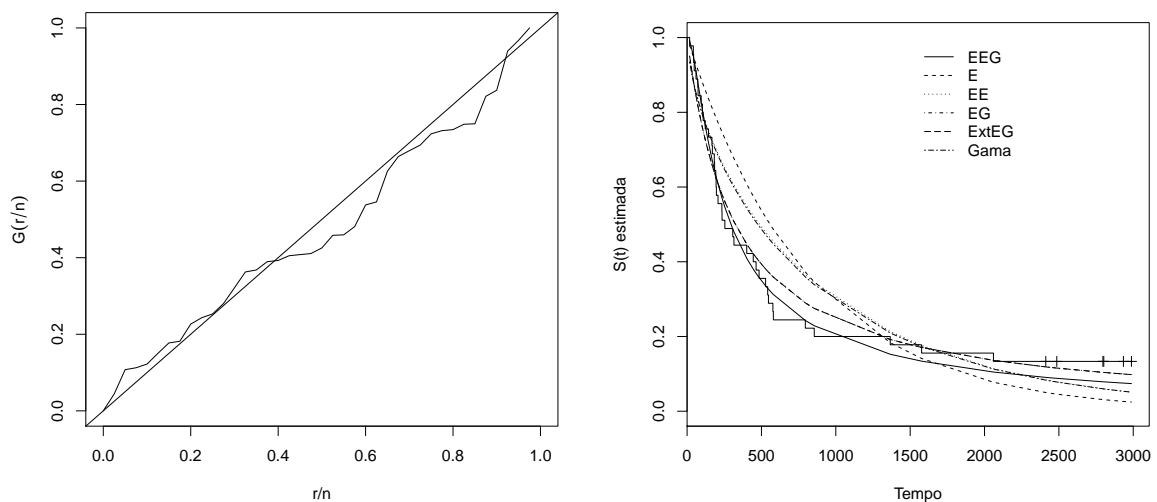


Figura 4.3: TTT plot (painel esquerdo), Kaplan Meier e curva de sobrevivência ajustada para T2 (painel direito).

4.6 Comentários Finais

A distribuição EEG é uma generalização da distribuição EG proposta por Adamidis & Loukas (1998), que acomoda função de risco crescente, decrescente e unimodal. Propriedades como a função de densidade, função de risco, função de sobrevivência, momentos, estatísticas de ordem e média residual foram apresentados. A estimação via máxima verossimilhança foi realizada em conjuntos de dados reais e um estudo de simulação para verificar o erro quadrático médio e a probabilidade de cobertura.

Capítulo 5

Distribuição Exponencial geométrica complementar exponenciada considerando mais de um sistema

Neste capítulo apresenta-se a origem da distribuição ECEGN e suas características, tais como: função acumulada, função de sobrevivência, função densidade, função quantil, função de risco, função característica e a função geradora de momentos. Em seguida são deduzidas a função densidade, a função característica e alguns momentos da k -ésima estatística de ordem de uma amostra aleatória de tamanho n da distribuição ECEGN. O logaritmo da verossimilhança referente a uma amostra aleatória de tamanho n da distribuição ECEGN é apresentado na Seção (5.3). A distribuição ECEGN é ajustada a dados reais e verificada sua performance por meio dos critérios $-\log(\cdot)$, AIC , BIC e teste de Kolmogorov Smirnov na seção (5.5).

5.1 Origem

Suponha que uma companhia tenha N sistemas funcionando independentemente e cada sistema siga uma distribuição Geométrica com função de probabilidade dada por:

$$P[N = n] = \theta(1 - \theta)^{n-1}. \quad (5.1)$$

Suponha que cada sistema possua α unidades funcionando em paralelo independentemente e que a falha do sistema ocorra quando todas as unidades do sistema falhem. Assuma que cada unidade do i -ésimo sistema possua distribuição CEG, ou seja, $Z_{i1}, Z_{i2}, \dots, Z_{i\alpha}$, sejam os tempos de vida de cada unidade do sistema com parâmetro de escala λ e de forma θ . Seja Y_i o tempo de falha do i -ésimo sistema. O i -ésimo sistema falha quando todas as

unidades pararem de funcionar, isto é, $Y_i = \max(Z_{i1}, Z_{i2}, \dots, Z_{i\alpha})$. Considere X a variável aleatória que verifica o primeiro sistema a falhar, ou seja, $X = \min(Y_1, Y_2, \dots, Y_N)$. A distribuição condicional de X dado N é dada por:

$$\begin{aligned} G(x|N) &= 1 - Pr(X > x|N) \\ &= 1 - Pr^N(Y_1 > x) \\ &= 1 - \left[1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right]^N I_{[0, \infty)}(x). \end{aligned}$$

Então a acumulada de X é encontrada como:

$$\begin{aligned} G(x) &= 1 - \theta \sum_{n=1}^{\infty} (1 - \theta)^{n-1} \left[1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right]^N \\ &= 1 - \left[\frac{\theta \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right)}{\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha} \right] I_{[0, \infty)}(y). \end{aligned} \quad (5.2)$$

A função de sobrevivência ou confiabilidade é dada por:

$$S(x) = I_{(-\infty, 0]}(y) + \left(\frac{\theta \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda y}(1 - \theta) + \theta} \right)^\alpha \right)}{\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha} \right) I_{[0, \infty)}(y). \quad (5.3)$$

A função de densidade da ECEGN é dada por:

$$f(x) = \frac{\alpha \theta \left(\frac{(1 - e^{-\lambda x})\theta}{e^{-\lambda x}(1 - \theta) + \theta} \right)^{-1 + \alpha} \left(\frac{e^{-\lambda x} \theta \lambda}{(e^{-\lambda x}(1 - \theta) + \theta)^2} \right)}{\left(\theta + (1 - \theta) \left(\frac{(1 - e^{-\lambda x})\theta}{e^{-\lambda x}(1 - \theta) + \theta} \right)^\alpha \right)^2} I_{[0, \infty)}(y). \quad (5.4)$$

5.2 Propriedades

A inversa da função de sobrevivência da distribuição ECEGN é dada por:

$$Q(p) = -\frac{1}{\lambda} \ln \left(\frac{\theta \left(-1 + \left(\frac{(-1+p)\theta}{-\theta - p + p\theta} \right)^{\frac{1}{\alpha}} \right)}{-\theta - \left(\frac{(-1+p)\theta}{-\theta - p + p\theta} \right)^{\frac{1}{\alpha}} + \left(\frac{(-1+p)\theta}{-\theta - p + p\theta} \right)^{\frac{1}{\alpha}} \theta} \right) I_{[0, 1)}(p). \quad (5.5)$$

Por meio das expressões (5.3) e (5.4), temos que a função de risco é dada por

$$h(x) = \frac{\alpha \left(\frac{(1 - e^{-\lambda x})\theta}{e^{-\lambda x}(1 - \theta) + \theta} \right)^{-1 + \alpha} \left(\frac{e^{-\lambda x} \theta \lambda}{(e^{-\lambda x}(1 - \theta) + \theta)^2} \right)}{\left(\theta + (1 - \theta) \left(\frac{(1 - e^{-\lambda x})\theta}{e^{-\lambda x}(1 - \theta) + \theta} \right)^\alpha \right) \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda x}(1 - \theta) + \theta} \right)^\alpha \right)} I_{[0, \infty)}(y). \quad (5.6)$$

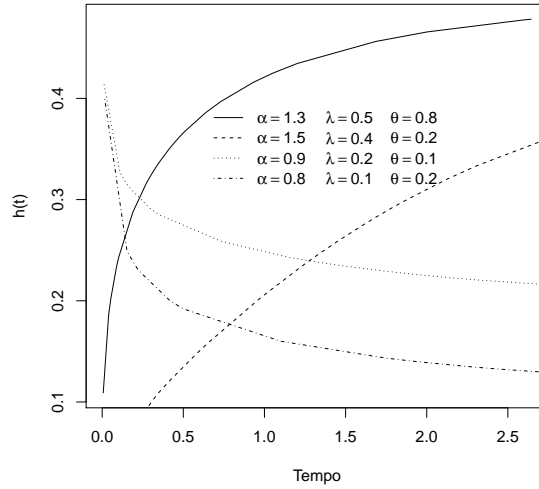


Figura 5.1: Funções de risco da ECEGN

$$\begin{aligned}
 \text{Defina } \eta(x) &= -\frac{f'(x)}{f(x)} \\
 &= -\frac{2(1-\theta)e^{-\lambda x}}{e^{-\lambda x}(1-\theta)+\theta} + \frac{\lambda}{(e^{-\lambda x}(1-\theta)+\theta)^4} + \frac{\lambda\theta e^{-\lambda x}(1+\theta)}{(1-e^{-\lambda x})\theta(e^{-\lambda x}(1-\theta)+\theta)} \\
 &\quad - \frac{\alpha\theta\lambda e^{-\lambda x}(1-\theta)}{(1-e^{-\lambda x})\theta(e^{-\lambda x}(1-\theta)+\theta)} \\
 &\quad + \frac{2e^{-\lambda x}(1-\theta)\theta(1-e^{-\lambda x})^{\alpha-1}\theta^{\alpha-1}\lambda}{\theta(e^{-\lambda x}(1-\theta)+\theta)+(1-\theta)((1-e^{-\lambda x})\theta)^\alpha(e^{-\lambda x}(1-\theta)+\theta)^{1-\alpha}}.
 \end{aligned}$$

Assim,

$$\begin{aligned}
 \eta'(x) &= \frac{2e^{-\lambda x}(1-\theta)\lambda}{e^{-\lambda x}(1-\theta)+\theta} \left(1 - \frac{(1-\theta)e^{-\lambda x}}{e^{-\lambda x}(1-\theta)+\theta} + \frac{2\lambda}{(e^{-\lambda x}(1-\theta)+\theta)^4} \right) \\
 &\quad - (\alpha-1) \frac{e^{-2\lambda x}(1-\theta)^2\lambda^2}{(1-e^{-\lambda x})(e^{-\lambda x}(1-\theta)+\theta)^2} + \frac{e^{-\lambda x}(1-\theta)\alpha\lambda^2}{(1-e^{-\lambda x})(e^{-\lambda x}(1-\theta)+\theta)} \\
 &\quad \left(\frac{e^{-\lambda x}}{1-e^{-\lambda x}} + 1 \right) - \frac{e^{-\lambda x}(1+\theta)\lambda^2}{(1-e^{-\lambda x})(e^{-\lambda x}(1-\theta)+\theta)} \left(\frac{e^{-\lambda x}}{1-e^{-\lambda x}} + 1 \right) \\
 &\quad - \frac{2e^{-\lambda x}(1-e^{-\lambda x})^{-1+\alpha}\theta^\alpha(1+\theta)\lambda^2}{\theta(e^{-\lambda x}(1-\theta)+\theta)+(1-\theta)((1-e^{-\lambda x})\theta)^\alpha(e^{-\lambda x}(1-\theta)+\theta)^{1-\alpha}} \\
 &\quad \times \left(1 + \frac{e^{-\lambda x}}{1-e^{-\lambda x}}(1-\alpha) \right) \\
 &\quad - \frac{2e^{-\lambda x}(1-e^{-\lambda x})^{-1+\alpha}\theta^\alpha(1+\theta)\lambda(-e^{-\lambda x}(1-\theta)\theta\lambda+e^{-\lambda x}\alpha(1-\theta)\theta)}{(\theta(e^{-\lambda x}(1-\theta)+\theta)+(1-\theta)((1-e^{-\lambda x})\theta)^\alpha(e^{-\lambda x}(1-\theta)+\theta)^{1-\alpha})^2} \\
 &\quad \times ((1-e^{-\lambda x})\theta)^{-1+\alpha} \frac{\lambda+e^{-\lambda x}(1-\alpha)(1-\theta)^2((1-e^{-\lambda x})\theta)^\alpha(e^{-\lambda x}(1-\theta)+\theta)^{-\alpha}\lambda}{(\theta(e^{-\lambda x}(1-\theta)+\theta)+(1-\theta)((1-e^{-\lambda x})\theta)^\alpha(e^{-\lambda x}(1-\theta)+\theta)^{1-\alpha})^2} \quad (5.7)
 \end{aligned}$$

Se $\alpha > 1$ e $y > 0$, $\eta'(x) > 0$, dessa forma pelo Teorema de Glaser (1980) a função de risco

é crescente.

Para verificar que a função de risco é crescente, considere $(\alpha, \lambda, \theta) = (3,5; 2,5; 0,5)$. Verifica-se a condição $\eta(y) > 0$ para $y > 0$. De Glaser (1980), a função de risco é crescente.

Se $0 < \alpha < 1$ e $y > 0$ $\eta'(x) < 0$, logo pelo Teorema de Glaser (1980) a função de risco é decrescente.

Para verificar que a função de risco é decrescente, considere $(\alpha, \lambda, \theta) = (0,8; 2,5; 0,9)$. Verifica-se a condição $\eta(y) < 0$ para $y > 0$. De Glaser (1980), conclui-se que a função de risco é decrescente.

Proposição 5.2.1. *Considere $T \sim F_T(\cdot)$ e $U \sim U(0, 1)$. Então $S_T^{-1}(u)$ tem a distribuição de T , $(F_T(\cdot))$.*

Demonstração.

$$\begin{aligned}
 P(S_T^{-1}(U) > a) &= P(U > S_T(a)) \\
 &= S_U(S_T(a)) \\
 &= 1 - F_U(S_T(a)) \\
 &= 1 - S_T(a) \\
 &= 1 - (1 - F_T(a)) \\
 &= F_T(a).
 \end{aligned}$$

□

Proposição 5.2.2. *A função característica da variável aleatória ECEGN é dada por:*

$$\begin{aligned}
 \phi_X(t) &= \int_0^\infty e^{itx} f(x) dx \\
 &= \sum_{k=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \frac{(2)_k}{k!} \frac{(-k)_l}{l!} \frac{(\frac{1}{\lambda}(it - \lambda) + 1)_m}{m!} \frac{((1 - \theta))_n}{n!} \theta^{k(1-\alpha)+1-\alpha+\alpha l - \alpha k} \\
 &\times \alpha (-1)^k \frac{1}{m - \alpha + \alpha l - \alpha k + n}.
 \end{aligned} \tag{5.8}$$

Demonstração. Utilizando a expressão (1) e considerando $u = (1 - e^{-\lambda y})$, temos que:

$$\begin{aligned}
\phi_X(t) &= \int_0^\infty e^{itx} f(x) dx \\
&= \int_0^\infty e^{x(it-\lambda)} \theta^2 \lambda \alpha \left(\frac{(1 - e^{-\lambda x}) \theta}{e^{-\lambda x} (1 - \theta) + \theta} \right)^{-1+\alpha} \left((e^{-\lambda x} (1 - \theta) + \theta) \right)^{-2} \\
&\quad \times \left(\theta + (1 - \theta) \left(\frac{(1 - e^{-\lambda x}) \theta}{e^{-\lambda x} (1 - \theta) + \theta} \right)^\alpha \right)^{-2} dx \\
&= \int_0^\infty e^{x(it-\lambda)} \theta^2 \lambda \alpha \left(\frac{(1 - e^{-\lambda x}) \theta}{e^{-\lambda x} (1 - \theta) + \theta} \right)^{-1+\alpha} \left(e^{-\lambda x} (1 - \theta) + \theta \right)^{-2} \\
&\quad \times \left((1 - \theta) \left(\frac{(1 - e^{-\lambda x}) \theta}{e^{-\lambda x} (1 - \theta) + \theta} \right)^\alpha \right)^{-2} \left(\frac{\theta}{(1 - \theta) \left(\frac{(1 - e^{-\lambda x}) \theta}{e^{-\lambda x} (1 - \theta) + \theta} \right)^\alpha + 1} \right)^{-2} dx \\
&= \sum_{k=0}^\infty \sum_{l=0}^\infty \frac{(2)_k}{k!} \frac{(-k)_l}{l!} \theta^{k(1-\alpha)+2} \alpha (-1)^k \int_0^1 (1-u)^{\frac{-1}{\lambda}(it-\lambda)-1} (u\theta)^{-1-\alpha+\alpha l-\alpha k} \\
&\quad \times (1-u(1-\theta))^{-1-\alpha+\alpha l-\alpha k} du.
\end{aligned}$$

□

Seja X_1, \dots, X_n uma amostra aleatória da distribuição ECEGN e $X_{1:n}, \dots, X_{n:n}$ as respectivas estatísticas de ordem. Então, a função de densidade $f_{k:n}(x)$ da k -ésima estatística de ordem $X_{k:n}$ é dada por:

$$f_{k:n}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{i-1} (1-F(x))^{n-i} f(x) I_{[0,\infty)}(x). \quad (5.9)$$

O momento da k -ésima estatística de ordem $X_{k:n}$, pode ser representado por:

$$E[X_{k:n}^r] = r \sum_{p=n-k+1}^n (-1)^{p-n+k-1} \binom{p-1}{n-k} \binom{n}{p} \int_0^\infty x^{r-1} [S(x)]^p dx. \quad (5.10)$$

Proposição 5.2.3. *Para a variável aleatória X com distribuição ECEGN temos que o r -ésimo momento da k -ésima estatística de ordem é dada por:*

$$\begin{aligned}
E[X_{k:n}^r] &= \frac{r}{\lambda^r} \sum_{p=n-k+1}^n (-1)^{p-n+k-1} \sum_{b=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \sum_{t=0}^\infty \sum_{a=0}^\infty \binom{p-1}{n-k} \binom{n}{p} \\
&\quad \times \frac{(-p)_b}{b!} \frac{(p)_l}{l!} \frac{(-l)_m}{m!} \frac{(\alpha b - \alpha p + \alpha m - \alpha l)_n}{n!} \frac{(1)_t}{t!} \frac{(-\alpha b + \alpha p - \alpha l - n - t)_a}{a!} \\
&\quad \times \theta^{p+l+\alpha b+\alpha p+\alpha m-\alpha l} (1-\theta)^{p+n} \frac{(-1)^l}{\lambda^r} \frac{\Gamma(r)}{(1+a)^r}.
\end{aligned}$$

Demonstração. Resolvendo a integral:

$$\begin{aligned}
\int_0^\infty x^{r-1}[S(x)]^p dx &= \sum_{b=0}^\infty \sum_{l=0}^\infty \frac{(-p)_b (p)_l}{b! l!} (1-\theta)^p \theta^l \int_0^\infty x^{r-1} \theta^p \left(\frac{\theta(1-e^{-\lambda x})}{e^{-\lambda x}(1-\theta)+\theta} \right)^{\alpha b - \alpha p} \\
&\times \left(1 - \left(\frac{e^{-\lambda x}(1-\theta)+\theta}{\theta(1-e^{-\lambda x})} \right)^\alpha \right)^l = \sum_{b=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \sum_{t=0}^\infty \frac{(-p)_b (p)_l}{b! l!} \\
&\times \frac{(-l)_m (\alpha b - \alpha p + \alpha m - \alpha l)_n (1)_t}{m! n! t!} (-1)^l \frac{\theta^{p+l+\alpha b - \alpha p + \alpha m - \alpha l}}{\lambda^r} \\
&\times (1-\theta)^n \int_0^1 u^{\alpha b - \alpha p + \alpha m + \alpha l + n + t} (-\ln(1-u))^{r-1} du \\
&= \sum_{b=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \sum_{t=0}^\infty \sum_{a=0}^\infty \frac{(-p)_b (p)_l (-l)_m (\alpha b - \alpha p + \alpha m - \alpha l)_n}{b! l! m! n!} \\
&\times \frac{(1)_t (-1)^l (1-\theta)^n \theta^{p+l+\alpha b - \alpha p + \alpha m - \alpha l} (-\alpha b + \alpha p - \alpha m - \alpha l - n - t)_a}{t! \lambda^r a!} \\
&\times \int_0^\infty v^{r-1} e^{-v(1+a)} dv = \sum_{b=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \sum_{t=0}^\infty \sum_{a=0}^\infty \frac{(-p)_b (p)_l}{b! l!} \\
&\times \frac{(-l)_m (\alpha b - \alpha p + \alpha m - \alpha l)_n (1)_t}{m! n! t!} (-1)^l \frac{\theta^{p+l+\alpha b - \alpha p + \alpha m - \alpha l}}{\lambda^r} \\
&\times (1-\theta)^n \frac{(-\alpha(b+p-m-l) - n - t)_a}{a!} \frac{\Gamma(r)}{(1+a)^r}.
\end{aligned} \tag{5.11}$$

Substituindo (5.11) em (5.10) termina a prova. \square

A média residual de tempos de vida com função de sobrevivência $S(x)$ é dada por:

$$\mu(t) = E(X - t | X > t) = \frac{1}{S(t)} \int_t^\infty S(x) dx. \tag{5.12}$$

Para a variável aleatória Y com distribuição ECEGN, temos que a média residual é dada por:

$$\begin{aligned}
\mu(t) &= \frac{\theta + (1-\theta) \left(\frac{\theta(1-e^{-\lambda x})}{\theta+(1-\theta)e^{-\lambda x}} \right)^\alpha}{\theta \left(1 - \left(\frac{\theta(1-e^{-\lambda x})}{e^{-\lambda x}(1-\theta)+\theta} \right)^\alpha \right)} \\
&\times \sum_{k=0}^\infty \sum_{l=0}^\infty \sum_{m=0}^\infty \sum_{n=0}^\infty \sum_{t=0}^\infty \frac{(-1)_k (1)_l (-l)_m (1)_t (\alpha k + \alpha m - \alpha l - \alpha)_n}{k! l! m! t! n!} \\
&\times \frac{\theta^{l+1+\alpha k + \alpha m - \alpha l - \alpha}}{\lambda} (1-\theta)^n \left(\frac{1 - (1-e^{-\lambda x})^{\alpha k + \alpha m - \alpha l - \alpha + t + n + 1}}{\alpha k + \alpha m - \alpha l - \alpha + t + n + 1} \right).
\end{aligned}$$

Demonstração. Temos que:

$$\begin{aligned}
\frac{1}{S(t)} \int_t^\infty S(x) dx &= \frac{\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha}{\theta \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda x}(1 - \theta) + \theta} \right) \right)} \sum_{k=0}^{\infty} \frac{(-1)_k}{k!} \int_y^\infty \theta \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda x}(1 - \theta) + \theta} \right)^{\alpha k} \\
&\times \left(\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha \right)^{-1} dx \\
&= \frac{\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha}{\theta \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda x}(1 - \theta) + \theta} \right) \right)} \\
&\times \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{t=0}^{\infty} \frac{(-1)_k (1)_l (-l)_m (1)_t (\alpha k + \alpha m - \alpha l - \alpha)_n}{k! l! m! t! n!} \\
&\times \frac{\theta^{l+1+\alpha k+\alpha m-\alpha l-\alpha}}{\lambda} (1 - \theta)^n \int_{1-e^{-\lambda y}}^1 u^{\alpha k+\alpha m-\alpha l-\alpha+t+n} (-1)^l du \\
&= \frac{\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda x})}{\theta + (1 - \theta)e^{-\lambda x}} \right)^\alpha}{\theta \left(1 - \left(\frac{\theta(1 - e^{-\lambda x})}{e^{-\lambda x}(1 - \theta) + \theta} \right) \right)} \\
&\times \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{t=0}^{\infty} \frac{(-1)_k (1)_l (-l)_m (1)_t (\alpha k + \alpha m - \alpha l - \alpha)_n}{k! l! m! t! n!} \\
&\times \frac{\theta^{l+1+\alpha k+\alpha m-\alpha l-\alpha}}{\lambda} (1 - \theta)^n \left(\frac{1 - (1 - e^{-\lambda x})^{\alpha k+\alpha m-\alpha l-\alpha+t+n+1}}{\alpha k + \alpha m - \alpha l - \alpha + t + n + 1} \right). \tag{5.13}
\end{aligned}$$

Utilizando a expressão (1) e considerando $u = 1 - e^{-\lambda x}$ o resultado é encontrado substituindo (5.13) em (5.12). □

5.3 Inferência

A função logaritmo da verossimilhança é dada por:

$$\begin{aligned}
\ell(\vartheta, \underline{x}, \underline{\delta}) &= n \log(\theta) + \sum_{i=1}^n \delta_i \log(\alpha \lambda \theta^\alpha) + \sum_{i=1}^n (1 - \delta_i) \log \left(1 - \left(\frac{\theta(1 - e^{-\lambda y_i})}{e^{-\lambda y_i}(1 - \theta) + \theta} \right)^\alpha \right) \\
&- \sum_{i=1}^n (1 + \delta_i) \log \left(\theta + (1 - \theta) \left(\frac{\theta(1 - e^{-\lambda y_i})}{\theta + (1 - \theta)e^{-\lambda y_i}} \right)^\alpha \right) - \lambda \sum_{i=1}^n \delta_i y_i \\
&+ \sum_{i=1}^n \delta_i \log \left(\frac{(1 - e^{-\lambda y_i})^{\alpha-1}}{(e^{-\lambda y_i}(1 - \theta) + \theta)^{\alpha+1}} \right), \tag{5.14}
\end{aligned}$$

em que $\vartheta = (\alpha, \lambda, \theta)$, δ_i é a variável indicadora de censura definida em 1.1.

As EMV de α , λ e θ são soluções das seguintes equações:

$$\begin{aligned}
& \frac{\sum_{i=1}^n (1-\delta_i) X_i^\alpha \ln(X_i)}{1-X_i^\alpha} + \frac{\sum_{i=1}^n (1-\delta_i)(1-\theta) X_i^\alpha \ln(X_i)}{\theta+(1-\theta)X_i^\alpha} = \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\alpha}{\theta} \right) + \sum_{i=1}^n \ln \left(\frac{X_i}{\theta} \right), \\
& \frac{\sum_{i=1}^n (1-\delta_i) \alpha X_i^{\alpha-1} \theta y_i Z_i}{1-X_i^\alpha} (1+X_i(1-\theta)) + \frac{\sum_{i=1}^n (1+\delta_i)(1-\theta) \alpha y_i Z_i X_i^{\alpha-1}}{\theta+(1-\theta)X_i^\alpha} (\theta-X_i) \\
& + \sum_{i=1}^n \delta_i y_i = \frac{\sum_{i=1}^n \delta_i}{\lambda} + \sum_{i=1}^n y_i ((\alpha-1)K_i + (\alpha+1)(1-\theta)Z_i), \\
& \frac{n}{\theta} + \frac{\sum_{i=1}^n \delta_i}{\lambda \theta} + \frac{\sum_{i=1}^n (1-\delta_i) X_i (1-X_i)}{(1-X_i^\alpha)\theta} = \frac{\sum_{i=1}^n (1+\delta_i)}{\theta+(1-\theta)X_i^\alpha} \left(1-X_i^\alpha + \frac{\alpha(1-\theta)X_i^\alpha(1-X_i)}{\theta^2+(1-\theta)\theta X_i^\alpha} \right) \\
& + \sum_{i=1}^n (\alpha+1)A_i.
\end{aligned}$$

em que $X_i = \frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$, $Z_i = \frac{e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta)+\theta}$, $K_i = \frac{e^{-\lambda y_i}}{1-e^{-\lambda y_i}}$ e $A_i = \frac{(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$. A matriz de informação de Fisher para $(\alpha, \lambda, \theta)$ é dada por:

$$I_F(\alpha, \theta, \lambda) = - \left(\begin{array}{ccc} I_{\alpha,\alpha} & I_{\alpha,\theta} & I_{\alpha,\lambda} \\ I_{\theta,\alpha} & I_{\theta,\theta} & I_{\theta,\lambda} \\ I_{\lambda,\alpha} & I_{\lambda,\theta} & I_{\lambda,\lambda} \end{array} \right) \Bigg|_{(\alpha,\theta,\lambda)=(\hat{\alpha},\hat{\theta},\hat{\lambda})}, \quad (5.15)$$

em que os elementos da matriz $I_F(\alpha, \lambda, \theta)$ são dados no apêndice A.

5.4 Estudo de simulação

5.4.1 Estudo das propriedades dos estimadores por M.V dos parâmetros

O estudo se baseia na geração de 1000 conjuntos de dados com distribuição ECEGN para diferentes valores de α , λ e θ para $n = 30, 50, 100, 200$ e 500 . Para o parâmetro θ a seguinte transformação foi considerada: $\theta = \frac{e^{\theta^*}}{(1+e^{\theta^*})}$, em que $\theta^* \in (0, 1)$ e para λ e α se considerou a transformação exponencial. Para o calculo das variâncias foi considerado o método Delta. Os resultados estão na Tabela (5.1), no qual estão as médias das estimativas de máxima verossimilhança ($Av(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$), junto com as probabilidades de cobertura com intervalo de confiança de 95% $ECEGN[C(\hat{\alpha}, \hat{\lambda}, \hat{\theta})]$ e calculado as variâncias e encontrado o EQM. $[EQM((\hat{\alpha}), (\hat{\lambda}), (\hat{\theta}))]$.

Na Tabela 5.1, observa-se que o EQM para todos os parâmetros diminuem à medida que se aumenta o tamanho amostral. A probabilidade de cobertura para α está melhor do que λ e θ pois a menor probabilidade de cobertura foi de 0,810 para o cenário de $\alpha = 5$, $\lambda = 2$, $\theta = 0,9$. Apenas para θ a probabilidade de cobertura está abaixo de 50% em alguns cenários.

Tabela 5.1: Média das EMV, probabilidade de cobertura e EQM.

| n | $(\alpha, \lambda, \theta)$ | $Av(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$ | $C(\alpha, \lambda, \theta)$ | $EQM(\hat{\alpha})$ | $EQM(\hat{\lambda})$ | $EQM(\hat{\theta})$ |
|-----|-----------------------------|---|------------------------------|---------------------|----------------------|---------------------|
| 30 | (0,500;1,500;0,700) | (0,605;1,426;0,500) | (0,902;0,926;0,409) | 0,033 | 0,254 | 0,208 |
| | (0,700;0,500;0,600) | (0,795;0,472;0,460) | (0,864;0,913;0,454) | 0,036 | 0,029 | 0,197 |
| | (2,000;2,000;0,800) | (2,123;2,041;0,707) | (0,885;0,926;0,422) | 0,733 | 0,251 | 0,164 |
| | (5,000;2,000;0,900) | (4,729;1,949;0,685) | (0,810;0,939;0,554) | 6,937 | 0,147 | 0,188 |
| 50 | (0,500;1,500;0,700) | (0,576;1,380;0,567) | (0,868;0,915;0,441) | 0,022 | 0,138 | 0,177 |
| | (0,700;0,500;0,600) | (0,776;0,446;0,482) | (0,872;0,892;0,463) | 0,024 | 0,015 | 0,193 |
| | (2,000;2,000;0,800) | (1,943;1,957;0,642) | (0,825;0,921;0,438) | 0,347 | 0,122 | 0,192 |
| | (5,000;2,000;0,900) | (4,517;1,890;0,685) | (0,815;0,885;0,545) | 4,696 | 0,105 | 0,180 |
| 100 | (0,500;1,500;0,700) | (0,555;1,334;0,592) | (0,882;0,842;0,505) | 0,013 | 0,091 | 0,150 |
| | (0,700;0,500;0,600) | (0,754;0,446;0,525) | (0,868;0,890;0,444) | 0,014 | 0,008 | 0,180 |
| | (2,000;2,000;0,800) | (1,890;1,891;0,662) | (0,829;0,896;0,467) | 0,196 | 0,070 | 0,169 |
| | (5,000;2,000;0,900) | (4,253;1,821;0,708) | (0,848;0,840;0,574) | 2,392 | 0,068 | 0,146 |
| 200 | (0,500;1,500;0,700) | (0,531;1,324;0,666) | (0,896;0,772;0,519) | 0,006 | 0,060 | 0,110 |
| | (0,700;0,500;0,600) | (0,738;0,445;0,557) | (0,883;0,811;0,495) | 0,008 | 0,005 | 0,158 |
| | (2,000;2,000;0,800) | (1,896;1,856;0,681) | (0,848;0,835;0,483) | 0,125 | 0,046 | 0,145 |
| | (5,000;2,000;0,900) | (4,366;1,828;0,747) | (0,855;0,765;0,576) | 1,679 | 0,054 | 0,114 |
| 500 | (0,500;1,500;0,700) | (0,517;1,324;0,704) | (0,918;0,610;0,649) | 0,002 | 0,042 | 0,071 |
| | (0,700;0,500;0,600) | (0,722;0,443;0,590) | (0,893;0,593;0,569) | 0,004 | 0,004 | 0,127 |
| | (2,000;2,000;0,800) | (1,910;1,832;0,727) | (0,880;0,640;0,554) | 0,063 | 0,038 | 0,098 |
| | (5,000;2,000;0,900) | (4,336;1,814;0,760) | (0,856;0,663;0,654) | 1,165 | 0,047 | 0,081 |

5.5 Aplicação

5.5.1 Ajuste a conjunto de dados reais da distribuição ECEGN e outras distribuições

Nesta seção o ajuste da distribuição ECEGN é comparada com outras distribuições em dois conjuntos de dados extraídos da literatura. O primeiro conjunto de dados, Kalbfleisch & Prentice (1980) consiste em tempos de vida do diagnóstico até a morte de pacientes com câncer orofaríngeo que pode afetar a base da língua, amígdalas, palato e faringe. O segundo conjunto de dados, Lawless (2003) consiste em milhas de 96 locomotivas fizeram antes de quebrarem. As locomotivas que quebraram depois de 135 milhas foram consideradas censuradas.

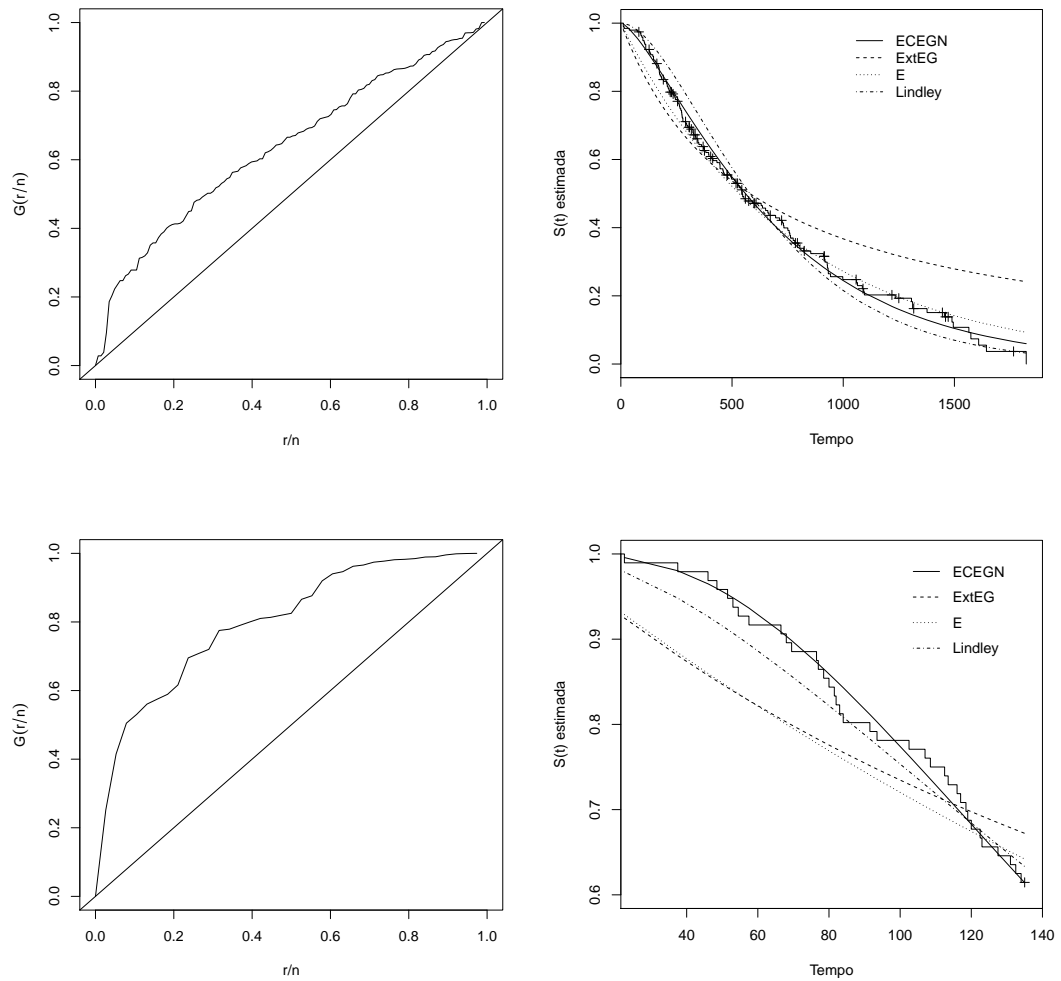


Figura 5.2: TTT plot (painel esquerdo) e Kaplan Meier com função de sobrevivência ajustada para cada distribuição (painel direito).

A Figura 5.2 mostra o TTT plot e o ajuste das distribuições sobrepostas a curva de sobrevivência Kaplan Meier. Os dois conjuntos de dados possuem taxa de risco crescente e possuem um bom ajuste da distribuição ECEGN comparadas as outras distribuições.

Tabela 5.2: Ajuste de quatro distribuições ao conjunto de dados T1: estimativas dos parâmetros dos modelos ECEGN, ExtEG, E, e Lindley

| Parâmetros | ECEGN | | ExtEG | | E | | Lindley | |
|-------------------------|------------------|--------------------|------------------------------------|-----------------------|------------------|--------------------|------------------|--------------------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 1,479 | 0,168 | - | - | - | - | - | - |
| λ | 0,001 | 1×10^{-4} | $2,3 \times 10^{-13}$ | $2,3 \times 10^{-10}$ | 0,001 | 1×10^{-4} | 0,002 | 1×10^{-4} |
| θ | 0,992 | 0,216 | $1,3 \times 10^{-10}$ | $1,3 \times 10^{-7}$ | - | - | - | - |
| $-\log(L(\vartheta_g))$ | 1079,308 | | 1106,125 | | 1085,318 | | 1084,387 | |
| AIC | 2164,617 | | 2216,249 | | 2172,636 | | 2172,774 | |
| BIC | 2174,436 | | 2222,795 | | 2175,909 | | 2179,320 | |
| Estatística KS | 0,059 (0,491) | | 0,241 ($2,7 \times 10^{-10}$) | | 0,092 (0,068) | | 0,094 (0,061) | |

Tabela 5.3: Ajuste das quatro distribuições ao conjunto de dados T2: estimativas dos parâmetros dos modelos ECEGN, ExtEG, E, e Lindley

| Parâmetros | ECEGN | | ExtEG | | E | | Lindley | |
|-------------------------|------------------|-------|----------------------|----------------------|------------------|--------------------|------------------|--------------------|
| | EMV | D.P. | EMV | D.P. | EMV | D.P. | EMV | D.P. |
| α | 3,550 | 0,897 | - | - | - | - | - | - |
| λ | 0,010 | 0,002 | $1,3 \times 10^{-9}$ | $7,2 \times 10^{-8}$ | 0,003 | 5×10^{-4} | 0,009 | 9×10^{-4} |
| θ | 0,999 | 0,002 | $3,6 \times 10^{-7}$ | 2×10^{-5} | - | - | - | - |
| $-\log(L(\vartheta_g))$ | 237,077 | | 251,958 | | 248,610 | | 239,516 | |
| AIC | 480,154 | | 507,917 | | 499,221 | | 483,033 | |
| BIC | 487,847 | | 513,045 | | 501,785 | | 488,162 | |
| Estatística KS | 0,041 (0,996) | | 0,111 (0,186) | | 0,108 (0,205) | | 0,042 (0,995) | |

As Tabelas 5.2 e 5.3 apresentam as EMV para os parâmetros de cada distribuição estudada, os critérios e o teste de Kolmogorov Smirnov para cada distribuição ajustada. No qual a ECEGN possui os menores critérios e o maior p -valor do teste de Kolmogorov Smirnov para os conjuntos de dados T1 e T2.

5.6 Comentários Finais

Neste capítulo uma nova distribuição capaz de modelar função de falha crescente ou decrescente é apresentada. Denotada por ECEGN é baseada em um contexto de N sistemas em que cada um possui α componentes no qual a variável estudada X verifica o primeiro sistema a falhar. Propriedades são exploradas e uma análise em conjunto de dados reais mostra que a distribuição ECEGN se ajusta melhor em relação a extensão da distribuição Exponencial Geométrica, Exponencial e Lindley.

Capítulo 6

Conclusão

Neste trabalho foram criadas e apresentadas quatro distribuições provindas da distribuição EG. Apresentamos a distribuição ECEG, nova distribuição provinda da CEG, com um parâmetro novo α que generaliza a situação de n (inteiro) componentes num sistema. Propriedades dessa distribuição foram explorados assim como 1000 simulações para diferentes tamanhos amostrais para verificar o vício e o EQM. Distribuição que é capaz de modelar taxa de risco crescente, decrescente e em forma de banheira. Foi explorada a situação em que covariáveis influenciavam a ocorrência do evento assumidas no parâmetro de escala. Uma análise dos resíduos global foi realizada utilizando o resíduos de Cox Snell, no qual considera como modelo bem ajustado quando os resíduos seguem distribuição exponencial padrão. Em todas as análises com covariáveis foi verificado que a nova distribuição se ajusta melhor em relação as particulares (CEG e E), pois além dos resíduos seguirem distribuição exponencial padrão a curva de sobrevivência ajustada dos resíduos se aproximava mais perto da curva de sobrevivência dos resíduos Kaplan Meier.

Foi apresentado o modelo de mistura padrão para a distribuição ECEG, chamado de LECEG. Distribuição capaz de modelar taxa de risco decrescente e unimodal. A versalidade dessa nova distribuição consiste em considerar parte da população imune a ocorrência do evento. O mesmo estudo de simulação foi realizado para verificar o vício e o EQM dos parâmetros. Assim como considerar covariáveis no parâmetro de escala e de longa duração. Verificou-se pela análise de resíduos que o modelo de regressão LECEG se ajusta melhor em relação as particulares (LCEG e LE) pois os resíduos de Cox Snell seguiam distribuição Exponencial Padrão e possuíam a curva de sobrevivência ajustada mais perto da curva de sobrevivência dos resíduos de Kaplan Meier.

Em seguida uma generalização da EG chamada de EEG com um parâmetro de

escala λ e dois de forma α e θ , em que α generaliza a situação de n (inteiro) componentes num sistema, uma distribuição capaz de modelar taxa de risco crescente, decrescente e unimodal. Propriedades foram explorados assim como 1000 simulações para diferentes tamanhos amostrais para verificar o vício e o EQM das EMV dos parâmetros.

Por fim a distribuição ECEGN foi mostrada, baseada em um contexto com mais de um sistema, no qual cada sistema possui α componentes. Essa distribuição se mostrou melhor do que outras no ajuste em um conjunto de dados reais.

Extensões deste trabalho seria trabalhar com as distribuições EEG e ECEGN incluindo covariáveis no estudo. Assim como trabalhar todos esses modelos apresentados do ponto de vista bayesiano. Abordar o caso de amostras com população imune para a distribuição EEG.

Referências Bibliográficas

- Adamidis, K. & Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics Probability Letters*, **39**, 35–42.
- Adamidis, K., Dimitrakopoulou, T. & Loukas, S. (2005). On an extension of the exponential-geometric distribution. *Statistics Probability Letters*, **73**, 259–269.
- Allison, P. (1995). *Survival Analysis Using SAS a practical guide*. SAS.
- Anscombe, F. J. (1961). Estimating a mixed-exponential response law. *Journal of American Statistical Association*, **56**, 493–502.
- Bakouch, H. S., Al-Zahrani, B. M., Al-Shomrani, A. A., Marchi, V. A. A. & Louzada, F. (2011). An extended lindley distribution. *Journal of the Korean Statistical Society*, **41**, 75–85.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*. John Wiley & Sons.
- Barriga, G. D., Louzada-Neto, F. & Cancho, V. G. (2011). The complementary exponential power lifetime model. *Computational Statistics and Data Analysis*, **54**(5), 1250–1259.
- Basu, A. & Klein, J. (1982). Some recent development in competing risks theory. *Survival Analysis*, **1**(1), 216–229.
- Berkson, J. & Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501–515.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society*, **B 11**, 15–44.

- Cancho, V. G., Ortega, E. M. M. & Bolfarine, H. (2009). The log-exponentiated-weibull regression models with cure rate: local influence and residual analysis. *Journal of Data Science*, **7**, 433–458.
- Cancho, V. G., Louzada-Neto, F. & Barriga, G. D. C. (2011). The poisson-exponential lifetime distribution. *Computational Statistics & Data Analysis*, **55**(1), 677–686.
- Chahkandi, M. & Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics & Data Analysis*, **53**(1), 4433–4440.
- Chen, M., Ibrahim, J. & Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**(447), 909–919.
- Colosimo, E. A. & Giolo, S. R. (2006). *Análise de Sobrevida Aplicada*. Editora Edgard Blucher.
- Cox, D. R. & Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society*, **30**, 248–275.
- Crowder, M., Kimber, A., Smith, R. & Sweeting, T. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall.
- Crowley, J. & Hu, M. (1977). Covariance analysis of heart transplant data. *Journal of American Statistic Association*, **72**, 27–36.
- Draper, N. & Smith, H. (1998). *Applied Regression Analysis*. John Wiley and Sons.
- Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrics*, **64**, 43–46.
- Ghitany, M. E., Maller, R. A. & Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of Multivariate Analysis*, **49**, 218–241.
- Glaser, R. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, **75**, 667–672.
- Gulland, J. A. (1955). On the estimation of population parameters from masked members. *Biometrika*, **42**, 269–270.

- Gupta, R. D. & Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, **41**(2), 173–188.
- Hemmati, F., Khorram, E. & Rezakhan, S. (2011). A new three-parameter ageing distribution. *Journal of Statistical Planning & Inference*, **141**(1), 2266–2275.
- Hirotagu, A. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.
- Hosmer, D. & Lemeshow, S. (1999). *Applied Survival Analysis*. John Wiley and Sons.
- Hosmer, D. W. & Lemeshow, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition*. John Wiley and Sons Inc.
- Kalbfleisch, J. & Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. Wiley & Sons.
- Karlis, D. (2003). An EM algorithm for multivariate poisson distribution and related models. *Journal of Applied Statistics*, **30**, 63–77.
- Kus, C. (2007). A new lifetime distribution. *Computational Statistics Data Analysis*, **11**, 4497–4509.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, second edition.
- Louzada, F., Roman, M. & Cancho, V. G. (2011). The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart. *Computational Statistics & Data Analysis*, **55**(8), 2516–2524.
- Louzada-Neto (1999). Poly-hazard regression models for lifetime data. *Biometrics*, **55**, 1121–1125.
- Maller, R. A. & Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, **51**, 1197–1205.
- Marshall, A. & Olkin, I. (2007). *Life distributions*. Springer.
- Mudholkar, G., Srivastava, D. & Freimer, M. (1995). The exponentiated weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, **37**(4), 436–445.

- Nadarajah, S. & Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97–111.
- Ng, H. K. T., Chan, P. S. & Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using em algorithm. *Computational Statistical Data Analysis*, **39**, 371–386.
- Perdoná, G. C. & Louzada-Neto, F. (2011). A general hazard model for lifetime data in the presence of cure rate. *Journal of Applied Statistics*, **38**, 1395–1405.
- Perperoglou, A., Keramopoulos, A. & van Houwelingen, H. C. (2007). Approaches in modelling long-term survival: an application to breast cancer. *Statistics in Medicine*, **26**, 2666–2685.
- Pons, O. & Lemdani, M. (1969). On waiting time distributions. *Acta Sociologica*, **12**, 132–143.
- Pons, O. & Lemdani, M. (2003). Estimation and test in long-term survival mixture models. *Computational Statistics and Data Analysis*, **41**, 465–479.
- Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variable. *Biometrika*, **60**, 279–288.
- Ristic, M. M. & Nadarajah, S. (2012). A new lifetime distribution. *Journal of Statistical Computation and Simulation*, pages 1–16.
- Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009). On the unification of long-term survival models. *Statistics & Probability Letters*, **79**(6), 753–759.
- Roman, M., Louzada, F., Cancho, V. G. & Leite, J. G. (2012). A new long-term survival distribution for cancer data. *Journal of Data Science*, **10**, 241–258.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**(2), 461–464.
- Smith, R. M. & Bain, L. J. (1975). An exponential power life-testing distribution. *Communications in Statistics, Theory and Methods*, **4**, 469–481.

- Stablein, D. & Koutrouvelis, I. (1985). A two sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics*, **41**, 643–652.
- Tahmasbi, R. & Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistic Data Analysis*, **52**(8), 3889–3901.
- Team, R. D. C. (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.
- Yakovlev, A., Tsodikov, A. & Asselain, B. (1996). Stochastic models of tumor latency and their biostatistical applications. *World Scientific Pub Co Inc*.

Apêndice A

Expansão em série exponencial usada

$$(1-x)^{-s} = \sum_{r=0}^{\infty} \frac{(s)_r}{r!} x^r, \quad (1)$$

em que $(r)_k$ é o símbolo Pochhammer, dado por $(r)_k = r(r+1)\dots(r+k-1)$, se $|x| < 1$ a série converge, $(-r)_k = (-1)^k(r-k+1)_k$.

A seguinte equação é utilizada:

$$\int_0^1 z^{p-1}(1-z)^{n-1}(1+bz^m)^l dz = \Gamma(n) \sum_{k=0}^{\infty} \binom{l}{k} \frac{b^k \Gamma(p+km)}{\Gamma(p+n+km)}, \quad (2)$$

em que $\Gamma(a, b)$ é a função gama incompleta dada por $\Gamma(a, b) = \int_b^{\infty} t^{a-1} e^{-t} dt$.

Os elementos da matriz de informação de Fisher (2.23) apresentados no Capítulo 2 são dados por:

$$\begin{aligned} I_{\alpha, \alpha} &= \sum_{i=1}^n (1 + \delta_i) \left(\frac{X_i^{2\alpha} (\ln(X_i))^2}{(1 - X_i^\alpha)^2} + \frac{X_i^\alpha (\ln(X_i))^2}{1 - X_i^\alpha} \right) + \sum_{i=1}^n \left(\left(-\frac{1}{\alpha^2} + \frac{\ln(X_i)}{\alpha} \right) \delta_i - \frac{\delta_i}{\alpha} (2 + (\ln(X_i))^2) \right); \\ I_{\lambda, \lambda} &= \sum_{i=1}^n (1 - \delta_i) \left(\frac{-\alpha X_i^{-2+\alpha} \theta^2 y_i^2 Z_i^2 (A_i(1-\theta) + 1)^2}{1 - X_i^\alpha} \left(\frac{\alpha X_i^\alpha}{1 - X_i^\alpha} + 2 - \alpha \right) - \frac{\alpha X_i^{-1+\alpha}}{1 - X_i^\alpha} \right. \\ &\times \left(2y_i^2 Z_i (1-\theta) \theta \left(\frac{(1-\theta)A_i^3}{Z_i(1-e^{-\lambda y_i})^2} + Z_i - \frac{A_i}{2} - \frac{1}{(1-\theta)2} \right) \right) + \sum_{i=1}^n \left(\delta_i \left(\frac{1}{\lambda} + \frac{2(1-\theta)y_i Z_i}{e^{-\lambda y_i}} \right. \right. \\ &- y_i + (-1 + \alpha)y_i(Z_i(1-\theta) + K_i) \left(-\frac{1}{\lambda} + y_i \right) - \frac{2(1-\theta)\delta_i y_i Z_i \alpha \theta \lambda}{\alpha^2 \theta^2 \lambda^2} \left(\frac{1}{\lambda} + 2 \frac{(1-\theta)y_i Z_i e^{-\lambda y_i}}{K_i} - \frac{e^{-\lambda y_i} y_i}{Z_i} \right. \\ &+ e^{-\lambda y_i} (-1 + \alpha) \left((1-\theta)y_i + \frac{K_i y_i}{Z_i} \right) + \frac{(1-\alpha)\delta_i \theta y_i e^{-\lambda y_i}}{X_i^\alpha \alpha^2 \theta^2 \lambda^2} \left(\frac{1-\theta}{K_i} + \frac{1}{Z_i} \right) \left(\frac{Z_i \alpha X_i^{-1+\alpha}}{e^{-\lambda y_i} (e^{-\lambda y_i} (1-\theta) + \theta) y_i} \right. \\ &\times \left(\frac{e^{\lambda y_i}}{y_i \theta} + \frac{2(1-\theta)}{\theta e^{-\lambda y_i}} - \frac{\lambda}{\theta e^{-\lambda y_i}} + \frac{(-1+\alpha)\lambda}{\theta y_i e^{-\lambda y_i}} (Z_i(1-\theta)y_i + y_i K_i) \right) + \frac{c_i}{\alpha^2 \theta^2 \lambda^2} \left(4\alpha(1-\theta) - \frac{2\alpha}{Z_i} + 6\alpha \lambda y_i Z_i \right. \\ &- \frac{6\alpha(1-\theta)\lambda}{\theta} + \frac{\alpha \theta \lambda y_i}{Z_i \theta} + 2(-1+\alpha)\alpha \theta \left(\frac{1-\theta}{\theta} + \frac{K_i}{Z_i \theta} \right) + 4(-1+\alpha)\alpha(1-\theta)\lambda (Z_i(1-\theta) + K_i) y_i - 2(-1+\alpha)\alpha \theta \lambda \\ &\times \left(\frac{(1-\theta)y_i}{\theta} + \frac{K_i}{Z_i \theta} \right) + e^{-\lambda y_i} (-2+\alpha)(-1+\alpha) \frac{\lambda y_i}{Z_i} (Z_i(1-\theta) + K_i)^2 + (-1+\alpha)\alpha \lambda \\ &\times \left. \left(2(1-\theta)^2 y_i Z_i + 2(1-\theta)y_i K_i - \frac{(1-\theta)y_i}{1-e^{-\lambda y_i}} - \frac{y_i K_i}{Z_i \theta} \right) y_i Z_i \theta \right); \end{aligned}$$

$$\begin{aligned}
I_{\theta,\theta} &= \sum_{i=1}^n (1 + \delta_i) \left(\frac{-\alpha X_i^{-1+\alpha} 2A_i^2 (A_i\theta - 1)}{1 - X_i^\alpha} - \frac{\alpha^2 X_i^{-2+2\alpha} A_i^2 (-\theta A_i + 1)^2}{(1 - X_i^\alpha)^2} - \frac{(-1 + \alpha)\alpha X_i^{-2+\alpha} A_i^2 (-A_i\theta + 1)^2}{1 - X_i^\alpha} \right) \\
&+ \sum_{i=1}^n \left(-4A_i^2 \delta_i + \frac{A_i \delta_i}{\theta \lambda} + \frac{\delta_i(-1 + \alpha)}{\theta} \left(-\frac{A_i^2 \theta}{1 - e^{-\lambda y_i}} + A_i \right) + \frac{2Z_i \delta_i}{\theta K_i} - \alpha \lambda \delta_i - \frac{-1 + \alpha}{\theta} \delta_i \left(-\frac{Z_i}{K_i} + \frac{1}{\theta} \right) \right) \\
&+ \frac{e^{\lambda y_i} (1 - \alpha)}{\alpha \theta \lambda} \left(-\frac{e^{-\lambda y_i} \theta}{K_i} + \frac{e^{-\lambda y_i}}{Z_i} \right) \frac{Z_i}{\theta e^{-\lambda y_i}} \left(\frac{-2Z_i \alpha \theta \lambda}{K_i e^{-\lambda y_i}} + e^{-\lambda y_i} \alpha \lambda + e^{-\lambda y_i} (-1 + \alpha) \alpha \theta \left(-\lambda + \frac{K_i \lambda}{Z_i \theta^2} \right) \right) \delta_i \\
&+ \left(6A_i^2 - \frac{4Z_i}{\theta K_i} + (-1 + \alpha) \alpha \theta^2 2A_i^3 (\theta A_i - 1) \lambda - 4(-1 + \alpha) \alpha \theta^2 \left(\frac{Z_i^2}{K_i^4} e^{-2\lambda y_i} \theta + A_i^3 (e^{-\lambda y_i} (1 - \theta) + \theta)^2 \right) \lambda \right. \\
&+ \left. 2(-1 + \alpha) \alpha \left(-A_i + \frac{1}{\theta} \right) \lambda + (-2 + \alpha) (-1 + \alpha) \alpha \theta \left(\frac{K_i}{Z_i \theta} A_i^2 (-A_i \theta + 1)^2 \lambda \right) \right) \delta_i, \\
I_{\alpha,\theta} = I_{\theta,\alpha} &= -\sum_{i=1}^n (1 - \delta_i) \left(\frac{\alpha X_i^{\alpha-1} (A_i(1 - X_i)) \ln(X_i)}{(1 - X_i^\alpha)^2} + \frac{X_i^{\alpha-1} (A_i(1 - X_i))}{(1 - X_i^\alpha)^2} \right) (1 - X_i^\alpha) + \sum_{i=1}^n \delta_i \left(\frac{A_i(1 - X_i)}{X_i} \right) \\
&+ \frac{\sum_{i=1}^n (1 - \delta_i) X_i^\alpha \ln(X_i) (-\alpha X_i^{\alpha-1} (A_i(1 - X_i)))}{(1 - X_i^\alpha)^2}, \\
I_{\lambda,\theta} = I_{\theta,\lambda} &= -\sum_{i=1}^n (1 - \delta_i) \alpha \left((-1 + \alpha) X_i^\alpha \left(\frac{X_i^{-1}}{\theta} - \theta \right) Z_i \theta y_i (A_i(1 - \theta) + 1) + X_i^{-1+\alpha} y_i Z_i (1 - X_i) (A_i(1 - \theta) + 1) \right. \\
&+ \left. X_i^{-1+\alpha} \theta y_i Z_i A_i (-1 - (1 - \theta) A_i) \right) + \sum_{i=1}^n \delta_i (2y_i (-1 - (1 - \theta) A_i) Z_i + (-1 + \alpha) K_i y_i A_i (-1 - (1 - \theta) A_i)) \\
I_{\alpha,\lambda} = I_{\lambda,\alpha} &= -\sum_{i=1}^n (1 - \delta_i) \left(\frac{\alpha X_i^{\alpha-1} \theta y_i Z_i \left(1 + \frac{(1-\theta)\theta Z_i}{K_i} \right) \ln\left(\frac{\theta Z_i}{K_i}\right)}{(1 - X_i^\alpha)^2} + \frac{X_i^{\alpha-1} Z_i \left(y_i + \frac{Z_i}{K_i} (1 - \theta) \right) \theta}{1 - X_i^\alpha} \right. \\
&+ \left. \sum_{i=1}^n (1 - \delta_i) X_i^\alpha \ln(X_i) \left(-\alpha X_i^{\alpha-1} \theta y_i Z_i \left(e^{\lambda y_i} + \frac{Z_i(1 - \theta)}{K_i e^{-\lambda y_i}} \right) (1 - X_i^\alpha)^{-2} + \sum_{i=1}^n \delta_i \left(K_i y_i \left(1 + \frac{(1 - \theta) Z_i}{K_i} \right) \right) \right) \right),
\end{aligned}$$

em que $X_i = \frac{\theta(1 - e^{-\lambda y_i})}{e^{-\lambda y_i}(1 - \theta) + \theta}$, $Z_i = \frac{e^{-\lambda y_i}}{e^{-\lambda y_i}(1 - \theta) + \theta}$, $K_i = \frac{e^{-\lambda y_i}}{1 - e^{-\lambda y_i}}$ e $A_i = \frac{(1 - e^{-\lambda y_i})}{e^{-\lambda y_i}(1 - \theta) + \theta}$.

Os elementos da matriz de informação de Fisher (3.15) do Capítulo 3 são dados por:

$$\begin{aligned}
I_{\alpha,\alpha} &= \sum_{i=1}^n \left(-\frac{(1-p)X_i^\alpha (\ln(X_i))^2 (1 - \delta_i)}{(1 - (1-p)X_i^\alpha)} \left(\frac{(1-p)X_i^\alpha}{1 - (1-p)X_i^\alpha} - 1 \right) \right) \\
&+ \left(\sum_{i=1}^n \left(-\frac{1}{\alpha} + \ln(\theta) \right) c_i - \frac{\ln(\theta)}{\alpha} (1 + \ln(\theta)) \delta_i + \frac{2\ln(\theta)\delta_i}{\alpha} + (\ln(\theta))^2 \delta_i \right); \\
I_{\lambda,\lambda} &= -\frac{\sum_{i=1}^n \delta_i}{\lambda^2} - \sum_{i=1}^n K_i (-1 + \alpha) \delta_i y_i^2 (K_i + 1) - \sum_{i=1}^n Z_i (1 + \alpha) (1 - \theta) \delta_i y_i^2 (-Z_i (1 - \theta) + 1) \\
&+ \sum_{i=1}^n \left(-(1-p)^2 \alpha^2 X_i^{-2+2\alpha} (1 - \delta_i) Z_i^2 \theta^2 y_i^2 \left(\frac{Z_i}{K_i} (1 - \theta) + 1 \right)^2 \right. \\
&- \frac{(1-p)(-1 + \alpha) \alpha X_i^{-2+\alpha} (1 - \delta_i)}{1 - (1-p)X_i^\alpha} Z_i^2 \theta^2 y_i^2 \left(\frac{Z_i}{K_i} (1 - \theta) + 1 \right)^2 \\
&- \left. (1-p) \alpha X_i^{-1+\alpha} (1 - \delta_i) Z_i \theta y_i^2 \left(\frac{2Z_i(1 - e^{-\lambda y_i})(1 - \theta)^2 + 2Z_i(1 - \theta) - 1}{1 - (1-p)X_i^\alpha} \right) \right); \\
I_{\theta,\theta} &= \sum_{i=1}^n \left(\frac{(1-p)\alpha X_i^{-1+\alpha} A_i^2 (-A_i\theta + 1)(1 - \delta_i)}{1 - (1-p)X_i^\alpha} \left(2 - \frac{-A_i\theta + 1}{X_i} \left(\frac{(1-p)\alpha X_i^\alpha}{1 - (1-p)X_i^\alpha} - \alpha + 1 \right) \right) \right) \\
&- \frac{\alpha}{\theta} \sum_{i=1}^n \delta_i + \sum_{i=1}^n A_i^2 (1 + \alpha) \delta_i; \\
I_{\alpha,\theta} = I_{\theta,\alpha} &= \sum_{i=1}^n \left(\frac{-(1-p)X_i^{-1+\alpha} A_i (-A_i\theta + 1)(1 - \delta_i)}{1 - (1-p)X_i^\alpha} \left(1 - \frac{(1-p)X_i^\alpha \ln(X_i)}{1 - (1-p)X_i^\alpha} - \alpha \ln(X_i) \right) \right) \\
&- \sum_{i=1}^n A_i \delta_i + \sum_{i=1}^n \left(\left(\frac{2}{\theta} + \frac{\alpha \ln(\theta)}{\theta} \right) \delta_i - \frac{(1 + \alpha \ln(\theta)) \delta_i}{\theta} \right);
\end{aligned}$$

$$\begin{aligned}
I_{\alpha,\lambda} = I_{\lambda,\alpha} &= \sum_{i=1}^n \delta_i y_i (K_i + Z_i(1-\theta)) + \sum_{i=1}^n \left(\frac{-(1-p)X_i^{-1+\alpha}(1-\delta_i)\theta y_i Z_i (A_i(1-\theta) + 1)}{1 - (1-p)X_i^\alpha} \right. \\
&\quad \left. - \frac{(1-p)\alpha X_i^{-1+\alpha} \ln(X_i)(1-\delta_i)Z_i \theta y_i (A_i(1-\theta) + 1)}{1 - (1-p)X_i^\alpha} \right) \left(\frac{(1-p)X_i^\alpha}{1 - (1-p)X_i^\alpha + 1} \right); \\
I_{\alpha,p} = I_{p,\alpha} &= \sum_{i=1}^n \left(\frac{X_i^\alpha \ln(X_i)(1-\delta_i)}{1 - (1-p)X_i^\alpha} \left(\frac{(1-p)X_i^\alpha}{1 - (1-p)X_i^\alpha} + 1 \right) \right); \\
I_{\lambda,\theta} = I_{\theta,\lambda} &= \sum_{i=1}^n -Z_i(1+\alpha)\delta_i y_i (A_i(1-\theta) + 1) + \sum_{i=1}^n \left(-(1-p)\alpha X_i^{-1+\alpha}(1-\delta_i)Z_i A_i(1-\theta)y_i \right. \\
&\quad \times \left(\frac{-2A_i\theta + 1}{1 - (1-p)X_i^\alpha} - \frac{2\theta}{1-\theta} + \frac{1}{A_i(1-\theta)} \right) - \frac{(1-p)^2 \alpha^2 X_i^{-2+2\alpha} A_i(-A_i\theta + 1)(1-\delta_i)Z_i \theta y_i (1 + A_i(1-\theta))}{(1 - (1-p)X_i^\alpha)^2} \\
&\quad \left. - \frac{(1-p)(-1+\alpha)\alpha X_i^{-2+\alpha} A_i(-A_i\theta + 1)(1-\delta_i)Z_i \theta}{1 - (1-p)X_i^\alpha} (A_i(1-\theta) + y_i) \right); \\
I_{\lambda,p} = I_{p,\lambda} &= \sum_{i=1}^n \left(\frac{\alpha X_i^{-1+\alpha}(1-\delta_i)Z_i \theta y_i (A_i(1-\theta) + 1)}{1 - (1-p)X_i^\alpha} \left(\frac{(1-p)X_i^\alpha}{1 - (1-p)X_i^\alpha} + 1 \right) \right); \\
I_{\theta,p} = I_{p,\theta} &= \sum_{i=1}^n \frac{\alpha X_i^{-1+\alpha} A_i(-A_i\theta + 1)(1-\delta_i)}{1 - (1-p)X_i^\alpha} \left(\frac{X_i^\alpha(1-p)}{1 - (1-p)X_i^\alpha} + 1 \right); \\
I_{p,p} &= -\sum_{i=1}^n \frac{X_i^{2\alpha}(1-\delta_i)}{(1 - (1-p)X_i^\alpha)^2} - \sum_{i=1}^n \frac{\delta_i}{(1-p)^2};
\end{aligned}$$

em que $X_i = \frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$, $Z_i = \frac{e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta)+\theta}$, $K_i = \frac{e^{-\lambda y_i}}{1-e^{-\lambda y_i}}$ e $A_i = \frac{(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$.

Os elementos da matriz de informação de Fisher (4.19) apresentados no Capítulo 4 são dados por:

$$\begin{aligned}
I_{\alpha,\alpha} &= -\frac{\delta_i}{\alpha^2} - \sum_{i=1}^n (1-\delta_i) \ln(B_i)^2 \frac{B_i^\alpha}{1-B_i^\alpha} \left[1 + \frac{B_i^\alpha}{1-B_i^\alpha} \right]; \\
I_{\lambda,\lambda} &= -\sum_{i=1}^n \frac{\delta_i}{\lambda^2} + \sum_{i=1}^n \delta_i y_i X_i \left[-\frac{(\alpha-1)}{K_i^2} + \frac{(\alpha+1)\theta}{Z_i^2} \right] \\
&\quad - \sum_{i=1}^n \frac{(1-\delta_i)}{(1-B_i^\alpha)^2} \frac{\alpha^2 B_i^\alpha K_i^{\alpha-1}}{B_i Z_i^{\alpha+3}} X_i^2 (1-\theta)^2 + \sum_{i=1}^n \frac{(1-\delta_i)}{1-B_i^\alpha} \alpha B_i^\alpha \\
&\quad \times \left[\frac{X_i y_i (1-\theta)}{K_i Z_i} - \frac{(\alpha-1)}{B_i^2} \left(\frac{X_i^2 (1-\theta)^2}{Z_i^4} + \frac{2\theta(1-\theta)X_i^2 e^{-\lambda y_i}}{Z_i^2} \right) + \frac{2}{B_i} \frac{X_i^2 (1-\theta)\theta}{Z_i^3} \right]; \\
I_{\theta,\theta} &= -\sum_{i=1}^n \frac{\delta_i}{(1-\theta)^2} + (\alpha+1) \sum_{i=1}^n \frac{\delta_i e^{-2\lambda y_i}}{Z_i} + \alpha^2 \sum_{i=1}^n (1-\delta_i) B_i^{2\alpha-2} \frac{B_i^2}{Z_i^2 (1-B_i^\alpha)^2} e^{-2\lambda y_i} \\
&\quad - \alpha \sum_{i=1}^n (1-\delta_i) \frac{B_i^{\alpha-1}}{1-B_i^\alpha} \left[\frac{(\alpha-1) K_i^2 e^{-2\lambda y_i}}{B_i Z_i^4} + \frac{2e^{-2\lambda y_i} K_i}{Z_i^3} \right]; \\
I_{\lambda,\alpha} = I_{\alpha,\lambda} &= I_{\alpha,\lambda} = \sum_{i=1}^n \delta_i X_i \left[\frac{1}{K_i} - \frac{\theta}{Z_i} \right] \sum_{i=1}^n \frac{(1-\delta_i)}{(1-B_i^\alpha)^\alpha} \frac{B_i^\alpha}{B_i} \frac{X_i(1-\theta)}{Z_i^2} \\
&\quad \times \left(1 + \frac{B_i^\alpha}{(1-B_i^\alpha)} \ln(B_i) \alpha + \ln(B_i) \right); \\
I_{\lambda,\theta} = I_{\theta,\lambda} &= I_{\theta,\lambda} = -(\alpha+1) \sum_{i=1}^n \delta_i X_i \left[\frac{1}{Z_i} + \frac{\theta e^{-\lambda y_i}}{Z_i^2} \right] + \sum_{i=1}^n \frac{(1-\delta_i)}{(1-B_i^\alpha)^2} \frac{X_i e^{-\lambda y_i} \alpha (1-\theta) K_i^\alpha}{Z_i^{\alpha+1}} \\
&\quad - \sum_{i=1}^n \frac{(1-\delta_i)}{1-B_i^\alpha} \alpha B_i^{\alpha-1} \left[\frac{e^{-\lambda y_i} X_i (1-\theta)}{Z_i^3} - X_i \left(\frac{-1 + 2e^{-\lambda y_i} - \theta e^{-2\lambda y_i} (2-\theta)}{Z_i^4} \right) \right]; \\
I_{\theta,\alpha} = I_{\alpha,\theta} &= I_{\alpha,\theta} = \sum_{i=1}^n \frac{\delta_i e^{-\lambda y_i}}{Z_i} - \sum_{i=1}^n \frac{(1-\delta_i)}{1-B_i^\alpha} \frac{B_i^\alpha}{Z_i} e^{-\lambda y_i} \left[1 - \frac{\alpha B_i^\alpha}{1-B_i^\alpha} \ln(B_i) - \alpha \ln(B_i) \right];
\end{aligned}$$

em que $K_i = 1 - e^{-\lambda y_i}$, $Z_i = 1 - \theta e^{-\lambda y_i}$, $B_i = \frac{K_i}{Z_i}$ e $X_i = y_i e^{-\lambda y_i}$.

Os elementos da matriz de informação de Fisher (5.15) do Capítulo 5 são dados

por:

$$\begin{aligned}
I_{\alpha,\alpha} &= \sum_{i=1}^n \delta_i \left(\frac{1}{\theta} - \frac{1}{\alpha^2} \right) - \frac{(1-\delta_i)(\ln X_i)^2 X_i^\alpha}{1-X_i^\alpha} \left(1 + \frac{X_i^\alpha}{1-X_i^\alpha} \right) - \frac{\sum_{i=1}^n (1+\delta_i)(1-\theta) X_i^\alpha (\ln(X_i))^2}{\theta + (1-\theta) X_i^\alpha}; \\
I_{\lambda,\lambda} &= -\frac{\sum_{i=1}^n \delta_i}{\lambda^2} - \frac{\sum_{i=1}^n (1-\delta_i) \alpha \theta y_i}{1-X_i^\alpha} \left[(\alpha-1) X_i^{\alpha-2} \theta^2 Z_i^2 y_i^2 (1+(1-\theta) A_i) + \alpha X_i^{\alpha-1} y_i Z_i (Z_i (1-\theta) - 1) \right] \\
&\quad - \frac{\sum_{i=1}^n (1-\delta_i) \alpha^3 \theta^2 y^2 X_i^{2\alpha-2} Z_i^2 \theta y (1+(1-\theta) A_i)}{(1-X_i^\alpha)^2} - \frac{\sum_{i=1}^n (1+\delta_i)(1-\theta) \theta y_i \alpha}{\theta + (1-\theta) X_i^\alpha} (((1-\theta) Z_i - 1) \\
&\quad \times y_i Z_i X_i^{\alpha-1} + (\alpha-1) X_i^{\alpha-2} Z_i^2 y \theta (1+(1-\theta) A_i)) - \frac{\sum_{i=1}^n (1-\delta_i) \alpha \theta y_i (1-\theta) (\alpha-1) X_i^{\alpha-2}}{(1-X_i^\alpha)^2} \\
&\quad \times \left(\theta y_i Z_i^2 A_i + \theta y_i (1-\theta) Z_i^2 A_i^2 + X_i^{\alpha-1} Z_i^2 \left(-\frac{y}{e^{-\lambda y_i}} + 2y_i + \frac{2y_i(1-\theta) Z_i}{(e^{-\lambda y_i})^2} + 2y_i(1-\theta) Z_i \right) \right) \\
&\quad - \frac{\sum_{i=1}^n (\alpha-1) y^2 K_i^2}{e^{-\lambda y_i}} + \frac{\sum_{i=1}^n (1-\delta_i)(1-\theta)^2 \alpha \theta^2 y_i^2 X_i^{2\alpha-2} Z_i^2 (1+(1-\theta) A_i)}{(\theta + (1-\theta) X_i^\alpha)^2} \\
&\quad - \frac{\sum_{i=1}^n (1+\delta_i)(1-\theta)^2 \alpha \theta^2 y^2 (\alpha-1) A_i X_i^{\alpha-2} (1+(1-\theta) A_i)}{(\theta + (1-\theta) X_i^\alpha)^2 Z_i^2} \\
&\quad + \frac{\sum_{i=1}^n (1+\delta_i)(1-\theta)^2 \alpha \theta y^2 Z_i^2 X_i^{\alpha-1} \left(2 - \frac{1}{e^{-\lambda y_i}} + \frac{2(1-\theta) Z_i (1-e^{-2\lambda y_i})}{(e^{-\lambda y_i})^2} \right)}{(\theta + (1-\theta) X_i^\alpha)^2} \\
&\quad + \sum_{i=1}^n (\alpha+1)(1-\theta) y_i^2 Z_i (-1 + (1-\theta) Z_i); \\
I_{\theta,\theta} &= -\frac{n}{\theta^2} - \frac{\sum_{i=1}^n \delta_i}{\lambda \theta^2} + \frac{(1-\delta_i)}{1-X_i^\alpha} \left(-A_i^2 - 2\theta A_i^2 + 2\theta^2 A_i^3 - \alpha X_i^{\alpha-1} \left(\frac{A_i - \theta X_i^2}{1-X_i^\alpha} \right)^2 \right) \\
&\quad + \frac{\sum_{i=1}^n (1+\delta_i)}{\theta + (1-\theta) X_i^\alpha} \left(\frac{1-X_i^\alpha}{\theta + (1-\theta) X_i^\alpha} \left(1-X_i^\alpha + (1-\theta) \alpha \theta^{\alpha-1} X_i^\alpha - (1-\theta) \alpha \theta^\alpha X_i^{\alpha+1} \right) + A_i^\alpha \alpha \theta^\alpha (-\theta^{-1} + A_i) \right) \\
&\quad - \theta^{\alpha-1} \alpha A_i^\alpha \left(1 - \frac{(\alpha-1)(1-\theta)}{\theta} \right) - \alpha \frac{X_i^{\alpha+1}}{\theta} \left(-1 + \frac{(1-\theta)}{\theta} - \frac{(1-\theta)}{\theta} \left(\frac{\alpha}{\theta} - (\alpha+1) A_i \right) \right) + (\alpha+1) \sum_{i=1}^n A_i^2; \\
I_{\alpha,\lambda} = I_{\lambda,\alpha} &= -\frac{\sum_{i=1}^n (1-\delta_i) X_i^\alpha}{1-X_i^\alpha} \left(\frac{\alpha \ln(X_i)}{(1-e^{-\lambda y_i})} + \alpha y_i (1-\theta) Z_i \ln(X_i) + y_i K_i + \theta (1-\theta) y_i Z_i \right) \\
&\quad + \frac{\ln(X_i) X_i^{\alpha-1} \theta y_i Z_i (1+(1-\theta) A_i)}{(1-X_i^\alpha)} - \frac{\sum_{i=1}^n (1-\delta_i) Z_i y_i}{\theta + (1-\theta) X_i^\alpha} \left((1-\theta) \alpha X_i^{\alpha-1} \theta (1+(1-\theta) A_i) \ln(X_i) + (1-\theta) \theta^\alpha \right) \\
&\quad \times \left((1-e^{-\lambda y_i})^{\alpha-1} + A_i^\alpha \right) + \sum_{i=1}^n \frac{(1+\delta_i)}{(\theta + (1-\theta) X_i^\alpha)^2} \left((1-\theta)^2 \alpha X_i^{2\alpha-1} \ln(X_i) \theta y_i Z_i (1+(1-\theta) A_i) \right) \\
&\quad + \sum_{i=1}^n y_i (K_i + (1-\theta) Z_i); \\
I_{\alpha,\theta} = I_{\theta,\alpha} &= -\frac{\sum_{i=1}^n \delta_i \alpha}{\theta^2} - \frac{\sum_{i=1}^n (1-\delta_i) X_i^{\alpha-1}}{(1-X_i^\alpha)} \left(A_i (1-\theta A_i) (\alpha \ln(X_i) + 1) + \frac{\alpha X_i^\alpha}{1-X_i^\alpha} \ln(X_i) \left(\frac{X_i}{\theta} (1-X_i) \right) \right) \\
&\quad - \frac{\sum_{i=1}^n (1+\delta_i)}{\theta + (1-\theta) X_i^\alpha} \left(-X_i^\alpha \left(\ln(X_i) + (1-\theta) \alpha \left(\frac{1}{\theta} - \frac{X_i}{\theta} \right) \ln(X_i) + (1-\theta) \left(\frac{1-X_i}{\theta} \right) \right) \right) \\
&\quad + \frac{\sum_{i=1}^n (1-\delta_i)}{(\theta + (1-\theta) X_i^\alpha)^2} X_i^\alpha \left((1-\theta) \ln(X_i) (1-X_i^\alpha) + (1-\theta) \alpha \left(\frac{X_i}{\theta} \right) \left(\frac{1-X_i}{X_i} \right) \right) - \sum_{i=1}^n \frac{X_i}{\theta}; \\
I_{\lambda,\theta} = I_{\theta,\lambda} &= -\frac{\sum_{i=1}^n (1-\delta_i) \alpha X_i^{\alpha-1} y_i Z_i}{1-X_i^\alpha} \left((\alpha-1)(1-X_i) + (1-\theta) A_i \right) \left(1 - \frac{(\alpha X_i^{\alpha-1}) A_i}{(1-X_i^\alpha)} \right) \\
&\quad - \sum_{i=1}^n (\alpha+1) y_i Z_i (1+(1-\theta) A_i) - \frac{(1+\delta_i) y_i X_i^{\alpha-1} \alpha}{\theta + (1-\theta) X_i^\alpha} \left[-\theta Z_i + e^{-\lambda y_i} (1-\theta) \left(\frac{Z_i}{e^{-\lambda y_i}} - \frac{\theta A_i^2}{1-e^{-\lambda y_i}} \right) \right] \\
&\quad + (1-\theta) Z_i (\alpha-1)(1-X_i) - \frac{(1-\theta) \theta Z_i}{\theta + (1-\theta) X_i^\alpha} \times \left(1-X_i^\alpha + (1-\theta) \alpha \frac{X_i^\alpha}{\theta} (1-X_i) \right) + \frac{\theta (1-\theta)^2 e^{-\lambda y_i} A_i^2}{e^{-\lambda y_i} (1-\theta) + \theta} \\
&\quad - (1-\theta) Z_i A_i \times \left(2 - \frac{(1-\theta)}{\theta} - (\alpha-1) X_i^{\alpha-2} \left(A_i - \frac{X_i^2}{\theta} \right) - \frac{1}{\theta + (1-\theta) X_i^\alpha} \left(1 + X_i^\alpha + (1-\theta) \alpha X_i^{\alpha-1} \right) \right. \\
&\quad \times \left. (A_i - X_i^2) \right) \Big] \tag{3}
\end{aligned}$$

em que $X_i = \frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$, $Z_i = \frac{e^{-\lambda y_i}}{e^{-\lambda y_i}(1-\theta)+\theta}$, $K_i = \frac{e^{-\lambda y_i}}{1-e^{-\lambda y_i}}$ and $A_i = \frac{(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta)+\theta}$.

Tabela 1: Valores do TRV e p-valor para cada ajuste ECEG

| Ajuste ECEG | -Log | TRV | Teste P-valor |
|-------------------------------|----------|---------|---------------|
| Intercepto | 2071,254 | - | - |
| X1 | 2050,566 | 41,376 | <0,001 |
| X2 | 2035,750 | 71,008 | 0 |
| X3 | 2070,781 | 0,946 | 0,330 |
| X4 | 2006,701 | 129,106 | 0 |
| X5 | 2060,453 | 21,602 | <0,001 |
| X6 | 2062,318 | 17,872 | <0,001 |
| X4 | 2006,701 | - | - |
| X4+X1 | 1998,589 | 16,224 | <0,001 |
| X4+X2 | 1955,056 | 103,290 | 0 |
| X4+X3 | 2005,355 | 2,692 | 0,100 |
| X4+X5 | 1998,716 | 15,970 | <0,001 |
| X4+X6 | 1999,454 | 14,494 | 0,0001 |
| X4+X2 | 1955,056 | - | - |
| X4+X2+X1 | 1713,071 | 483,970 | 0 |
| X4+X2+X3 | 1945,929 | 18,254 | <0,001 |
| X4+X2+X5 | 1947,318 | 15,476 | <0,001 |
| X4+X2+X6 | 1950,900 | 8,312 | 0,003 |
| X4+X2+X1 | 1713,071 | - | - |
| X4+X2+X1+X3 | 1708,040 | 10,062 | 0,001 |
| X4+X2+X1+X5 | 1710,874 | 4,394 | 0,036 |
| X4+X2+X1+X6 | 1712,927 | 0,288 | 0,591 |
| X4+X2+X1+X3 | 1708,04 | - | - |
| X4+X2+X1+X3+X5 | 1705,332 | 5,416 | 0,019 |
| X4+X2+X1+X3+X6 | 1707,828 | 0,424 | 0,514 |
| X4+X2+X1+X3+X5 | 1705,332 | - | - |
| X4+X2+X1+X3+X5+X6 | 1705,454 | -0,244 | 1,000 |
| X4+X2+X1+X3+X5 | 1705,332 | - | - |
| X4+X2+X1+X3+X5+X1*X2 | 1701,645 | 7,374 | 0,006 |
| X4+X2+X1+X3+X5+X1*X3 | 1696,301 | 18,062 | <0,001 |
| X4+X2+X1+X3+X5+X1*X4 | 1690,975 | 28,714 | <0,001 |
| X4+X2+X1+X3+X5+X1*X5 | 1704,956 | 0,752 | 0,385 |
| X4+X2+X1+X3+X5+X1*X6 | 1703,503 | 3,658 | 0,055 |
| X4+X2+X1+X3+X5+X2*X6 | 1725,806 | -40,948 | 1 |
| X4+X2+X1+X3+X5+X3*X6 | 1704,648 | 1,368 | 0,242 |
| X4+X2+X1+X3+X5+X4*X6 | 1704,981 | 0,702 | 0,402 |
| X4+X2+X1+X3+X5+X5*X6 | 1704,540 | 1,548 | 0,208 |
| X4+X2+X1+X3+X5+X1*X4 | 1690,975 | - | - |
| X4+X2+X1+X3+X5+X1*X4+X1*X2*X6 | 1688,733 | 4,484 | 0,034 |
| X4+X2+X1+X3+X5+X1*X4+X1*X3*X6 | 1687,360 | 7,230 | 0,007 |
| X4+X2+X1+X3+X5+X1*X4+X1*X4*X6 | 1690,781 | 0,388 | 0,533 |
| X4+X2+X1+X3+X5+X1*X4+X1*X5*X6 | 1690,055 | 1,84 | 0,174 |

Tabela 2: Valores do TRV e p-valor para cada ajuste

| Ajuste CEG | -Log | Teste TRV | P-valor |
|-------------------------------|----------|-----------|--------------------|
| Intercepto | 2211,662 | - | - |
| X1 | 2187,198 | 48,928 | <0,001 |
| X2 | 2158,285 | 106,754 | 0 |
| X3 | 2210,650 | 2,024 | 0,154 |
| X4 | 2135,846 | 151,632 | 0 |
| X5 | 2196,023 | 31,278 | <0,001 |
| X6 | 2199,112 | 25,1 | <0,001 |
| X4 | 2135,846 | - | - |
| X4+X1 | 2125,541 | 20,610 | <0,001 |
| X4+X2 | 2057,261 | 157,170 | 0 |
| X4+X3 | 2133,228 | 5,236 | 0,022 |
| X4+X5 | 2122,544 | 26,604 | <0,001 |
| X4+X6 | 2124,122 | 23,448 | <0,001 |
| X4+X2 | 2057,261 | - | - |
| X4+X2+X1 | 1745,13 | 624,262 | 0 |
| X4+X2+X3 | 2042,089 | 30,344 | <0,001 |
| X4+X2+X5 | 2044,909 | 24,704 | <0,001 |
| X4+X2+X6 | 2050,832 | 12,858 | 0,0003 |
| X4+X2+X1 | 1745,13 | - | - |
| X4+X2+X1+X3 | 1741,359 | 7,542 | 0,006 |
| X4+X2+X1+X5 | 1741,806 | 6,648 | 0,009 |
| X4+X2+X1+X6 | 1744,084 | 2,092 | 0,148 |
| X4+X2+X1+X3 | 1741,359 | - | - |
| X4+X2+X1+X3+X5 | 1737,909 | 6,9 | 0,008 |
| X4+X2+X1+X3+X6 | 1740,149 | 2,42 | 0,119 |
| X4+X2+X1+X3+X5 | 1737,909 | - | - |
| X4+X2+X1+X3+X5+X6 | 1736,740 | 2,338 | 0,126 |
| X4+X2+X1+X3+X5 | 1737,909 | - | - |
| X4+X2+X1+X3+X5+X1*X2 | 1731,704 | 12,410 | 4×10^{-4} |
| X4+X2+X1+X3+X5+X1*X3 | 1725,949 | 23,920 | <0,001 |
| X4+X2+X1+X3+X5+X1*X4 | 1723,338 | 29,142 | <0,001 |
| X4+X2+X1+X3+X5+X1*X5 | 1736,664 | 2,49 | 0,114 |
| X4+X2+X1+X3+X5+X1*X6 | 1735,093 | 5,632 | 0,017 |
| X4+X2+X1+X3+X5+X2*X6 | 1729,667 | 16,484 | <0,001 |
| X4+X2+X1+X3+X5+X3*X6 | 1736,346 | 3,126 | 0,077 |
| X4+X2+X1+X3+X5+X4*X6 | 1735,563 | 4,692 | 0,030 |
| X4+X2+X1+X3+X5+X5*X6 | 1736,556 | 2,706 | 0,099 |
| X4+X2+X1+X3+X5+X1*X4 | 1723,338 | - | - |
| X4+X2+X1+X3+X5+X1*X4+X1*X2*X6 | 1720,028 | 6,620 | 0,010 |
| X4+X2+X1+X3+X5+X1*X4+X1*X3*X6 | 1719,650 | 7,376 | 0,006 |
| X4+X2+X1+X3+X5+X1*X4+X1*X4*X6 | 1722,910 | 0,856 | 0,354 |
| X4+X2+X1+X3+X5+X1*X4+X1*X5*X6 | 1722,007 | 2,662 | 0,102 |

Tabela 3: Valores do TRV e p-valor para cada ajuste

| Ajuste E | -Log | Teste TRV | P-valor |
|-------------------------------|----------|-----------|---------|
| Intercepto | 2211,658 | - | - |
| X1 | 2187,198 | 48,92 | <0,001 |
| X2 | 2158,279 | 106,758 | 0 |
| X3 | 2210,646 | 2,024 | 0,154 |
| X4 | 2135,836 | 151,644 | 0 |
| X5 | 2196,023 | 31,270 | <0,001 |
| X6 | 2199,108 | 25,100 | <0,001 |
| X4 | 2135,836 | - | - |
| X4+X1 | 2125,531 | 20,610 | <0,001 |
| X4+X2 | 2057,259 | 157,154 | 0 |
| X4+X3 | 2133,228 | 5,216 | 0,022 |
| X4+X5 | 2122,584 | 26,504 | <0,001 |
| X4+X6 | 2124,115 | 23,442 | <0,001 |
| X4+X2 | 2057,259 | - | - |
| X4+X2+X1 | 1748,087 | 618,344 | 0 |
| X4+X2+X3 | 2042,062 | 30,394 | <0,001 |
| X4+X2+X5 | 2044,92 | 24,678 | <0,001 |
| X4+X2+X6 | 2050,828 | 12,862 | 0,0003 |
| X4+X2+X1 | 1748,087 | - | - |
| X4+X2+X1+X3 | 1745,12 | 5,934 | 0,014 |
| X4+X2+X1+X5 | 1744,761 | 6,652 | 0,009 |
| X4+X2+X1+X6 | 1746,603 | 2,968 | 0,084 |
| X4+X2+X1+X3 | 1744,761 | - | - |
| X4+X2+X1+X3+X5 | 1741,837 | 5,848 | 0,015 |
| X4+X2+X1+X3+X6 | 1743,305 | 2,912 | 0,087 |
| X4+X2+X1+X3+X5 | 1741,837 | - | - |
| X4+X2+X1+X3+X5+X6 | 1740,146 | 9,23 | 0,002 |
| X4+X2+X1+X3+X5 | 1740,146 | - | - |
| X4+X2+X1+X3+X5+X1*X2 | 1735,632 | 9,028 | 0,002 |
| X4+X2+X1+X3+X5+X1*X3 | 1731,187 | 17,918 | <0,001 |
| X4+X2+X1+X3+X5+X1*X4 | 1727,705 | 24,882 | <0,001 |
| X4+X2+X1+X3+X5+X1*X5 | 1740,148 | -0,004 | 1 |
| X4+X2+X1+X3+X5+X1*X6 | 1738,497 | 3,298 | 0,069 |
| X4+X2+X1+X3+X5+X2*X6 | 1734,052 | 12,188 | 0,0004 |
| X4+X2+X1+X3+X5+X3*X6 | 1739,557 | 1,178 | 0,277 |
| X4+X2+X1+X3+X5+X4*X6 | 1738,947 | 2,398 | 0,121 |
| X4+X2+X1+X3+X5+X5*X6 | 1739,993 | 0,306 | 0,580 |
| X4+X2+X1+X3+X5+X1*X4 | 1727,705 | - | - |
| X4+X2+X1+X3+X5+X1*X4+X1*X2*X6 | 1724,724 | 5,962 | 0,014 |
| X4+X2+X1+X3+X5+X1*X4+X1*X3*X6 | 1724,882 | 5,646 | 0,017 |
| X4+X2+X1+X3+X5+X1*X4+X1*X4*X6 | 1727,184 | 1,040 | 0,307 |
| X4+X2+X1+X3+X5+X1*X4+X1*X5*X6 | 1726,331 | 2,748 | 0,097 |

Tabela 4: Valores do TRV e p-valor para cada ajuste

| Ajuste LECEG | -Log | Teste TRV | P-valor |
|-------------------------|----------|--------------|----------------------|
| Intercepto | 463,111 | - | - |
| X1 | 461,133 | 3,954 | $4,6 \times 10^{-2}$ |
| X2 | 462,552 | 1,118 | $2,9 \times 10^{-1}$ |
| X4 | 458,263 | 9,694 | $1,8 \times 10^{-3}$ |
| X3 | 447,828 | 30,566 | $3,2 \times 10^{-8}$ |
| X3 | 447,828 | - | - |
| X3+X1 | 446,878 | 1,898 | 0,168 |
| X3+X2 | 446,567 | 2,521 | 0,112 |
| X3+X4 | 445,329 | 4,997 | 0,025 |
| X3+X4 | 445,3292 | - | - |
| X3+X4+X1 | 443,357 | 3,943 | 0,047 |
| X3+X4+X2 | 444,123 | 2,411 | 0,120 |
| X3+X4+X1 | 443,357 | - | - |
| X3+X4+X1+X2 | 442,515 | 1,684 | 0,194 |
| X3+X4+X1 | 443,357 | - | - |
| X3+X4+X1+X1*X4 | 442,214 | 2,285 | 0,130 |
| X3+X4+X1+X1*X3 | 437,613 | 11,488 | 7×10^{-4} |
| X3+X4+X1+X2*X4 | 442,726 | 1,262 | 0,261 |
| X3+X4+X1+X2*X3 | 443,239 | 0,235 | 0,627 |
| X3+X4+X1+X1*X3 | 437,6134 | - | - |
| X3+X4+X1+X1*X3+X1*X3*X4 | 437,5241 | 0,178 | 0,672 |
| X3+X4+X1+X1*X3+X2*X3*X4 | 437,457 | 0,312 | 0,575 |

Tabela 5: Valores do TRV e p-valor para cada ajuste

| Ajuste LCEG | -Log | Teste TRV | P-valor |
|-------------------------|----------|--------------|-----------------------|
| Intercepto | 511,1957 | - | - |
| X1 | 510,589 | 1,212 | $2,7 \times 10^{-1}$ |
| X2 | 511,156 | 0,079 | $7,7 \times 10^{-1}$ |
| X4 | 501,786 | 18,818 | $1,4 \times 10^{-5}$ |
| X3 | 485,875 | 50,640 | $1,1 \times 10^{-12}$ |
| X3 | 485,8755 | - | - |
| X3+X1 | 484,286 | 3,178 | 0,074 |
| X3+X2 | 479,831 | 12,087 | 5×10^{-4} |
| X3+X4 | 481,888 | 7,974 | 0,004 |
| X3+X2 | 479,8319 | - | - |
| X3+X2+X1 | 477,850 | 3,9635 | 0,046 |
| X3+X4+X2 | 476,503 | 6,6559 | 0,009 |
| X3+X4+X2 | 476,503 | - | - |
| X3+X4+X1+X2 | 475,366 | 2,275 | 0,131 |
| X3+X4+X2 | 476,503 | - | - |
| X3+X4+X2+X1*X4 | 474,210 | 4,586 | 0,032 |
| X3+X4+X2+X1*X3 | 481,861 | -10,715 | 1,000 |
| X3+X4+X2+X2*X4 | 476,351 | 0,304 | 0,581 |
| X3+X4+X2+X2*X3 | 476,297 | 0,412 | 0,520 |
| X3+X4+X2+X1*X4 | 474,2298 | - | - |
| X3+X4+X2+X1*X4+X1*X3*X4 | 474,027 | 0,403 | 0,525 |
| X3+X4+X2+X1*X4+X2*X3*X4 | 476,195 | -3,931 | 1 |

Tabela 6: Valores do TRV e p-valor para cada ajuste

| Ajuste LE | -Log | Teste TRV | P-valor |
|----------------------------|-------------|----------------------|-----------------------|
| Intercepto | 511,1855 | - | - |
| X1 | 510,572 | 1,226 | $2,6 \times 10^{-1}$ |
| X2 | 511,153 | 0,063 | $8,0 \times 10^{-1}$ |
| X4 | 501,783 | 18,804 | $1,4 \times 10^{-5}$ |
| X3 | 485,873 | 50,624 | $1,1 \times 10^{-12}$ |
| X3 | 485,873 | - | - |
| X3+X1 | 484,270 | 3,206 | 0,073 |
| X3+X2 | 479,830 | 12,085 | 5×10^{-4} |
| X3+X4 | 481,869 | 8,007 | 0,004 |
| X3+X2 | 479,830 | - | - |
| X3+X2+X1 | 477,8581 | 3,950 | 0,047 |
| X3+X4+X2 | 479,830 | 0 | 1,000 |
| X3+X2+X1 | 477,8581 | - | - |
| X3+X2+X1+X4 | 475,362 | 4,991 | 0,025 |
| X3+X2+X1+X4 | 475,362 | - | - |
| X3+X2+X1+X4+X1*X4 | 474,059 | 2,605 | 0,106 |
| X3+X2+X1+X4+X1*X3 | 471,892 | 6,939 | 0,008 |
| X3+X2+X1+X4+X2*X4 | 475,182 | 0,359 | 0,548 |
| X3+X2+X1+X4+X2*X3 | 475,339 | 0,046 | 0,829 |
| X3+X2+X1+X4+X1*X3 | 471,8926 | - | - |
| X3+X2+X1+X4+X1*X3+X1*X3*X4 | 472,916 | -2,048 | 1,000 |
| X3+X2+X1+X4+X1*X3+X2*X3*X4 | 475,174 | -6,564 | 1,000 |