RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 26/08/2021.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Bayesian and Classical Inference for Extensions of Geometric Exponential Distribution with Applications in Survival Analysis Under the Presence of the Data Covariated and Randomly Censored

Paulo Roberto de Lima Gianfelice

Academic Supervisor: Prof. Fernando Antonio Moala, Ph.D.

Post-Graduate Program in Applied and Computational Mathematics

Presidente Prudente, February 2020

UNIVERSIDADE ESTADUAL PAULISTA
'JÚLIO DE MESQUITA FILHO'
Faculdade de Ciências e Tecnologia de Presidente Prudente
Post-Graduate Program in Applied and Computational Mathematics

Bayesian and Classical Inference for Extensions of Geometric Exponential Distribution with Applications in Survival Analysis Under the Presence of the Data Covariated and Randomly Censored

Paulo Roberto de Lima Gianfelice

Academic Supervisor: Prof. Fernando Antonio Moala, Ph.D.

Master's Thesis presented to the Post-Graduate Program in Applied and Computational Mathematics of the Faculty of Science and Technology of the São Paulo State University "Júlio de Mesquita Filho", Presidente Prudente campus, in partial fulfillment of the requirements for title of Master in Applied and Computational Mathematics.

Presidente Prudente, February 2020

FICHA CATALOGRÁFICA

G433b

Gianfelice, Paulo Roberto de Lima

Bayesian and Classical Inference for Extensions of Geometric Exponential Distribution with Data Covariated and Randomly Applications in Survival Analysis Under the Presence of the Censored/ Paulo Roberto de Lima Gianfelice - Presidente Prudente, 2020

91 p.: il, tabs.

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, Presidente Prudente, 2020

Orientador: Fernando Antonio Moala

Inclui bibliografia

1. Censored and covariate data. 2. Maximum likelihood and bayesian estimation. 3. Extensions for Exponential Geometric distribution. 4. Statistical simulation. VI. Título.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Presidente Prudente

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Bayesian and classical inference for extensions of geometric exponential distribution with applications in survival analysis under the presence of the data covariated and randomly censored

AUTOR: PAULO ROBERTO DE LIMA GIANFELICE
ORIENTADOR: FERNANDO ANTONIO MOALA

Aprovado como parte das exigências para obtenção do Título de Mestre em MATEMÁTICA APLICADA E COMPUTACIONAL, pela Comissão Examinadora:

Prof. Dr. FERNANDO ANTONIO MOALA
Departamento de Estatística / Faculdade de Ciencias e Tecnologia de Presidente Prudente

Prof. Dr. LUIZ CARLOS BENINI
Departamento de Estatística / Faculdade de Ciencias e Tecnologia de Presidente Prudente

VIDEOCONFERÊNCIA

Prof. Dr. PEDRO LUIZ RAMOS
Pós-Doutorando / Universidade de São Paulo

Presidente Prudente, 26 de fevereiro de 2020

To all the people who supported, taught, instructed and guided me.
In particular, for Olézia Gianfelice is dedicad!

Acknowledgements

Acknowledgements first and foremost, to all gods for my soul, health, wisdom, deliverance from evil, courage and patience, without these items I would not have been not even able even win my self... not even human I would be. It is not possible for man, alone, to leave the most unlikely place, overcome all obstacles, go against all expectations and perform a feat like this, in a place like this.

Thank my parents for conceiving me, as the first and with the greatest love on the 15th day of January 1985. Without their care, examples and deeds I would never have taken this direction, arrived where I am and taken the steps decisions I made.

I will thank you until the end of my life for my three beautiful brothers, they are the primary inspiration and reason for my conquests. I thank my whole family from the bottom of my heart for their understanding, support and encouragement, without their support and advice I would have been lost me in the course of this journey. I especially thank from "Jão", my cousin-friend-brother, and my aunt "Cris", the first for always remembering and believing in me and the second for all the affection, support, help, understanding and value. I seek success in my achievements as an academic, and I will continue to seek in other areas of life, as a way of compensating and honoring them. Their care, guidance and incentives made me a person capable of overcoming anything.

To all my teachers, in particular to Gilcilene Sanchez De Paulo and Manoel Ivanildo Silvestre Bezerra. For the professional inclination and attendance, the first for teaching me to lead the learning of the most important concept in this line of the academy with affection and patience, and the second for believing in my ability, giving me freedom in the construction of my academic works, reading them and mainly trusting my proposals. In particular, I thank Professor Marta, for her Probability and Statistics classes in my 4th year of the Mathematics course opened my eyes and aroused interest in this career.

To my companions in study and unwinding, André Antunes, Bruno Gois, José Vanterler (Zé), Laisson, Leonardo Kengi (Léo), Marcos Viana (Osama) and Rafael Paulino (Pão), who besides being extraordinary people and academics, believed in my capabilities, tolerated me in unpleasant moments of tiredness and frustration. I am grateful to them for the guidance, influence, notes, advice, invitations and doubts taken about n situations of personal and academic life, especially because they are extremely nice people and exemplary academics I feel privileged to have lived with them at this stage.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, case number 88882.441640/2019-01, and FCT/UNESP, through Pos MAC, therefore, a recognition especially to for these institutions by the opportunity, credit, structure and privileges in the development of this work, training and academic career. Through these two institutions I also transfer my gratitude to each of the Brazilians who contributed with culture, kindness, serenity, trust, and work to pay yours taxes, fees and all tax contributions for development and outcome of scientific and technological work , as well as for the training and academic career of scientists/researchers in Brazil.

"E o gládio a erguer, que arrasa e que depreda,
E o olhar que ante a agnomínia não desmaia,
Luta! E é forçoso que ao lutar não caia,
Pois se cair o esmagarão na queda."
*(Aleluias, **Raimundo Correia**)*

Abstract

This work presents a study of probabilistic modeling, with applications to survival analysis, based on a probabilistic model called Exponential Geometric (EG), which offers great flexibility for the statistical estimation of its parameters based on samples of life time data complete and censored. In this study, the concepts of estimators and lifetime data are explored under random censorship in two cases of extensions of the EG model: the Extended Geometric Exponential (EEG) and the Generalized Extreme Geometric Exponential (GE2). The work still considers, exclusively for the EEG model, the approach of the presence of covariates indexed in the rate parameter as a second source of variation to add even more flexibility to the model, as well as, exclusively for the GE2 model, a analysis of the convergence, hitherto ignored, it is proposed for its moments. The statistical inference approach is performed for these extensions in order to expose (in the classical context) their maximum likelihood estimators and asymptotic confidence intervals, and (in the bayesian context) their a priori and a posteriori distributions, both cases to estimate their parameters under random censorship, and covariates in the case of EEG. In this work, bayesian estimators are developed with the assumptions that the prioris are vague, follow a Gamma distribution and are independent between the unknown parameters. The results of this work are regarded from a detailed study of statistical simulation applied to compare the estimation procedures approached under the pretext of evaluating these estimators based on the 95% coverage probability, mean square error, mean bias and the mean interval amplitude. At the end of each extension's approach, an application with real data is also presented to highlight the reach and particularities of the extended model addressed.

Keywords: Censored and covariate data, Maximum likelihood and bayesian estimation, Extensions for Exponential Geometric distribution, Statistical simulation.

Resumo

Este trabalho apresenta um estudo de modelagem probabilística, com aplicações à análise de sobrevivência, fundamentado em um modelo probabilístico denominado Exponencial Geométrico (EG), que oferece uma grande flexibilidade para a estimação estatística de seus parâmetros com base em amostras de dados de tempo de vida completos e censurados. Neste estudo são explorados os conceitos de estimadores e dados de tempo de vida sob censuras aleatórias em dois casos de extensões do modelo EG: o Exponencial Geométrico Estendido (EEG) e o Exponencial Geométrico Extremo Generalizado (GE2). O trabalho ainda considera, exclusivamente para o modelo EEG, a abordagem de presença de covariáveis indexadas no parâmetro de taxa como uma segunda fonte de variação para acrescentar ainda mais flexibilidade para o modelo, bem como, exclusivamente para o modelo GE2, uma análise de convergência até então ignorada, é proposta para seus momentos. A abordagem da inferência estatística é realizada para essas extensões no intuito de expor (no contexto clássico) seus estimadores de máxima verossimilhança e intervalos de confiança assintóticos, e (no contexto bayesiano) suas distribuições à priori e posteriori, ambos os casos para estimar seus parâmetros sob as censuras aleatórias, e covariáveis no caso do EEG. Neste trabalho os estimadores bayesianos são desenvolvidos com os pressupostos de que as prioris são vagas, seguem uma distribuição Gama e são independentes entre os parâmetros desconhecidos. Os resultados deste trabalho são resguardados de um estudo detalhado de simulação estatística aplicado para comparar os procedimentos de estimação abordados sob o pretexto de avaliar estes estimadores com base na probabilidade de 95% de cobertura, erro quadrático médio, vício médio e a amplitude intervalar média. Ao final da abordagem de cada extensão é apresentada ainda uma aplicação com dados reais para destacar o alcance e as particularidades do modelo estendido abordado.

Palavras-Chave: Dados censurados e covariados, Estimação de máxima verossimilhança e bayesiano, Extensões para a distribuição Exponencial Geométrica, Simulação estatística.

List of Figures

List of Tables

List of Acronyms

AIC: Akaike Information Criterion
AICC: Akaike Information Criterion Corrected
BIC: Bayesian Information Criterion
DIC: Deviance Information Criterion
E2G: Exponentiated Exponential Geometric.
EG: Exponential Geometric.
CE2G: Complementary Exponentiated Exponential Geometric.
CEG: Complementary Exponential Geometric
EEG: Extended Exponential Geometric.
ELL: Exponentiated Log-Logistic
EM: Expectation Maximization
EW: Weibull Exponentiated
GE2: Generalized Exponential Geometric Extreme
GEG: Generalized Exponential Geometric
GG: Generalized Gamma
HPD: Highest posterior Density
LCEG: Long-Term Complementary Exponential Geometric
LE2G: Long-Term Extended Exponential Geometric
LL: Log-Logistic
MBA: Mean Bias Absolute
MCMC: Monte Carlo Markov Chain
MH: Metropolis-Hastings
MIA: Mean Interval Amplitude
MLE: Maximum Likelihood Estimates
MOEE: Marshall-Olkin Extended Exponential
MOExp: Marshall-Olkin Exponential
MOExpExt: Marshall-Olkin Exponential Extension
MOGE: Marshall-Olkin Generalized Exponential
MSE: Mean Square Errors
TTT: Total Time in Teste
VIF: Variance Inflation Factor

Contents

CHAPTER

1

Introduction: Probabilistic Models, Genesis and Extended Summary

Statistics can be announced a science inherent to Probability Theory, a concept that in the most refined context is approached in the light of Analysis Mathematics in a branch of study known as Measure Theory and that is in constant discoveries since its origin as scientific knowledge in the first decade of the 19th century (see Laplace [32].

In the same vein, several applications are summarized in the shadow of the Probability Theory that is summarized the data analysis to explain the frequency of occurrence of the natural and/or artificial events and phenomena present in all fields of human knowledge, both in observational studies as in experiments to model randomness and uncertainty in order to estimate or enable the prediction of future observations of these phenomena, since even in the most basic data analysis the processes are still dependent on measures that can be interpreted as counting, size, mass, volume ... or any additive and multiplicative property through the fundamental principle of counting.

In this sense, the Probability Distributions (or Probabilistic Models) until then presented in the literature are intrinsic to the Theory of Probability and Statistics, and are consolidated as the platform for statistical analysis in practically all areas of study, whether in the context of the probabilistic modeling for estimate the probability of occurrence of events, as the same in the statistical inference for the decision making and/or also through statistical modeling under conclusions about a subset of representative values of a universe, the sample of a population.

1.1 Initial Considerations

Since Pierre Simon Laplace, who according to Hájek [21], is considered the precursor of probability theory, a wide range of probability distributions have been proposed to model a multitude of random phenomena, and in recent decades the statistical literature related to Survival Analysis and Reliability Theory, in addition to becoming more complex mathematically and computationally, according to Harris and Albert [23], it is enormously rich in useful results regarding not only analytical results, but also to the proposal of probability distributions due to the modeling of survival and reliability functions that, above all, are probabilistic models for a random variable T positive for a temporal phenomenon, a quantile t for the model, called failure time (or lifetime).

In survival and reliability studies, a given distribution function (or cumulative probability function), commonly denoted by F, is at the heart of the survival or reliability function, denoted respectively by S and R, which by in turn, with the probability density function (or mass probability function) f derived from F, it is the description of the hazard rate resulting from the effects of time exposures, the hazard rate function (or risk function) commonly denoted by h.

In this context, the common practice in modeling random phenomena through the results of the survival and reliability analysis consists of prior knowledge of the forms of the hazard rate function, since whatever the probability distribution from which a function S derives or R, such functions are strictly decreasing probability functions, while the h function contains five basic forms: constant, increasing, decreasing, unimodal and bathtub.

The richness of the diversity of forms that the hazard rate can take over time is the main feature in the modeling of the functions of lifetime random variables and the more shapes a model captures, the more indicated him it will be to model the failure time.

However, it is very common to observe that the greater the number of θ parameters in the parameter vector of a probability distribution, the greater the number of hazard rate forms captured by its hazard rate function.

Two-parameter models, such as those discussed in the recent studies of Ramos [50], Braguim [5] and Reis [53], respectively derived from the fundamental distributions Gama, Log-Logistic (LL) and Weibull, although limited in their characteristics and unable to show ample flexibility, they present under certain parametric conditions the forms of increasing and decreasing risk, and with the exception of LL, they still capture the constant form. But, as more parameters are inserted in these models, as developed in Stacy et al. [62], Mudholkar and Srivastava [44] and Rosaiah, Kantam and Kumar [56] to propose, respectively, the Generalized Gamma (GG), Weibull Exponentiated (EW) and Exponentiated Log-Logistic (ELL) distributions, more limitations are also inserts in its characteristics and flexibility, but more forms for the hazard rate function are now captured, such as unimodal and bathtub risks in both distributions.

Such searches have been developed over the years, either for a better exploration of the observed phenomena or to obtain distributions with more forms for hazard rate, such as obtaining bimodal risk models as developed in Mendoza, Ortega and Cordeiro [41] and, no less important, the discovery of hazard rate functions with reduced number of parameters but that captures as many forms of hazard rate as possible, as shown in the work of Ramos, Louzada and Ramos [52], where the approached distribution assumes the four basic forms in the parametric conditions of only two parameters.

For family Exponential distribution, for exemple, the obtaining procedures consist, in the most basic approach, as proposed by Gupta and Kundu [20] for a random variable T, in exponentialize the distribution function F of interest in the shape parameter $\pi > 0$ to be added, so that, with the parametric vector $\boldsymbol{\theta}$

$$\Psi(t|\boldsymbol{\theta},\pi) = [F(t|\boldsymbol{\theta})]^{\pi} \ , \tag{1.1}$$

in the domain $t, \boldsymbol{\theta}, \pi \in \mathbb{R}_+^*$ so that

$$\psi(t|\boldsymbol{\theta},\pi) = \frac{d}{d\pi}\Psi(t|\boldsymbol{\theta},\pi) = \pi[F(t|\boldsymbol{\theta})]^{\pi-1} f(t|\boldsymbol{\theta}) \ . \tag{1.2}$$

Another method, introduced by Marshall and Olkin [39], which in line that the method proposed in Gupta and Kundu [20] by introducing a new shape parameter, is more general because it does not consider restrictions for the shape parameter π, as well as for the random variable X. Generalization is also obtained from a F distribution function as

$$\Psi(x|\boldsymbol{\theta},\pi) = \frac{\pi f(x|\boldsymbol{\theta})}{1-\pi F(x|\boldsymbol{\theta})} \ , \tag{1.3}$$

in the domain of $t, \pi \in \mathbb{R}$ and $\boldsymbol{\theta} \in \Theta$, there Θ is parametric space for the parametric vector $\boldsymbol{\theta}$, and so that

$$\psi(x|\boldsymbol{\theta}, \pi) = \frac{\pi F(x|\boldsymbol{\theta})}{[1 - \pi F(x|\boldsymbol{\theta})]^2} \ . \tag{1.4}$$

See that ψ is the distribution function of a new probability distribution, provided that your first derivative exists in relation to π and in the sample space $\Omega \subseteq \mathbb{R}$, for every $\mathbb{X} \in \Omega$ event still occur

$$P(\mathbb{X}) \geq 0 \ , \tag{1.5}$$

$$P\left(\bigcup_{i=1}^{\infty} \mathbb{X}_i\right) = \sum_{i=1}^{\infty} P(\mathbb{X}_i) \ , \tag{1.6}$$

$$P(\Omega) = \int_0^{\infty} \psi(t|\boldsymbol{\theta}, \pi) dt = 1 \ . \tag{1.7}$$

Another widely applied artifice consists of mixing the f distribution with a given φ distribution to obtain a new probability distribution called composite models (or mix models).

In Tahir and Cordeiro [63], the authors scanned the literature on the analysis of survival and reliability in search of the models generated through this technique and cataloged a vast number of distributions composed since that this procedure was used in 1997 by the work of Adamidis and Loukas [3].

The authors in Tahir and Cordeiro [63] still point out that there is not only one procedure for obtaining a composite distribution, and that they follow three categories of composition of distributions that start from the simple generalization called the "first generalization approach" (or G-classes to denote the classes of generalized distributions), actually reaching the procedures properly called composite models in the "second generalization approach" (or composition) and arrive in the current procedures for obtaining composite models in the category called "recent trends in composition".

Above all, the first model composed in the literature, proposed Adamidis and Loukas [3], the Exponential Geometric model (EG), was conceived to model phenomena in function of time with three characteristics: (I) the time minimum t with (II) exponential evolution which (III) marks the n repetitions required until a failure occurs.

To model this phenomenon, the authors considered the auxiliary random variables $X \sim Exp(\lambda)$ and $N \sim Geo(\theta)$ representing, respectively, (II) any time with exponential evolution parameterized in $\lambda > 0$ and (III) the number of Bernoulli attempts required to achieve failure in analysis.

As the objective is to obtain (I) a model based on the minimum t time, the composite model (or baseline distribution as highlighted by Tahir and Cordeiro [63]) should be represented by the random variable defined as $T = \min(\{X_i\}), \forall i = 1, \ldots, n$.

In these conditions, the Exponential Geometric model comes from a population composed of 3 subpopulations, where each of them represents a particular characteristic, that is, for all $\lambda, x \in \mathbb{R}_+^*$, $n \in \mathbb{N}$ and $0 < \theta < 1$:

$$\delta(x|\lambda) = \lambda e^{-\lambda x} \qquad \text{e} \qquad \varepsilon(n|\theta) = \theta(1-\theta)^{n-1} \ . \tag{1.8}$$

As $T = \min(\{X_i\})$ has the distribution to the minimum of the set $\{X_i\}, \forall i = 1, \ldots, n$, due to the $X_i \sim Exp(\lambda)$ being independent and identically distributed, we obtain that $T \sim Exp(n\lambda)$, $\forall \ n\lambda, t \in \mathbb{R}_+^*$, so that

$$\delta_{\min}(t|n\lambda) = n\lambda e^{-n\lambda t} \ . \tag{1.9}$$

Then, if T is the baseline distribution of the composite model between it the discrete random variable N, noting that $e^{-n\lambda t} = e^{-\lambda t}e^{-(n-1)\lambda t}$, we can write

$$g(t|\theta,\lambda) = \sum_{n=1}^{+\infty} \delta(n|\theta)\epsilon(t|n\lambda) = \lambda\theta e^{-\lambda t} \sum_{n=1}^{+\infty} n[(1-\theta)e^{-\lambda t}]^{n-1} \ , \tag{1.10}$$

where $|(1-\gamma)e^{-\lambda t}| < 1$ induces the convergence of the infinite series indexed in n, so that

$$\sum_{n=1}^{+\infty} n[(1-\gamma)e^{-\lambda t}]^{n-1} = \frac{1}{[1-(1-\gamma)e^{-\lambda t}]^2} \ . \tag{1.11}$$

Therefore, the model composed of random variables $T \sim Exp(n\lambda)$ e $N \sim Geo(\theta)$, for $n\lambda, t > 0$ e $0 < \theta < 1$, respectively, has a probability density function given by

$$g(t|\theta,\lambda) = \frac{\lambda\theta e^{-\lambda t}}{[1-(1-\theta)e^{-\lambda t}]^2} \ , \tag{1.12}$$

for all $t, \lambda \in \mathbb{R}_+^*$ e $\theta \in \ ]0,1[$.

Many probabilistic distributions have been proposed as composite models based on the Geometric distribution, in addition, generalizations for these models are acquired and, as such models consider the parameter $\theta \in \ ]0.1[$ inherited from the geometric distribution, extensions for the composite models and their generalizations are also discovered when considering the complement of θ in $\mathbb{R}_+^*$.

As well as defined for the Exponential and Weibull distributions, the models derived from the EG distribution can also be understood as a family formed by the models Extended Geometric Exponential (EEG), Generalized Exponential Geometric (GEG), Complementary Exponential Geometric (CEG), Long-Term Complementary Exponential Geometric (LCEG), Complementary Exponentiated Exponential Geometric (CE2G), Exponentiated Exponential Geometric (E2G), Marshall-Olkin Generalized Exponential (MOGE) and the Generalized Exponential Geometric Extreme (GE2) distributions, proposed respectively by Adamidis, Dimitrakopoulou and Loukas [2], Silva, Barreto and Cordeiro [60], Louzada, Roman and Cancho [37], Louzada et al. [33], Louzada, Cancho and Carpenter [34], Louzada, Marchi and Roman [35], Ristic and Kundu [54] and Ristic and Kundu [55].

The family of distributions for the EG model can also be classified as distributions of two generations of models, the first with 2 parameters and the second with 3, where the first generation starts with the EG model and ends with the EEG model and the second generation starts with the E2G model and ends with the MOGE and GE2 models.

In Ristic and Kundu [54] and Ristic and Kundu [55] it is verified that, although the generalized models MOGE and GE2 (MOGE/GE2) are identified with different names and proposed on different dates, they are obtained under the same conditions, have the same parametric states and have the same expression for the distribution and probability density functions. Both works consider random variables strictly positive and parameterized in α, λ e β defined in $\mathbb{R}_+^*$.

Important extensions or generalizations were, and can be, introduced in the statistical literature, however, the EG distribution was, according to the records of Adamidis and Loukas [3], Gupta and Kundu [20] and Tahir and Cordeiro [63] the first obtained to model rates of decreasing risk under two parameters and started the modern era of probability distribution theory with the main objective of proposing and emphasizing the solution of problems faced by professionals and researchers in practically all areas of knowledge.

1.2 General Objectives

The model EEG, as previously discussed, is the extension of the EG distribution and has the function of probability density given by

$$g(t|\gamma,\lambda) = \frac{\gamma\lambda e^{-\lambda t}}{[1-(1-\gamma)e^{-\lambda t}]^2}\ , \tag{1.13}$$

for all $t, \gamma, \lambda \in \mathbb{R}_+^*$, as shown in Adamidis, Dimitrakopoulou and Loukas [2], where the decreasing and increasing forms for its function of derived hazard rate are presented. Above all, such a model can be applied in situations where the hazard rate shows little or no variation after a considerable period of evolution, as such behavior leads to the adoption of a risk model whose form captures none or the slow increase or decrease in the hazard rate.

That such form for the hazard rate can be captured by the Gamma and Weibull distributions, but in contrast to these cases for EEG, the EEG distribution may be much more appropriate due to its guaranteed flexibility in its basic properties that can be reflected for the computational gain when applied to adjust complete and censored lifetime data.

In Adamidis, Dimitrakopoulou and Loukas [2], the parameter estimation is performed using the classic approach via maximum likelihood estimators, in addition, analytical expressions for the mean and variance of the distribution are presented and the work implies that in that approach to the moments of distribution, such expressions are defined only in the case where $\gamma < 1$. The expectation maximization algorithm (EM) is presented to calculate the estimates of the parameters of this distribution, and in Ramos et al. [51] a bayesian analysis is developed considering complete data with different previous noninformative distributions, but in both works nothing is clarified about the r-th moment in the event that $\gamma > 1$.

Although the distribution is characterized by its mean residual lifetime for all $\gamma \in \mathbb{R}-\{1\}$, as the first objective of this work we will present a study for the moments of the EEG distribution in the general case, specifically when $\gamma \geq 1$.

It is obvious that, since its disclosure, the EEG distribution is already consolidated in the works identified in the analysis of survival and reliability, since under the consideration of censorship, as shown in the works Anwar et al. [4], Mirjalili [42], Dey et al. [11] and Abujarad et al. [1], respectively, with a type-II, progressive type-II, progressive type-I hybrid and censorship on the right approach, it is verified that this distribution is applied in a refined way in relation to the types of censorship considered both in the context of survival and reliability.

However, the approach according to random censorship is not found in the literature, and in none of these lines of application, therefore the second objective of this study is to obtain the estimates for the parameter assuming different estimation methods for the distribution parameters of EEG under randomly censored samples. In this case, the main objective is to study the bayesian estimates considering different noninformative prior distributions and to contrast them with the maximum likelihood estimate.

As the bayesian estimates and marginal posterior densities cannot be obtained in a closed form, we carry out a bayesian analysis for the EEG distribution using Markov Chain Monte Carlo (MCMC) methods (see , Gelfand and Smith [15]; or Chib and Greenberg [8]) to obtain the posterior summaries of interest. Since bayesian analysis remains a new approach proposal for this distribution, as verified by the main works cited here, in the present work we limit ourselves to applying inferences in this context only under non-informative priori.

Numerical integration based on stochastic simulation methods as the MCMC will be used to simulate samples of the marginal posterior distribution of interest and in particular, we will be using the Metropolis-Hastings (MH) algorithm to obtain the posterior summaries of interest.

As a final proposal for an approach to this distribution, we will consider the presence of covariates in conjunction with random censorship and with the objective of presenting the characteristics of this distribution under the presence of this double influence, in this approach we consider it essential to develop the same approach used for the distribution in the case randomly censored and without covariate, that is, the approach to the moments of the EEG distribution in the presence of random and convariable censorship, as well the obtaining their estimators will follow the same conditions, with the estimators obtained in the cases classic and bayesian in order to be contrasted possible applications.

With similar objectives, in this work we will also consider the distribution that generalizes the second generation of the EG distribution, as presented in the previous section, the MOGE-GE2 distributions, whose probability density function is given by

$$f(t|\alpha,\lambda,\beta) = \frac{\alpha\lambda\beta e^{-\lambda t}(1-e^{-\lambda t})^{\alpha-1}}{[\beta+(1-\beta)(1-e^{-\lambda t})^{\alpha}]^2}\ , \tag{1.14}$$

for all $t, \alpha, \lambda, \beta \in \mathbb{R}_+^*$.

As it is highlighted in this work that the MOGE/GE2 distributions are the same and are being approached as members of the EG family, we will from here on follow the common notation that the literature presents for the EG distribution family, as discussed in the previous section, we will follow referring to the MOGE/GE2 distributions only as GE2.

From the works of Ristic and Kundu [54] and Ristic and Kundu [55] to the most recent, such as the current ones Khan, Akhtar and Ali Khan [45], Dey et al. [11], Torabi, Bagheri and Mahmoudi [64] and Abujarad et al. [1], there is no significant approach to the moments of the GE2 distribution.

Although in Khan, Akhtar and Ali Khan [45] the authors have announced a study in the light of bayesian inference for Marshall-Olkin family of distributions, which has admirably addressed the moments in the shadows of Laplace Approximation, this is did for Marshall-Olkin Exponential (MOExp) and no reference is given for the GE2 distribution.

In Dey et al. [11], although a study is presented for the model considered through the refined application of type-II progressive censorship, the work considers an approach to the Marshall-Olkin Extended Exponential (MOEE) distribution and is also dont't move of no measure to the GE2 model.

In the work of Torabi, Bagheri and Mahmoudi [64], despite the fact that the authors actually approach the GE2 model, they do so with the parameters strictly limited to $\alpha = 1$ and $\beta \in \,]0,1[$ for all $\lambda > 0$, reducing the model $GE2(1,\lambda,\theta)$ for the GE distribution. Therefore, the authors present a study that is limited to an application for the GE distribution to complete data, where there is no study for the models of the GE2 distribution in the general case.

In the most recent work, presented by Abujarad et al. [1], although the GE2 distribution proposed by Ristic and Kundu [54] and Ristic and Kundu [55] is indeed explored, the authors invoke this model under the reparametrization $\lambda = \theta^{-1}$ for all $\theta > 0$, and insert a third name for this distribution as Marshall-Olkin Exponentiated Exponential (MOExp-Exp) and develop the approach, together with the MOExp models and the Marshall-Olkin Exponential Extension (MOExpExt), in line with a simulation study for censored data on the right, with application to real censored data. However, they also do not mention anything about the moments of the three models covered.

From what the literature provides, there is a low, or negligible, approach to the GE2 distribution, presenting a reasonable characteristic for this model. Therefore, in this work we bring a Mathematical approach on the r-th moment of this distribution, in order to characterize it as a probability distribution.

In addition, we will present an approach under random censorship for its estimators and, under the same conditions described above on the proposed objectives for the EEG distribution approach, this study will be developed considering the classical and bayesian estimators with vague prioris, with applications to available real data in the literature on survival analysis.

Both in the approach of the EEG distribution and in the GE2, the study of their respective estimators will be developed computationally by the justification that these models are semi-pathological, as we will show in the approach of their respective moments.

Therefore, we consider a simulation study for both models and following the forms of their respective hazard rate functions, that is, we will simulate the EEG in the parametric condition in which its risk is decreasing and increasing, while for the GE2 model the simulation it is performed for cases in which it manifests the decreasing, increasing, unimodal and bath shape for its hazard rate function.

1.3 Dissertation Structure

In detail, this work is organized as that is describe in the sequence.

In the chapter 2, in the section 2.1, the properties of the EEG model are reviewed in the context of survival analysis. A lemma that guarantees obtaining the mean and variance for this model, in the conditions presented by the author in Adamidis, Dimitrakopoulou and Loukas [2] and, consequently, a theorem that reinforces the semi-pathology result for EEG when $\gamma > 1$ são demonstrados, as well as a theorem that guarantees that this model tends to lose memory as time progresses, a result hitherto commented on previous works and, almost always, highlighted geometrically in the approaches for this distribution family.

In the section 2.2, the maximum likelihood estimators for the parameters are presented, as well as a reference to Fisher's information matrix obtained by Kitidamrongsuk et al. [30] in the specific case in which $\gamma < 1$ is described. Subsequently, a bayesian approach to this model is introduced under the application of vague prioris as a result of the product of independent Gamma distributions for its joint composition.

In the subsection 2.2.3, we present the simulation for the EEG model in the presence of random censorship, in 3 particular cases for this category of censorship, with 4 different cases of sample size and two parametric cases to simulate the manifestation of these scenarios with the two forms of the function of risk that this model assumes.

The subsection 2.2.4 describes an application for the EEG model in the presence of censorship for a modeling problem with a set of 137 observations from patients with lung cancer, where the parametric models Gamma, LL and Weibull are taken in competition with to the EEG model for modeling under the classical conditions of the model estimators. A random sample of 40 observations was also considered for modeling, however, under bayesian conditions to highlight a contrast between these two approaches for the model's estimators.

The section 2.3.2 presents the proposed distribution approach in the presence of covariates and censored times, so that a second source of variation is mathematically defined for the density function and, consequently, defined the survival and risk function highlighting that with two sources of variation these three functions geometrically represent surfaces in the space $\mathbb{R}^3$. Still in the section 2.3.2, a theorem is presented to show that the property of tendency to memory loss also manifests itself in the presence of covariates, however, in this context, the risk function tends for a curve in the $\mathbb{R}^3$.

Following the theorem, the maximum likelihood equation, as well as its system of maximum likelihood equations, for the $p+1$ parameters of the model in the presence of censorship are presented without significant change, maintaining the same properties that in the case considered with a source of variation as shown in the analytical expression presented in the section 2.2, which does not occur in the bayesian approach that is introduced with vague priori and also as the product between the independent Gamma distribution for the parameter γ and the Normal distribution for each of the $p+1$ parameters β_j in the covariate term.

In the subsection 2.3.1, the simulation for the EEG model in the presence of covariates and censored time is also developed in 3 particular cases of random censorship, with 4 cases of different sample size and two parametric cases for simulate the model according to forms for the risk function that remain in the same conditions as in the case of a source of variation, regardless of the composition of the second source in the model.

Subsequently, the subsection 2.3.2 deals with application that seeks to adjust survival and risk models for two groups of patients undergoing different lung cancer treatments. In this application, a group presents proportional risk, which is why the Cox regression model was used to contrast the adjustment implemented by the EEG model in the presence of censored and covariable data.

The application shows that, despite the flexibility that the Cox model offers, although improved with the application of the cubic spline to smooth the function of survival and risk, as we consider in two nodes, the EEG model still presented better results, adjusting efficiently to the empirical risk manifested by the censored data. In the second group, the same performance is observed in contrast to the Weibull model.

In the chapter 3, a generalization for the EEG model is reviewed in the theoretical context with a simulation and an application presented. This chapter has been divided into 2 sections.

In the first, the section 4.1 seeks to review the origin of this three-parameter model, highlighting its origin and consequent functions in the context of survival analysis, where the descending, crescent, unimodal and bathtub shapes are notable as illustrated in the figure 3. 1 displayed.

In the following subsection, in 3.1.1, a significant mathematical approach is presented for the function of r-th moment for the approached distribution, presenting as results three mottos, a proportion and a theorem to highlight a study of the moment of r-th order of this model, considering its possible parametric compositions and its memory loss tendency property inherited from its original distribution.

In the subsection 3.1.2 we will also see, in the classic aspect, the estimators for the parameters of this model and, in the same way, in the subsection 3.1.3, the bayesian approach is considered under vague prioris through the product independent range distributions.

In the sequence, the section 3.2 displays and discusses the results obtained in the simulation study that considered three particular cases of censorship, with five different sample size cases and four parametric cases to simulate and evaluate the estimators of this model approached according to each of the forms of risk function that the model provides.

The sequence application, shown in the section 3.3, again shows the competition between the three-parameter model approached with competing models. The models considered were obtained under the same conditions as the one approached, making the contrast in equality not only due to the parametric quantity and composition of the hazard rate function shape, but because they have different conditions of flexibility and analysis.

In this work, we try to show that the relevance of the EEG and GE2 models does not consist only of their good adjustments in statistical modeling applications, but that their relevance is highlighted due to the fact that their properties inherited from the Exponential and Geometric distributions, that this relevance consists of the fact the compositions, extensions or generalizations of models derived from Exponential and Geometric distribution. That although most of its analytical results are intractable as a result of the introduction of shape parameters, the flexibility inherited from its source distributions injects highly relevant benefits to the computational approach in real life-time data.

1.4 Other Considerations

The bayesian procedures include several statistical diagnostic tests witch seek to assess the convergence of the markovian chain, this work consider the Heidelberger-Welch stationarity test, known as the heidel diagnostics, which uses the Cramer-von-Mises statistic to test the hypothesis that the chain values provide a stationary distribution.

The diagnosis of heidel is based on the work of Heidelberger and Welch [24], Schruben [59] and heidelberger and Welch [25], and the results of this test with a given significance will shown in the tables by the columns test-stat, p-value, and test result columns, for show that all Markov chains from MCMC processes converged or not.

Since we consider adjustments by the performed using bayesian estimates, to compare these adjustments the most appropriate adjustment measure is the DIC (Deviance Information Criterion) which generalizes the Akaike, AAIC and Schwarz Information Criterion, respectively the AIC, AICC and BIC, to evaluate fitted models with calculated sample estimates via posterior distributions.

In the context of model comparison by the DIC measure, the best-fit model is indicated by penalties that the others suffer through the effective number of parameters and deviance, the statistics commonly indicated by p_D and D, respectivaly, and obtained by summing the squares of the linear regression residues which, according to Collett et al. [9], summarizes to what extent the fit of a current model of interest diverges from a model that is assumed to be a perfect fit to the data.

Suitable for selection problems of bayesian models in which the posterior distributions of the models are obtained via MCMC simulation, when the p_D decreases smaller is the complexity of the model andthe DIC decreases by pointing the best fit, then, the lower p_D and DIC, the better is adjustment. Then, defined

$$\bar{D}(\theta) = E[-2\log(p(y|\theta))] \quad \text{and} \quad D(\hat{\theta}) = -2\log(p(y|\hat{\theta})) \, , \tag{1.15}$$

from calculation of p_D as $p_D = \bar{D}(\theta) - D(\hat{\theta})$, the DIC can be obtained in two ways: according to Spiegelhalter et al. [61] as $DIC = p_D + \bar{D}(\theta)$, or according to Gelman et al. [16] as $DIC = 2p_D + D(\bar{\theta})$.

In the applications from the sections 2.2.4 and 2.3.2, purposing to fit an ideal probabilistic model for the censored lifetimes for the patients according to two chemotherapeutic agents, standard and test, based on the absence and presence of covariates, respectively, a same set of data is considered to discuss possible approaches and results around probabilistic modeling and to point out the benefits and penalties in modeling with data with and without variables.

The data was presented by Prentice [48], record the lifetime of 137 patients with advanced lung cancer and was recorded in the presence of covariates when the patient was taken to the study. Tumors are classified into four groups: squamous (1), small (2), adeno (3) and large (4).

The covariates considered in original study include performance status, a measure of overall medical status for the patient on a scale of 10 out of 10 units from 10 to 90, where 10, 20, and 30 indicate that the patient is fully hospitalized, 40, 50 and 60 indicate a partial confinement in the hospital and 70, 80 and 90 indicate that the patient is able to take care of himself.

Time in months from diagnosis to study initiation, age in years, and any therapy prior to the study (1-yes ou 0-not) were also considered with as covariates in study 2.

In applications, the objective is to model the lifetime of standard and test chemotherapeutic agents and, in the case of the presence of a vector $\boldsymbol{x}$ of covariates, is considers the tumor group as a new covariate. So, denoting as X_1, X_2, X_3, X_4 and X_5 the covariates of the study, where

- X_1: the tumor group with values 1, 2, 3 or 4
- X_2: the medical status with values 10, 20, ..., 80 or 90
- X_3: diagnostic time in months with values $x_{i,3} \in \mathbb{Z}_*^+$
- X_4: the age in years with values $x_{i,4} \in \mathbb{Z}_*^+$
- X_5: if you have had therapy with values 1 or 0

The covariate approach is developed with the covariates indexed in the scale parameter of the model under study. This form the presence of covariates in the probabilistic model integrates into its variability a source of variation with origin in additive effect index parametric that not require the scale parameter coupled in the model as a random variable, but a simple source of variation that integrates effect the covariate in the model.

Therefore, in this approach, the modeling concept considers that an additive model $\nu = \langle \boldsymbol{x}_k, \boldsymbol{\beta} \rangle + \epsilon$ contributes to the probabilistic variation of the model, together with the variable random that represents the lifetime, but as the scale parameter instead of being coupled in the model as a random variable.

The commonly used models that consider the presence of covariates, generally statistical models of regression or joint probability distribution, although powerful, require a delicate or complex approach before, during and after the modeling process.

To circumvent this situation and also find a more flexible model, Cox [10], proposed a semi-parametric model called the proportional risk model that became the most used in the analysis of survival data, under the covariable aspect, due its versatility. The general expression of the Cox model, for $\nu = \langle \boldsymbol{x}_k, \boldsymbol{\beta} \rangle$, is given by

$$h(t|\boldsymbol{x}, \boldsymbol{\beta}) = h_0(t)\phi_{\boldsymbol{\beta}}(\nu) \ , \tag{1.16}$$

on what $h_0(t|\boldsymbol{\beta})$ is the nonparametric component called the hazard rate function and $\phi_{\boldsymbol{\beta}}(\nu)$ is the parametric term, both non-negative on the condition that $\phi_{\boldsymbol{\beta}}(0) = 1$.

The parametric term is almost always applied in the multiplicative form as

$$\phi_{\boldsymbol{\beta}}(\nu) = e^{\nu}. \tag{1.17}$$

With the same objective, to make the survival data modeling more flexible, the use of spline functions is proposed, since this function also allows the investigation of nonlinear effects and the evaluation of time interactions in the presence the covariates. According to Heinzl and Kaider [26], modeling under the application of spline functions provides two main advantages: no specific functional form needs to be specified and its application is completely computational, excluding applications of algebraic concepts.

In addition, Durrleman and Simon [12] proposes the use of spline functions to detect nonlinear relationships between covariates and Cox model response, and in the same vein are used in Hess [27] and Heinzl, Kaider and Zlabinger [22] to evaluate and identify interactions between time and covariates.

As if the flexibility that Cox's model itself provides does not suffice, using spline functions on the Cox model, it is possible to obtain a parametric form from and which simulate survival times with adjust survival and risk functions with refined precision.

To obtain a flexible model for the hazard rate, Royston and Parmar [57] modeled the logarithm of the risk-risk function as a cubic spline function of the time logarithm as

$$ln[H(t|\boldsymbol{x},\boldsymbol{\beta})] = ln[H_0(t)] + \nu \ , \tag{1.18}$$

where H is the empirical cumulative risk function, $\boldsymbol{x}$ is the vector of covariates, $\boldsymbol{\beta}$ is the vector of parameters and $\nu = \langle \boldsymbol{x}_k, \boldsymbol{\beta} \rangle$ is the additive model.

In this context, according to Lambert and Royston [31], the s spline function is applied over the $ln(t)$ function with the nodes $\kappa_0 = (\kappa_1, \ldots, \kappa_m)$. In this case s can be written as $s[ln(t)|\boldsymbol{a}, \kappa_0]$, with node location κ_0 no coefficiente $\boldsymbol{a} = (a_1, \ldots, a_m)$, and then used for the logarithm of the accumulated baseline risk in the proportional accumulated risk model of the expression (1.16) to obtain (1.18) rewritten as

$$ln[H(t|\boldsymbol{x},\boldsymbol{\beta})] = s[ln(t)|\boldsymbol{a}, \kappa_0] + \nu \ , \tag{1.19}$$

that taken on the scale of survival and hazard rate, provides

$$S(t|\boldsymbol{x},\boldsymbol{\beta}) = e^{-e^{\nu}} \quad \text{and} \quad h(t|\boldsymbol{x},\boldsymbol{\beta}) = \left\{ \frac{d}{dt} s[ln(t)|\boldsymbol{a}, \kappa_0] \right\} e^{\nu} \ , \tag{1.20}$$

because the s spline function is a cubic polynomial and the relations between the functions h and S, in particular with $h(t) = H'(t)$, $S(t) = e^{-H(t)}$ and $H(t) = -ln[S(t)]$ under derivative and integral operators in relation to t.

Thus, according Rutherford, Crowther and Lambert [58] over the Cox models, the spline function can be considered a generalization of Weibull's proportional hazards rates model and is reduced to the Weibull model with only two nodes.

Therefore, in this application a contrast between the EEG model in the presence of covariates with the Cox/Spline$_2$ model, the Cox model under two-node spline function adjustment, it is considered.

In the application 2.3.2 the covariables X_3 and X_4 are integers, the 5 covariables of the problem can be considered categorical, therefore, graphical methods to find proportional risks are not feasible for the application, instead they are considered a statistical test based on Schoenfeld residues to test the overall proportionality for risk functions.

CHAPTER

4

Final Results and Conclusions

4.1 Final Results and Conclusions

The present work does not present a chapter for the theoretical framework or basic concepts used for the proposals made, but sought, when a given concept was used, to refer to it and describe its relevance with subtlety.

To compensate for this absence, in the introduction to this work, the most relevant concepts are described through four subsections. In particular, a bibliographic review that highlights the genesis of the models studied, mainly for the distribution from which they originate, was presented in the subsection 1.1.

The subsections 1.2 and 1.3 presented a detailed summary of the intended results in the present study and described the structure of how the work was developed, pointing out what and how the approaches were carried out. In the subsection 1.4 we sought to give a more detailed focus to the concepts that are less relevant to the study's proposals, but which were applied in parallel to the proposed objectives for obtaining the results and even for the applications considered.

Under valuable mathematical concepts, where theoretical points from different lines of study were approached such as those of Infinite Series, Special Functions and the Mathematical Analysis itself which somehow absorbs these fields, the properties of the models were demonstrated, proposing them exclusively to support in processes from the resoluction in real problens with these models.

However, the work sought to present the approach of these problems according to the assumed properties and, in some cases, demonstrated for the models approached here. The property of as constitutes the forms of hazard rate, derived from the density and survival functions of the models, conditioned to the scale parameter inherited from the Geometric distribution, was the main one.

It has also been seen, and demonstrated, that the extensions for the Geometric Exponential distribution provide hazard rate function that inherits the memory loss property of the Exponential and Geometric functions that generated them, as well as the relationships that their respective parameters have for a given form from the hazard rate function, making clear the parametric conditions under which they are generated.

Particularly, regarding the considered EEG models, varieting in time only and/or in the time with presence the covariables, it was clearly highlighted that, in both cases approached, whatever the value of the $\lambda > 0$ parameter assumed, the hazard function derived from this model has a decreasing behavior for cases where $\gamma < 1$ and increasing when $\gamma > 1$, where λ is the scale or rate parameter of the model that, becouse by construction is inherited from the Exponential distribution, and γ is its shape parameter. When $\gamma = 1$ the EEG model is reduced to the Geometric Exponential model.

Similarly, regarding the GE2 model, it was clarified that for any assumed $\lambda > 0$ value, when $\alpha < 1$ the resulting hazard rate function takes the decreasing form for any $\beta < 1$ and bathtub form for any $\beta > 1$, but in the case where $\alpha > 1$, the hazard rate is unimodal when $\beta < 1$ and increasing if $\beta > 1$.

Still on the hazard rate forms of the GE2 model, in the cases where (1) $\alpha = \beta = 1$, which (2) $\beta = 1$ for any $\alpha > 0$ with $\alpha \neq 1$, and that (3) $\alpha = 1$ for any $\beta > 0$ with $\alpha \neq 1$, the GE2 model is reduced to the Geometric Exponential model. In (1) and (2) this is immediate by substitution, as can be seen in [3], and in (3) this is guaranteed in [35].

Under the parametric conditions evidenced for the EEG and GE2 models, the computational approach is simplified, since in the bayesian case, for example, an appropriate initial kick is needed to calculate model estimates via MCMC, and knowing a favorable region for this kick, the markovian process reaches its convergence faster.

Therefore, knowing the hazard rate form for a given data set and, consequently, where the scale or rate parameter from model EEG converges, the initial kick to the γ estimate is simplified because of the region for this parameter to be previously known. The same way that is for the kicks for the estimates of α and β in the GE2 model, in the both cases the kicking λ are made according to the inherited memory loss property.

Also worth highlighting from the conclusions obtained about the properties demonstrated, the fact of the EEG and GE2 distributions are semi pathological in consequence from the yours parameters γ and β, respectivaly. Depending of the parametric condition that is considered for γ or β, exist or not a mean and variance for distribution EEG or GE2, and in this constatation we prove the parametric conditions under which we can guarantee that the expected value exists for the EEG and GE2 distribution, obtaining it in terms of infinite series.

The work shows that, equipped with these propertie, the use r-th moment it is eventually impracticable, mainly for distribution GE2 by virtue your expression, and even though it exists under certain parametric conditions the expected value and variance for EEG and GE2, presuppositions for important statistical concepts, such as expected Fisher's matrix, the Central Limit Theorem and the moments of order r are not achieved under in the complementary cases.

Based on elaborate computational resources to support the approaches considered in the study of these models, the work considers the outline of the geometric shapes of some functions, such as the confidence bands presented in the applications and in particular for the density, survival and hazard surface considered for the EEG model in the presence of covariables. Due to these computational resources, as a consequence of the obtained properties, the simulation study was concluded as follows.

The elaboration of the statistical simulation under the classical and bayesian approach via MLE and MCMC, respectively, required delicate computational processes for the extraction and calculation of statistics relevant to the results evaluation criteria. As described in chapter 1, this work was developed focusing on the study area of survival analysis and in these simulations were considered at least 4 samples sizes and 3 cases of censorships, one of them being the case for complete-time data.

As shown, we present the results of 3 simulation studies developed for (1) EEG model in the presence of censorship, (2) for the EEG model in the presence of censorship and multiple variables and (3) for the GE2 model in the presence of censorship. The 3 case studies were chosen according to the forms of the hazard function derived from the respective simulated models.

All 3 results were satisfactory, although with specific features. It is worth mentioning the slow convergence of the probability of empirical coverage in the case of the simulation (3) described in the subsection 3.2, where both bayesian and classical estimators showed equal conditions and properties over their estimates in any parametric case, censorship ratio and samples larger than 20 observations. Particularly, in small samples the bayesian estimators were superior in all the considered criteria.

In the subsection 2.2.3, the case of simulation (1) shows that, despite the high quality of the estimates obtained by the MLE, this quality was attributed to large samples and, in cases of low censorship, those lower than $tau = 0.20$. For these estimators, it was common to observe reasonable situations with relevant sample sizes and censorship ratios, such as $n = 50$ and $\tau = 0.2$, where estimates attributed high estimation errors, pointing to an estimated inefficient.

In this cases, the simulation study highlights that bayesian estimators are significantly superior to the classical one. Although it shows high estimation errors for high censorship ratios, it is still significantly better than MLE in these situations.

When the same model was simulated, also in the presence of censorship but considering multiple variables, as developed in simulation (2), the efficiency of the bayesian and classical estimators are analogous, and the results highlight that both the MLE and MCMC methods are satisfactorily accurates in the model in the presence of censorship and multiple variables, even with high proportion of censorship, showing that in both inference cases the presence of covariables in the model attributes high flexibility and excellent results in relation to their estimators.

With regard to the simulations developed, it is evident that while MLE is preferable for its speed in large samples, MCMC is preferable for its accuracy in small sample sizes, but what is noteworthy is that the presence of multiple variables compensates the penalty that estimates suffer by high censorship ratios, as shown in simulations (1) and (2).

Above all, the 3 studies allow to conclude that the estimators of these models are appropriate in the considered sample conditions and in view of their particular properties, as the biparametric EEG model in the presence of censorship and in the presence of covariables with $p + 1$ parameters in the presence of censorship or the GE2 model as a 3 parameter model and under censorship.

Focused in the EEG parametric model, the work presents the applications for this model in the presence of censorship and under censorship in the presence from covariable, respectively in the sub sections 2.2.4 and 2.3.2 . In these applications it is shown that this model is preferable in the contrasts performed, which was possible to conclude with its obtained properties, as described in 2.3.2 where, even though enhancing the specific competing model for a case of the proportional hazard rate, the EEG model presents a superior adjustment, because besides capturing the empirical hazard form presented by the data, it fits the empirical hazard with interpolation and adjustment satisfactory in the confidence range considered.

Similarly, in the case of the tri-parametric model, the application in the subsection 3.3 highlights that the GE2 model is also preferable over its competitors, that although they present excellent adjustments and capture the form of the risk rate function manifested empirically, all the information criteria presented, as well as the average distance between the survival function adjusted with the Kaplan-Meyer empirical, are the best because they are the smallest among all the models considered, which highlights the efficiency of the GE2 model over the others.

Bibliography

[1] M. H. AbuJarad and et al. Bayesian reliability analysis of marshall and olkin model. *Annals of Data Science*, pages 1–29, 2019.

[2] K. Adamidis, T. Dimitrakopoulou, and S. Loukas. On an extension of the exponential-geometric distribution. *Statistics & probability letters*, 73(3):259–269, 2005.

[3] K. Adamidis and S. Loukas. A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters*, 39(1):35–42, 1998.

[4] S. Anwar, S. Shahab, and A. U. Islam. Accelerated life testing design using geometric process for marshall-olkin extended exponential distribution with type ii censored data. *International Journal of Scientific Engineering and Technology*, 3(5):538–542, 2014.

[5] S. D. Braguim. Modelo hierárquico log-logístico: uma aplicação no estudo da linezolida. Master's thesis, Universidade Estadual de Maringá, 2015.

[6] T. J. I. Bromwich. *An Introduction To The Theory Of Infinite Series.* Macmillan and Company, 1908.

[7] S. Chakraborti, F. Jardim, and E. Epprecht. Higher-order moments using the survival function: The alternative expectation formula. *The American Statistician*, 73(2):191–194, 2019.

[8] S. Chib and E. Greenberg. Hierarchical analysis of sur models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, 68(2):339–360, 1995.

[9] D. Collett. *Modelling Survival Data in Medical Research.* Chapman and Hall/CRC, 2015.

[10] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[11] S. Dey and et al. Estimation and prediction of marshall–olkin extended exponential distribution under progressively type-ii censored data. *Journal of Statistical Computation and Simulation*, 88(12):2287–2308, 2018.

[12] S. Durrleman and R. Simon. Flexible regression models with cubic splines. *Statistics in medicine*, 8(5):551–561, 1989.

[13] A. Erdélyi. *Higher Transcendental Functions.* McGraw-Hill, 1953.

[14] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman and Hall/CRC, 2006.

[15] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

[16] A. Gelman and et al. *Bayesian Data Analysis.* Chapman and Hall/CRC, 2013.

[17] M. S. Goodman, Y. Li, and R. C. Tiwari. Survival analysis with change point hazard functions. 2006.

[18] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, 2014.

[19] J. Guillera and J. Sondow. Double integrals and infinite products for some classical constants via analytic continuations of lerchś transcendent. *The Ramanujan Journal*, 16(3):247–270, 2008.

[20] R. D. Gupta and D. Kundu. Exponentiated exponential family: an alternative to gamma and weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43(1):117–130, 2001.

[21] A. Hájek. Interpretations of probability. 2002.

[22] H. Harald, A. Kaider, and G. Zlabinger. Assessing interactions of binary time-dependent covariates with time in cox proportional hazards regression models using cubic spline functions. *Statistics in medicine*, 15(23):2589–2601, 1996.

[23] E. K. Harris and A. Albert. *Survivorship Analysis for Clinical Studies*, volume 114. CRC Press, 1990.

[24] P. Heidelberger and P. D. Welch. A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM*, 24(4):233–245, 1981.

[25] P. Heidelberger and P. D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983.

[26] H. Heinzl and A. Kaider. Gaining more flexibility in cox proportional hazards regression models with cubic spline functions. *Computer methods and programs in biomedicine*, 54(3):201–208, 1997.

[27] K. R. Hess. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in medicine*, 13(10):1045–1062, 1994.

[28] L. Hong. A remark on the alternative expectation formula. *The American Statistician*, 66(4):232–233, 2012.

[29] A. Jeffrey and D. Zwillinger. *Table of Integrals, Series, and Products.* Elsevier, 2007.

[30] P. Kitidamrongsuk. Discriminating between the extended exponential geometric distribution and the gamma distribution. *Unpublished Doctoral Dissertation). Assumption University of Thailand*, 2010.

[31] P. C. Lambert and P. Royston. Further development of flexible parametric models for survival analysis. *The Stata Journal*, 9(2):265–290, 2009.

[32] P. S. Laplace. *Philosophical Essay on Probabilities: Translated from the Fifth French Edition of 1825*, volume 13. Springer Science & Business Media, 1998.

[33] F. Louzada and et al. A new long-term lifetime distribution induced by a latent complementary risk framework. *Journal of Applied Statistics*, 39(10):2209–2222, 2012.

[34] F. Louzada, V. A. A. Marchi, and J. Carpenter. The complementary exponentiated exponential geometric lifetime distribution. *Journal of Probability and Statistics*, 2013, 2013.

[35] F. Louzada, V. A. A. Marchi, and M. Roman. The exponentiated exponential-geometric distribution: a distribution with decreasing, increasing and unimodal failure rate. *Statistics*, 48(1):167–181, 2014.

[36] F. Louzada, P. L. Ramos, and G. S. C. Perdoná. Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data. *Computational and Mathematical Methods in Medicine*, 2016, 2016.

[37] F. Louzada, M. Roman, and V. G. Cancho. The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart. *Computational Statistics & Data Analysis*, 55(8):2516–2524, 2011.

[38] J. Loxton. Special values of the dilogarithm function. *Acta Arithmetica*, 43(2):155–166, 1984.

[39] A. W. Marshall and I. Olkin. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84(3):641–652, 1997.

[40] C. A. McGilchrist and C. W. Aisbett. Regression with frailty in survival analysis. *Biometrics*, pages 461–466, 1991.

[41] N. V. R. Mendoza, E. M. M. Ortega, and G. M. Cordeiro. The exponentiated-log-logistic geometric distribution: Dual activation. *Communications in Statistics-Theory and Methods*, 45(13):3838–3859, 2016.

[42] M. Mirjalili, H. Torabi, and H. Nadeb. Stress-strength reliability of exponential distribution based on type-i progressively hybrid censored samples. *Journal of Statistical Research of Iran JSRI*, 13(1):89–105, 2016.

[43] M. Mood, D. C. Boes, and F. A. Graybill. *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.

[44] G. S Mudholkar and D. K. Srivastava. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2):299–302, 1993.

[45] K. Najrullah, A. Tanwir, and A. K. Atha. Bayesian analysis of marshall and olkin family of distributions. 8(7):18692–18699, 2017.

[46] P. J. Olver. *Classical Invariant Theory*, volume 44. Cambridge University Press, 1999.

[47] L Pochhammer. Über hypergeometrische funktionen höherer ordnungen. *Crelles J*, 71:316–352, 1870.

[48] R. L. Prentice. Exponential survivals with censoring and explanatory variables. *Biometrika*, 60(2):279–288, 1973.

[49] F. Qi, X. T. Shi, and F. F. Liu. Several identities involving the falling and rising factorials and the cauchy, lah, and stirling numbers. *Acta Universitatis Sapientiae, Mathematica*, 8(2):282–297, 2016.

[50] P. L. Ramos. Aspectos computacionais para inferência na distribuição gama generalizada. 2014.

[51] P. L. Ramos and et al. Bayesian analysis of the generalized gamma distribution using non-informative priors. *Statistics*, 51(4):824–843, 2017.

[52] P. L. Ramos, F. Louzada, and E. Ramos. Posterior properties of the nakagami-m distribution using noninformative priors and applications in reliability. *IEEE Transactions on reliability*, 67(1):105–117, 2017.

[53] T. C. S. Reis. Extensões da distribuição weibull aplicadas na análise de séries climatológicas. 2017.

[54] M. M. Ristić and D. Kundu. Marshall-olkin generalized exponential distribution. *Metron*, 73(3):317–333, 2015.

[55] M. M. Ristić and D. Kundu. Generalized exponential geometric extreme distribution. *Journal of Statistical Theory and Practice*, 10(1):179–201, 2016.

[56] K. Rosaiah, R. R. L. Kantam, and S. Kumar. Reliability test plans for exponentiated log-logistic distribution. *Economic Quality Control*, 21(2):279–289, 2006.

[57] P. Royston and M. K. B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, 21(15):2175–2197, 2002.

[58] M. J. Rutherford, M. J. Crowther, and P. C. Lambert. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, 85(4):777–793, 2015.

[59] L. W. Schruben. Detecting initialization bias in simulation output. *Operations Research*, 30(3):569–590, 1982.

[60] R. B. Silva, W. B. Souza, and G. M. Cordeiro. A new distribution with decreasing, increasing and upside-down bathtub failure rate. *Computational Statistics & Data Analysis*, 54(4):935–944, 2010.

[61] D. J. Spiegelhalter and et al. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

[62] E. W. Stacy. A generalization of the gamma distribution. *The Annals of mathematical statistics*, 33(3):1187–1192, 1962.

[63] M. H. Tahir and G. M. Cordeiro. Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications*, 3(1):13, 2016.

[64] H. Torabi, F. L. Bagheri, and E. Mahmoudi. Estimation of parameters for the marshall–olkin generalized exponential distribution based on complete data. *Mathematics and Computers in Simulation*, 146:177–185, 2018.

[65] D. Zagier. The dilogarithm function. In *Frontiers in number theory, physics, and geometry II*, pages 3–65. Springer, 2007.